

The Value of Human Expertise

Bradley Sturt

August 23, 2026

Abstract

We consider optimization applications with unknown parameters where the decision maker believes that the optimal value of the nominal problem—the optimization problem they would have solved if the true parameters were known—is unlikely to be large. This belief derives from information that humans have that is not captured in datasets, obtained from domain knowledge and interacting with the physical world. We propose an approach to evaluating policies that provides tighter performance guarantees if the decision maker’s belief happens to be correct. Our main result shows that if computing a policy’s worst-case performance is a convex program, then the *value of human expertise*—the maximum improvement in performance guarantees that can be obtained from the belief about the nominal problem—is equal to the minimax gap of a max-min problem. We illustrate our developments in assortment optimization and shortest path problems.

Keywords— Optimization under uncertainty; hyperlocal revenue management; private information.

1 Introduction

Consider a parent company that seeks to optimize product assortments at individual brick-and-mortar stores. The parent company has access to the transactional sales data generated by a store’s past assortments. However, because the number and variety of the store’s past assortments are limited, the parent company has insufficient data to estimate an accurate store-specific discrete choice model or identify an assortment that can be trusted to outperform the store’s best past assortment. That said, the past assortments were not chosen *randomly*: rather, the store manager chose the past assortments using private information about local consumer preferences obtained from observing and interacting with customers that visit the store. As a result, the parent company believes it is unlikely that the store’s best past assortment was highly suboptimal. *Can that belief help the parent company identify a new assortment that outperforms the store’s best past assortment?*

The above example is an instance of a new class of problems that we refer to as *optimization with human expertise*. These are optimization applications with unknown parameters where the decision maker believes that the optimal value of the nominal problem—the maximization problem they would have solved if the true parameters were known—is unlikely to be large. This belief can originate from various sources, such as a belief that a past policy was not highly suboptimal. For example, in the setting described above, if the store’s best past assortment generated an expected revenue of \$21 per customer, then the parent company may believe it is unlikely that the expected revenue of the optimal assortment is much larger than \$21. A decision maker’s belief about the nominal problem can also derive from other sources such as domain knowledge and intuition.

There are fundamental challenges with trying to incorporate human beliefs into optimization applications. First, a decision maker’s belief may be *incorrect*. For example, in the assortment optimization setting, it may be the case that the store’s best past assortment was in fact highly suboptimal, perhaps because it was chosen by a store manager with an objective other than maximizing revenue, such as maximizing market share. Second, in practice, a decision maker’s belief often comes in the form of a *vibe* rather than a numerical assertion. For example, a parent company may believe that it is unlikely that the store’s best

past assortment was highly suboptimal, but may be hard pressed to transform that belief into a claim that the expected revenue of the store’s best past assortment is guaranteed to be within, say, 31.2% of optimal. Third, even if the decision maker could turn their belief into a numerical bound on the optimal value of the nominal problem, that partial information is generally too crude for identifying the true parameters of the nominal problem and insufficient for recovering the optimal policy.

1.1 Contributions

The objective of this paper is to propose an approach to incorporating human beliefs that contends with the above challenges, and in doing so, to shed light on the value that human expertise can provide in this era of algorithm-driven decision making. The main contributions of this paper are the following.

Model. We formalize optimization with human expertise as follows. A decision maker faces an optimization application defined by a set of feasible policies and an objective function with an unknown parameter. All of the information from available datasets about the true but unknown parameter has been summarized into an uncertainty set of parameters.¹ In addition, the decision maker believes that the optimal value of the nominal problem—the problem they would have solved if the true parameter were known—is unlikely to be large. The decision maker seeks a policy that can be trusted to perform well across the parameters from the uncertainty set, but is guaranteed to perform even better if the optimal value of the nominal problem happens to be small.

We propose a general approach to evaluating and selecting policies in this class of problems. Rather than trying to elicit details about the belief of the decision maker in advance, the approach instead provides a menu of policies to a decision maker, along with a particular object for each policy that we call a *nominal curve*. In a nutshell, a nominal curve is a set of worst-case performance guarantees for a policy that hold under every possible upper bound on the optimal value of the nominal problem, or under every possible suboptimality gap for the best past policy.² By comparing the policies through their nominal curves, the decision maker can obtain an understanding of how the policies in the menu can be trusted to perform under different scenarios of the optimal value of the nominal problem, allowing them to apply their own confidence in their belief and their judgment of what scenarios are plausible to select a policy from the menu to use in practice.

Nominal curves may be viewed as attractive from a practical standpoint for several reasons. First, they address the fact that the decision maker’s belief about the nominal problem may be incorrect by containing worst-case performance guarantees for a policy that hold even if the optimal value of the nominal problem is arbitrarily large. Second, it is shown under mild assumptions that nominal curves are convex, continuous, and nonincreasing (Proposition 1 in §2.4), which makes them relatively simple to visualize, as we show in §§4.3 and 5.3. Third, a nominal curve does not require the decision maker to transform their belief into a fixed upper bound on the optimal value of the nominal problem, which may be inaccurate or difficult to elicit.

Theory. For the nominal curve of a policy to be useful to a decision maker, it must show that the worst-case performance of a policy improves when the optimal value of the nominal problem happens to be small. This raises many practical questions: for example, in what settings do nominal curves contain worst-case performance guarantees that are less conservative than those that would have been obtained in the absence of a belief about the optimal value of the nominal problem? By how much can the worst-case performance guarantees of a policy be improved if the decision maker’s belief about the nominal problem turns out to be correct?

To answer these questions, we introduce a quantity that we call the *value of human expertise* (§2.5). The value of human expertise is the maximum difference between the optimal values of two problems. The first is a robust optimization problem, that is, the problem of selecting a policy that performs best under the worst-case parameters from the uncertainty set. The second is a robust optimization problem where the uncertainty

¹Our setup (§2) is general and can address applications in which policies and parameters are infinite-dimensional.

²For example, in the assortment optimization setting from the beginning of §1, a nominal curve shows the worst-case expected revenue of a new assortment if the expected revenue of the store’s best past assortment happened to be within X% of optimal, for every possible value of X.

set has been augmented with constraints that remove all parameters that, if true, would have led to a nominal problem with an optimal value that exceeds a threshold. As such, the value of human expertise captures the maximum improvement in the optimal value of a robust optimization problem that can be obtained from a belief that the optimal value of the nominal problem is unlikely to be large. Stated alternatively, the value of human expertise is large if and only if there exist nominal curves that provide worst-case performance guarantees that are much less conservative than those that can be obtained by solving a robust optimization problem.

Our main result (Corollary 2 in §3.2) shows that the value of human expertise has a simple characterization. Specifically, we prove that if the problem of computing the worst-case performance of a policy over the uncertainty set is a convex problem, then the value of human expertise is equal to the gap between the optimal values of the min-max and max-min formulations of a robust optimization problem. This result, which holds for infinite-dimensional policy and parameter spaces, shows that a decision maker’s belief about the nominal problem is valuable precisely in the applications where minimax duality does not hold. The proof of our main result follows from a simple—and, as best we can tell, novel—theorem about max-min problems (Theorem 1 in §3.2). The theorem shows that if a max-min problem has an inner problem that is convex, then a pure strategy of the outer problem that performs best against all of the optimal solutions of the min-max problem yields an objective value that is equal to the optimal value of the min-max problem.

Applications. We illustrate our approach in two applications. First, we consider the hyperlocal assortment optimization setting that was described at the beginning of §1, in which a parent company seeks to optimize assortments at local brick-and-mortar stores (§4). We consider a nonparametric setting in which the parent company’s goal is to identify an assortment that performs well across all of the random utility maximization models that are consistent with the transactional sales data generated by the store’s past assortments (Farias et al. 2013, Sturt 2025). In a stylized numerical example, we show that nominal curves make it possible to identify new assortments to recommend to the store that in the worst case do not perform much worse than the store’s best past assortment, but are guaranteed to strictly outperform the store’s best past assortment if the expected revenue of the store’s best past assortment happens to be close to optimal (Figure 1 in §4.3).

Second, we consider combinatorial optimization problems (e.g., shortest path, bipartite matching, traveling salesman problems) with uncertain cost coefficients and budget uncertainty sets (§5). Our analysis of this setting is motivated by situations where decision makers themselves have domain knowledge. For example, experienced emergency dispatchers may have knowledge about traffic delays that commonly occur when there are flash floods; during a severe storm, the dispatcher may not precisely know the location of flooding, but may believe it is likely that all shortest routes will experience delays. For this setting, we establish several theoretical conditions under which relatively loose beliefs about the optimal value of the nominal problem can lead to improved performance guarantees. We highlight these theoretical results through numerical experiments on a class of randomly generated shortest path problems from Bertsimas and Sim (2003).

To facilitate the deployment of our proposed approach in applications, we develop two methods that can be applied in assortment and combinatorial optimization problems for computing a menu of policies that are *pointwise optimal*, meaning that their nominal curves give the best possible performance guarantees in at least one scenario (§§2.3 and 6.1). The first method is a compact mixed-integer programming formulation that applies to zero-one network flow problems (such as shortest path and bipartite matching) and cardinality-constrained assortment optimization under the multinomial logit model with polyhedral uncertainty sets (§6.1.1). The second method is a simple cutting plane method that can be applied in more general settings (§6.1.2). We propose other approaches for generating menus of policies in §6.2.

1.2 Related Literature

This paper draws on and contributes to several fields including robust optimization, inverse optimization, assortment optimization, and minimax theory.

Robust optimization. There is a vast literature on designing uncertainty sets in robust optimization based on probabilistic guarantees, risk aversion, historical data, and machine learning models; see Kuhn et al.

(2025), Lou et al. (2026), and references therein. Most of our main results hold for general uncertainty or ambiguity sets that are nonempty, compact, and convex (see §§2.4 and 3). As such, our main results can be applied concurrently with many existing methods for designing uncertainty and ambiguity sets. Compared to papers such as Iancu and Trichakis (2014) that reduce the conservatism of robust optimization for a given uncertainty set, we propose a new approach to evaluating policies by studying the relationship between the parameters in an uncertainty set and the possible optimal values of the nominal problem.

Our paper relates to literature on reducing the conservatism of uncertainty sets by incorporating information about an auxiliary optimization problem. Long et al. (2023) and Sim et al. (2025) introduce robust satisficing and propose computing the target in that model by solving an empirical optimization problem to contend with the optimizer’s curse. Wang et al. (2023) tune the parameters of an uncertainty set to perform well across a contextual family of different problems. Closer to the present paper is the work of Bennouna et al. (2025, 2026), who add constraints into uncertainty sets based on linear projections of the linear objective function with uncertain coefficients and analyze when these constraints are sufficient for identifying the optimal solution of the nominal problem. In our work, the auxiliary problem is the nominal problem, i.e., the problem we would have solved if the true parameters were known, and we incorporate bounds on the optimal value of the nominal problem by adding constraints into an uncertainty set.

Private information and inverse optimization. An example of an application where humans have access to private information that is not available to the decision maker is hyperlocal revenue management, whereby a parent company or distributor seeks to optimize pricing or assortment decisions at local stores. In these settings, there is literature documenting that managers have access to private information about local consumer preferences that influences their decisions (Kök et al. 2008, Farias et al. 2017), which relates to a much broader literature on the value of private information in human-AI collaboration (Kesavan and Kushwaha 2020, Balakrishnan et al. 2026). Our paper studies how such private information can be incorporated indirectly through beliefs about the quality of past policies that were selected by humans.

A related stream of research is the field of inverse optimization, which focuses on using partial information about the optimal solutions of a nominal problem to infer the parameters of its objective function (Ahuja and Orlin 2001, Chan et al. 2025). Typical motivating examples for inverse optimization are situations where one observes decisions made by experts (such as routing choices of drivers or medical decisions made by doctors) and wishes to “reverse engineer” the objective or utility function that the experts were trying to optimize. Many variants have been proposed, including those where the optimal solutions are data-driven and subject to noise (Aswani et al. 2018, Mohajerin Esfahani et al. 2018) and where information about the optimal value is available (Ahmed and Guan 2005).

Our paper studies a problem setting that lies somewhere between traditional robust optimization and inverse optimization. Similar to robust optimization, we start with an uncertainty set of parameters of the objective function, and like inverse optimization, we additionally have partial information about the nominal problem with the true parameters. However, we are interested in cases where that partial information is too crude for identifying the true parameters of the nominal problem and insufficient for recovering the optimal policy. One of the takeaways of the numerical examples of this paper (§§4.3 and 5.3) is that intersecting a “conservative” uncertainty set with “crude” information about the optimal value of the nominal problem can lead to robust optimization problems that are practically useful.

Assortment optimization. A central problem in the field of assortment optimization is identifying the discrete choice model from aggregated transactional data generated by a store’s past assortments. The challenge is that there are often many stochastically rational (i.e., random utility maximization) choice models that can fit such data perfectly. To contend with this, a stream of literature (Rusmevichientong and Topaloglu 2012, Farias et al. 2013, Bertsimas and Mišić 2017, Jin et al. 2022, Désir et al. 2024, Wang et al. 2024, Sturt 2025, Ruan et al. 2026) has focused on constructing uncertainty sets of discrete choice models and finding an assortment that does best in the worst case by solving a robust assortment optimization problem. This work fits within a broader literature in revenue management and economics on prior-free and non-Bayesian approaches to pricing and mechanism design; see Bergemann and Schlag (2011), Koçyiğit et al. (2020), Anunrojwong et al. (2024), Bahamou et al. (2024), and references therein.

Our paper contributes to the above revenue management literature in three ways. First, we propose a new robustness criterion that is designed for settings where a decision maker believes that the optimal value of the nominal problem is unlikely to be large. Second, we provide numerical evidence that uncertainty sets that are constructed using nonparametric modeling techniques from [Farias et al. \(2013\)](#) can contain random utility maximization models that, if true, would imply that past assortments were highly suboptimal (see §4). Third, we show that eliminating those parameters can induce a phase transition of a robust assortment optimization problem from being overly conservative to practically useful.

Minimax theory. There is a rich history in operations research and game theory of establishing properties of max-min problems that ensure that the minimum and maximum can be interchanged. Classic conditions for minimax theorems include [Fan \(1953\)](#) and [Sion \(1958\)](#), and conditions that arise in the context of robust optimization applications can be found in [Nilim and El Ghaoui \(2005\)](#), [Iyengar \(2005\)](#), [Wei and Zhang \(2024\)](#), [Désir et al. \(2024\)](#), [Zhen et al. \(2025\)](#), [Shafiee et al. \(2026\)](#). Our paper differs from the above literature because we prove our main results under assumptions that are insufficient for minimax theorems (Assumptions 1 and 2 in §2.4). The study of mixed strategies in the context of stochastic programming and robust optimization includes [Delage et al. \(2019\)](#), [Wang et al. \(2024\)](#), [Guan and Mišić \(2026\)](#). Our work focuses on high-stakes applications where a decision maker is risk-averse and randomized policies are impossible to implement or undesirable to the decision maker; examples of such applications can be found throughout this paper.

As best we can tell, the closest related work to our Theorem 1 is [Terkelsen \(1972, p. 406\)](#), which presents proof techniques based on finite intersections that are similar to ours, specifically our proof steps that we relegate to Appendix A.2. While our proof techniques for Theorem 1 are elementary, we are not aware of a prior statement of our theorem in the literature, and our theorem does not appear to follow immediately from classical results from papers such as [Fan \(1953\)](#) and [Sion \(1958\)](#). The value of our Theorem 1 lies in its generality: it does not require convexity of the outer problem, does not assume that sets of optimal strategies for the inner or outer problems are singletons, and extends to inner and outer problems that are infinite-dimensional.

2 Optimization with Human Expertise

2.1 Problem Setting

We consider optimization problems of the form

$$\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \bar{\boldsymbol{\theta}}) \tag{1}$$

where the policies are chosen from a nonempty compact set $\mathcal{X}$ and where the true parameter of the objective function $\bar{\boldsymbol{\theta}}$ is unknown. To ensure that optimums are attained, we assume throughout that $f(\cdot, \cdot)$ is bounded, upper semicontinuous in its first argument, and lower semicontinuous in its second argument.³ The parameter $\bar{\boldsymbol{\theta}}$ can correspond, for example, to a probability measure, a random utility model, or edge costs in a network.

We focus on the problem setting in which all of the information from available datasets about the true but unknown parameter $\bar{\boldsymbol{\theta}}$ has been summarized by a nonempty compact set of parameters $\mathcal{U}$, referred to as an uncertainty set. We assume that the uncertainty set was constructed with the goal of being small while containing the true parameter $\bar{\boldsymbol{\theta}} \in \mathcal{U}$. The study of data-driven techniques for constructing uncertainty sets with rigorous guarantees has rapidly evolved into what is now a relatively mature discipline (see §1.2). In assortment optimization, for example, uncertainty sets can be constructed as a set of random utility maximization models that are consistent with transactional sales data generated by a store’s past assortments.

In this paper, we assume that in addition to having an uncertainty set, the decision maker believes that it is unlikely that the optimal value of the nominal problem (1) is large. The belief is derived from information that is not captured in datasets, obtained from domain knowledge and interacting with the physical world.

³Note that these assumptions allow for the policies $\mathbf{x}$ and the parameter $\bar{\boldsymbol{\theta}}$ to be infinite-dimensional.

The decision maker seeks a policy that can be trusted to perform well across the parameters from the uncertainty set, but performs even better if the optimal value of the nominal problem (1) happens to be small.

2.2 A New Approach to Evaluating Policies

We propose the following approach to evaluating policies in the above problem setting. For each $\eta \in \mathbb{R}$, define the *reduced uncertainty set* corresponding to η as

$$\mathcal{U}_\eta := \left\{ \boldsymbol{\theta} \in \mathcal{U} : \max_{\mathbf{x}' \in \mathcal{X}} f(\mathbf{x}', \boldsymbol{\theta}) \leq \eta \right\} = \{ \boldsymbol{\theta} \in \mathcal{U} : f(\mathbf{x}', \boldsymbol{\theta}) \leq \eta \ \forall \mathbf{x}' \in \mathcal{X} \}.$$

In words, the reduced uncertainty set is equal to the original uncertainty set with constraints that exclude all parameters that, if true, would lead the nominal problem to have an optimal value that exceeds η . Instead of evaluating a policy $\mathbf{x}$ by its worst-case performance, we propose evaluating the policy by its *nominal curve*

$$\left\{ \left(\eta, \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) \right) : \eta \geq \underline{\eta} \right\} \quad (2)$$

where $\underline{\eta} := \min\{\eta : \mathcal{U}_\eta \neq \emptyset\}$ is the smallest scalar for which the reduced uncertainty set is nonempty.⁴ The nominal curve provides the usual worst-case performance guarantee for the policy for all sufficiently large η and gives performance guarantees that may be less pessimistic if $\boldsymbol{\theta} \in \mathcal{U}$ and the optimal value of the nominal problem (1) happens to be less than or equal to η , for each $\eta \geq \underline{\eta}$. As we show below, the nominal curve thus offers the decision maker a rigorous way to transform a belief that it is unlikely that the optimal value of the nominal problem (1) is large into performance guarantees for a policy that are practically useful.

As a motivating example, it is common in assortment optimization for a parent company to have access to historical sales data generated by a store's past assortments, and for the store manager to be risk-averse and reluctant to experiment with new assortments that might lead to a decline in expected revenue. Given the historical sales data, the parent company can construct an uncertainty set of random utility maximization models that are consistent with the historical sales data, and then solve a robust optimization problem to identify an assortment that can be trusted to outperform the store's best past assortment across all random utility models in the uncertainty set (Farias et al. 2013, Sturt 2025). In §4.3, we show that such data-driven uncertainty sets can contain random utility maximization models that, if true, would imply that the store's past assortments were highly suboptimal. This would be surprising in many practical settings in which store managers did not choose the past assortments *randomly*, but rather were informed by private information about the types of customers that typically visit the store. The nominal curve (2) makes it possible to, for example, identify new assortments to recommend to the store that in the worst case do not perform much worse than the store's best past assortment, but are guaranteed to strictly outperform the store's best past assortment if the expected revenue of the store's best past assortment happens to be close to optimal.

The nominal curve (2) is not a scalar. This is motivated by the fact that the decision maker's belief that it is unlikely that the optimal value of (1) is large may be subjective, and so it may be difficult for a decision maker to transform their belief into, for example, an accurate upper bound on the optimal value of (1). A nominal curve thus allows a decision maker to obtain performance guarantees for a policy under a variety of possible scenarios about the optimal value of the nominal problem (1) that the decision maker believes may be plausible, and to compare policies under those possible scenarios by plotting their nominal curves; see §4.3 and §5.3. By capturing the worst-case performance of a policy over the entire uncertainty set in the case of sufficiently large η ⁵, the nominal curve also shows the decision maker how a policy can be trusted to perform even if their belief is incorrect.

⁴An explicit formula for $\underline{\eta}$ is found in Proposition 2 in §3.1.

⁵It follows from the boundedness of $f(\cdot, \cdot)$ that $\mathcal{U}_\eta = \mathcal{U}$ for all sufficiently large η .

2.3 Generating Policies

Because (2) is not a scalar, there may not exist a policy that is *universally optimal* with respect to (2). Specifically, we observe that the nominal curves (2) corresponding to feasible policies are pointwise upper bounded by

$$\left\{ \left(\eta, \max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) \right) : \eta \geq \underline{\eta} \right\}. \quad (3)$$

If there exists a policy whose nominal curve (2) coincides with the upper bound (3) for every η , then that policy would be considered universally optimal. Because such a policy may not exist, one can generate a menu of policies to offer to the decision maker.

A simple approach to generating a menu of policies is to solve the *reduced robust optimization* problem

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) \quad (4)$$

across a discrete range of values of $\eta \in [\underline{\eta}, \infty)$. The optimal policies for those reduced robust optimization problems have nominal curves that intersect the upper bound (3); we refer to these as *pointwise optimal* nominal curves. Given the menu of policies obtained by solving (4) for different values of η , the decision maker can compare the nominal curves (2) of those policies and apply their own judgment about what scenarios are plausible to select a single policy. We apply this approach to generating menus of policies in the numerical examples in §§4.3 and 5.3, and alternative approaches to generating a menu of policies can be found in §6.2.

2.4 Computation

There are two relevant computational tasks to consider: computing a nominal curve (2) for a fixed policy, and finding a menu of policies through solving (4) for a range of values of η . Neither task is trivial in general in light of the minimal assumptions stated at the beginning of §2.1, which are repeated below:

Assumption 1. $\mathcal{X}, \mathcal{U}$ are compact, nonempty sets, and $f(\cdot, \cdot)$ is a bounded function that is upper semicontinuous in its first argument and lower semicontinuous in its second argument.

Those two computational tasks simplify, however, if the following assumption is also satisfied:

Assumption 2. $\mathcal{U}$ is a convex set and $\boldsymbol{\theta} \mapsto f(\mathbf{x}, \boldsymbol{\theta})$ is a convex function for all $\mathbf{x} \in \mathcal{X}$.

The above assumption, which is common in the literature and often satisfied in real-world applications, does not guarantee that nominal curves (2) are easy to compute, nor does it imply that the reduced robust optimization problem (4) is easy to solve. For example, Assumptions 1 and 2 do not preclude $\mathcal{U}$ from being an infinite-dimensional set of parameters. That said, Assumptions 1 and 2 ensure for each $\mathbf{x} \in \mathcal{X}$ and $\eta \geq \underline{\eta}$ that

$$\min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta})$$

is a convex optimization problem over a compact convex set $\mathcal{U}_\eta$. As such, the nominal curve (2) for every fixed $\mathbf{x} \in \mathcal{X}$ can be calculated to any accuracy by solving multiple convex optimization problems, one for each η in a sufficiently fine grid. The above assumptions also ensure that nominal curves (2) have a simple structure:

Proposition 1. *If Assumptions 1 and 2 hold and $\mathbf{x} \in \mathcal{X}$, then the function $v_{\mathbf{x}}(\eta) := \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta})$ is nonincreasing, convex, and continuous for $\eta \in [\underline{\eta}, \infty)$.*

The above proposition implies that nominal curves (2) for fixed policies as well as the upper bound curve (3) have a structure that is relatively easy to visualize, as we show in §§4.3 and 5.3. Proposition 1 also implies that if a policy $\mathbf{x}$ satisfies $\min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) > \min_{\boldsymbol{\theta} \in \mathcal{U}} f(\mathbf{x}, \boldsymbol{\theta})$ for some η , and if $\bar{\boldsymbol{\theta}} \in \mathcal{U}$, then the smaller the optimal value of the nominal problem (1) happens to be, the better the policy is guaranteed to perform in the worst case. The proof of Proposition 1, which is found in Appendix B, makes use of the closed-form expression of $\underline{\eta} := \min\{\eta : \mathcal{U}_\eta \neq \emptyset\}$ that is established in Proposition 2 in §3.1.

2.5 The Value of Human Expertise

To evaluate the capacity of nominal curves to provide performance guarantees that are practically useful, we introduce a quantity that we call the *value of human expertise*. This quantity, defined below, is the maximum improvement in the optimal value of a robust optimization problem that can be achieved by eliminating parameters from the uncertainty set that, if true, would have led the nominal problem to have a large optimal value. Equivalently, it is the difference between the maximum and minimum of the upper bound curve (3):

$$\begin{aligned}\Delta &:= \max_{\eta \geq \underline{\eta}} \left\{ \max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) \right\} - \min_{\eta \geq \underline{\eta}} \left\{ \max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) \right\} \\ &= \max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) - \max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}} f(\mathbf{x}, \boldsymbol{\theta})\end{aligned}$$

The second equality holds because $\mathcal{U}_\eta$ is nonempty, $\mathcal{U}_\eta \subseteq \mathcal{U}_\eta$ for all $\eta \geq \underline{\eta}$, and $\mathcal{U}_\eta = \mathcal{U}$ for all sufficiently large η by the boundedness of $f(\cdot, \cdot)$. Note that the attainment of the maxima and minima in each of the above optimization problems follows from the assumptions at the beginning of §2.1, which were restated as Assumption 1 in §2.4.

The value of human expertise can be interpreted as the maximum improvement in performance guarantees that can be obtained from a belief that the optimal value of (1) is unlikely to be large. Indeed, if this quantity is equal to zero, then it is not possible for nominal curves to provide performance guarantees that are less conservative than those that would be obtained by solving a robust optimization problem. If the value of human expertise is large, then it guarantees the existence of policies with non-trivial nominal curves, that is, the existence of policies $\mathbf{x} \in \mathcal{X}$ for which $\min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta})$ is much larger than $\min_{\boldsymbol{\theta} \in \mathcal{U}} f(\mathbf{x}, \boldsymbol{\theta})$ as well as the optimal value of a robust optimization problem for some values of η . It follows from §2.3 that such policies can be obtained by solving the reduced robust optimization problem (4) for small values of η . In sum, a large Δ is a necessary and sufficient condition for there to exist nominal curves that offer much stronger performance guarantees than can be obtained by robust optimization.

There are many questions related to the value of human expertise that are relevant from theoretical and practical perspectives. In what applications can the value of human expertise be strictly positive? How does the value of human expertise relate to the structure of a traditional robust optimization problem $\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}} f(\mathbf{x}, \boldsymbol{\theta})$? More generally, for there to be a non-zero gap between the optimal values of the reduced robust optimization problem (4) and a traditional robust optimization problem, must η be a tight bound on the optimal value of the nominal problem (1)? In addition to shedding light on those questions, our main goal of this paper is to provide answers to the following two questions:

Question 1. What is the maximum possible value of Δ ?

Question 2. In what settings does Δ attain that maximum possible value?

3 The Value of Human Expertise and the Minimax Gap

In this section, we show that there are simple answers to Questions 1 and 2. In §3.1, we answer Question 1 by developing an upper bound on the value of human expertise that holds under Assumption 1. In §3.2, we present the main result of this paper, Corollary 2, which answers Question 2 by proving that the value of human expertise is always equal to the upper bound from §3.1 if Assumption 2 is also satisfied.

3.1 Simple Bounds

We begin by establishing an upper bound on Δ and other simple preliminary results that hold under Assumption 1. These results will be obtained by relating the following three problems:

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) \quad (4)$$

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}} f(\mathbf{x}, \boldsymbol{\theta}) \quad (5)$$

$$\min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) \quad (6)$$

The first problem is a restatement of the reduced robust optimization problem (4) from §2. The second problem (5) is a traditional robust optimization problem. The third problem (6) is obtained from (5) by interchanging the maximum and minimum.

It follows from Assumption 1 that the optimums in the outer and inner problems of the above three optimization problems are always attained. Moreover, it follows from the standard minimax inequality that the optimal value of the min-max problem (6) is an upper bound on the optimal value of the max-min problem (5). Note that Assumptions 1 and 2 are insufficient for minimax theorems and do not ensure the existence of saddle points for (5); for example, results like Sion's minimax theorem that ensure the equality of the optimal values of (5) and (6) hold only under additional assumptions such as convexity of $\mathcal{X}$ and quasi-concavity of $f(\cdot, \boldsymbol{\theta})$.

In what follows, we develop an upper bound on Δ by showing that the optimal value of the min-max problem (6) is also an upper bound on the optimal value of the reduced robust optimization problem (4). To establish that upper bound, we first characterize the viable choices for the scalar η in the reduced robust optimization problem (4), that is, the values of the scalar for which the reduced uncertainty set is nonempty. In the following proposition, we formalize the viable choices for η by relating the requirement that $\mathcal{U}_\eta \neq \emptyset$ to the optimal value of the min-max problem (6).

Proposition 2. *If Assumption 1 holds, then $\mathcal{U}_\eta \neq \emptyset$ if and only if $\eta \geq \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta})$.*

Proof. Suppose that $\eta \geq \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta})$. Then

$$\mathcal{U}_\eta \supseteq \left\{ \boldsymbol{\theta} \in \mathcal{U} : \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) \leq \min_{\hat{\boldsymbol{\theta}} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \right\} = \arg \min_{\hat{\boldsymbol{\theta}} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \neq \emptyset,$$

where the set inclusion follows from the definition of $\mathcal{U}_\eta$ and the supposition on η , the equality follows from algebra, and the nonemptiness follows from the fact that the optimum of (6) is attained by Assumption 1. The converse follows from similar reasoning. $\square$

In view of the above, the following Proposition 3 establishes that the optimal value of (6) is an upper bound on the optimal value of the reduced robust optimization problem (4). It is followed by Corollary 1, which follows immediately from Proposition 3 and the definition of Δ .

Proposition 3. *If Assumption 1 holds and $\eta \geq \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta})$, then $\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) \leq \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta})$.*

Proof. It follows from Proposition 2 and the fact that $\eta \geq \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta})$ that $\mathcal{U}_\eta \neq \emptyset$. Therefore,

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta}) \leq \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) = \left[\begin{array}{ll} \min_{\boldsymbol{\theta} \in \mathcal{U}} & \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) \\ \text{s.t.} & \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) \leq \eta \end{array} \right] = \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}),$$

where the inequality is the minimax inequality, the first equality follows from the definition of $\mathcal{U}_\eta$, and the second equality follows from the fact that the constraint $\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) \leq \eta$ can be removed without affecting the outer minimization problem. $\square$

Corollary 1. *If Assumption 1 holds, then $\Delta \leq \min_{\theta \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta) - \max_{\mathbf{x} \in \mathcal{X}} \min_{\theta \in \mathcal{U}} f(\mathbf{x}, \theta)$.*

The above proposition and corollary show that the value of human expertise is linked to the gap between the min-max problem (6) and the max-min problem (5). Specifically, Proposition 3 shows that for the optimal value of the reduced robust optimization problem (4) to be greater than the optimal value of the robust optimization problem (5), it is *necessary* for there to be a gap between the optimal values of the max-min problem (5) and the min-max problem (6). Corollary 1 shows that the value of human expertise is at most the gap between the optimal values of (6) and (5).

Corollary 1 is useful because it furnishes us with a simple test for determining whether it is *not* worthwhile to consider nominal curves (2). Indeed, Corollary 1 implies that the worst-case performance $\min_{\theta \in \mathcal{U}_\eta} f(\mathbf{x}, \theta)$ for a policy $\mathbf{x}$ cannot be strictly greater than the optimal value of the robust optimization problem (5) in settings where the assumptions for minimax theorems are satisfied, such as problems where $\mathcal{X}, \mathcal{U}$ are convex sets and $f(\cdot, \cdot)$ is concave in its first argument and convex in its second argument. Corollary 1 also shows that the worst-case performance $\min_{\theta \in \mathcal{U}_\eta} f(\mathbf{x}, \theta)$ for a policy $\mathbf{x}$ cannot be much greater than the optimal value of (5) if the gap between the optimal values of (5) and (6) is small. On the positive side, Corollary 1 leaves open the possibility that the value of human expertise can be large in settings where pure strategies for the robust optimization problem (5) are highly suboptimal, i.e., settings where the gap between the optimal values of (5) and (6) is large.

3.2 Main Result

Corollary 1 provides a simple upper bound on the value of human expertise which holds whenever Assumption 1 is satisfied. In particular, Corollary 1 shows that the value of human expertise has the potential to be strictly positive—equivalently, the optimal value of the reduced robust optimization problem (4) can be larger than the optimal value of the robust optimization problem (5)—when the gap between the optimal values of the max-min problem (5) and the min-max problem (6) is large. But is that upper bound on Δ always attainable, or is the best-case value of human expertise less in practice?

In this subsection, we answer that question by showing that the upper bound from Corollary 1 is tight and always attained under Assumption 2. Our main result is a consequence of the following theorem, which states a property about optimal solutions for max-min problems.

Theorem 1. *If Assumptions 1 and 2 hold and $\eta = \min_{\theta \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$, then $\max_{\mathbf{x} \in \mathcal{X}} \min_{\theta \in \mathcal{U}_\eta} f(\mathbf{x}, \theta) = \min_{\theta \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$.*

The above theorem establishes that a policy that performs best against all of the optimal solutions of the min-max problem (6) yields an optimal value that is equal to the optimal value of the min-max problem (6). The value of Theorem 1 lies in its generality: it does not require convexity in $\mathcal{X}$ or quasi-concavity of $f(\cdot, \theta)$, does not assume that $\mathcal{U}^*$ is a singleton, and does not require that the sets $\mathcal{X}$ and $\mathcal{U}$ are finite-dimensional. Theorem 1 thus extends to a wide range of applications, such as settings where the parameter is a probability measure or the policies are infinite-dimensional in the context of dynamic optimization. The following proof of Theorem 1 is simple and based on elementary facts about topological spaces and convexity.

Proof of Theorem 1. Let $\eta = \min_{\theta \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$ and $\mathcal{U}^* := \arg \min_{\theta \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$. It follows from Assumption 2 that $\mathcal{U}^*$ is convex, and it follows from algebra that $\mathcal{U}^* = \{\theta \in \mathcal{U} : \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta) \leq \eta\} = \mathcal{U}_\eta$. We have two cases to consider:

- Case 1: Suppose that there exists an $\hat{\mathbf{x}} \in \mathcal{X}$ that satisfies $f(\hat{\mathbf{x}}, \theta) = \eta$ for all $\theta \in \mathcal{U}^*$. Then

$$\min_{\theta \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta) = \min_{\theta \in \mathcal{U}^*} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta) = \min_{\theta \in \mathcal{U}^*} f(\hat{\mathbf{x}}, \theta) \leq \max_{\mathbf{x} \in \mathcal{X}} \min_{\theta \in \mathcal{U}^*} f(\mathbf{x}, \theta).$$

Combining the above inequality with the fact that $\mathcal{U}^* = \mathcal{U}_\eta$ and Proposition 3, we have

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\theta \in \mathcal{U}^*} f(\mathbf{x}, \theta) = \min_{\theta \in \mathcal{U}^*} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta).$$

- **Case 2:** Suppose that there does not exist an $\hat{\mathbf{x}} \in \mathcal{X}$ that satisfies $f(\hat{\mathbf{x}}, \boldsymbol{\theta}) = \eta$ for all $\boldsymbol{\theta} \in \mathcal{U}^*$. Then it follows from Assumption 1 and from a direct application of basic results about topological spaces (see Appendix A.2) that there must exist a finite subset $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K\} \subseteq \mathcal{U}_\eta$ that satisfies

$$\bigcap_{k=1}^K \{\mathbf{x} \in \mathcal{X} : f(\mathbf{x}, \boldsymbol{\theta}_k) = \eta\} = \emptyset. \quad (7)$$

Since $\mathcal{U}_\eta$ is convex, it must be the case that $\hat{\boldsymbol{\theta}} := \frac{1}{K} \sum_{k=1}^K \boldsymbol{\theta}_k$ is an element of $\mathcal{U}^*$. Moreover, since $\hat{\boldsymbol{\theta}} \in \mathcal{U}^*$, it must be the case that there exists $\hat{\mathbf{x}} \in \mathcal{X}$ that satisfies $f(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \eta$. However, we observe that

$$\eta = f(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = f\left(\hat{\mathbf{x}}, \frac{1}{K} \sum_{k=1}^K \boldsymbol{\theta}_k\right) \leq \frac{1}{K} \sum_{k=1}^K f(\hat{\mathbf{x}}, \boldsymbol{\theta}_k) < \frac{1}{K} \sum_{k=1}^K \eta = \eta,$$

where the first inequality follows from Assumption 2, and the strict inequality follows from line (7) (which implies that there exists $k \in \{1, \dots, K\}$ such that $f(\hat{\mathbf{x}}, \boldsymbol{\theta}_k) < \eta$) and from the fact that $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K \in \mathcal{U}^*$ (which implies for each $k \in \{1, \dots, K\}$ and for all $\mathbf{x} \in \mathcal{X}$ that $f(\mathbf{x}, \boldsymbol{\theta}_k) \leq \eta$). We thus have a contradiction, which implies we cannot be in Case 2.

Because those cases are exhaustive, our proof of Theorem 1 is complete. $\square$

Combining the above theorem with Proposition 2 and Corollary 1, we obtain the main result of this paper:

Corollary 2. *If Assumptions 1 and 2 hold, then $\Delta = \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) - \max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}} f(\mathbf{x}, \boldsymbol{\theta})$.*

The above corollary shows under Assumptions 1 and 2 that the upper bound from Proposition 3 is always attained. That is, it shows that if η is the smallest possible scalar for which the reduced uncertainty set is nonempty (Proposition 2), then the gap between the optimal values of the reduced robust optimization problem (4) and the robust optimization problem (5) is equal to the gap between the optimal values of (6) and (5).

Corollary 2 has a number of practical implications. It shows affirmatively that if there is a gap between the optimal values of (5) and (6), and if Assumptions 1 and 2 are satisfied, then there exist policies with worst-case performance guarantees that are strictly greater than the optimal value of (5) if $\bar{\boldsymbol{\theta}} \in \mathcal{U}$ and the optimal value of (1) is small. In other words, Corollary 2 provides an exact characterization of the problems for which nominal curves (2) can be practically informative. More generally, Theorem 1 shows that incorporating an upper bound on the optimal value of the nominal problem (1) into the uncertainty set of a robust optimization problem (5) can induce an interpolation between optimal pure strategies and mixed strategies. Specifically, Theorem 1 shows that incorporating an upper bound on (1) modifies an uncertainty set in such a way that can make the optimal pure strategy for the reduced robust optimization problem (4) equal in optimal value to the optimal mixed strategy for the original robust optimization problem (5). This interpretation is demonstrated through the following toy example.

Example 1. Let $\mathcal{X} = \{-1, 1\}$, $\mathcal{U} = [-1, 1]$, and $f(x, \theta) = \theta x$. Then the optimal value over pure strategies and the optimal pure strategies are

$$\max_{x \in \mathcal{X}} \min_{\theta \in \mathcal{U}} \theta x = -1 \text{ and } \arg \max_{x \in \mathcal{X}} \min_{\theta \in \mathcal{U}} \theta x = \{-1, 1\}.$$

Letting $\mathcal{P}(\mathcal{X})$ denote the set of mixed strategies for the outer problem, and letting δ_x denote the Dirac delta measure at x , we observe that the optimal value over mixed strategies and the optimal mixed strategy are

$$\max_{\mu \in \mathcal{P}(\mathcal{X})} \min_{\theta \in \mathcal{U}} \mathbb{E}_{X \sim \mu}[\theta X] = \min_{\theta \in \mathcal{U}} \max_{x \in \mathcal{X}} \theta x = 0 \text{ and } \arg \max_{\mu \in \mathcal{P}(\mathcal{X})} \min_{\theta \in \mathcal{U}} \mathbb{E}_{X \sim \mu}[\theta X] = \{1/2\delta_{-1} + 1/2\delta_1\}.$$

It follows from Proposition 2 that $\mathcal{U}_\eta \neq \emptyset$ if and only if $\eta \in [\min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{x \in \mathcal{X}} \theta x, \infty) = [0, \infty)$, and for each

such η the reduced uncertainty set is

$$\mathcal{U}_\eta = \left\{ \theta \in [-1, 1] : \max_{x \in \{-1, 1\}} \theta x \leq \eta \right\} = \begin{cases} [-1, 1], & \text{if } \eta \geq 1, \\ [-\eta, \eta], & \text{if } 0 \leq \eta < 1. \end{cases}$$

Therefore, for all $\eta \in [0, \infty)$, the optimal value over pure strategies and the optimal pure strategies for the reduced robust optimization problem are

$$\max_{x \in \mathcal{X}} \min_{\theta \in \mathcal{U}_\eta} \theta x = \begin{cases} -1, & \text{if } \eta \geq 1, \\ -\eta, & \text{if } 0 \leq \eta < 1 \end{cases} \quad \text{and} \quad \arg \max_{x \in \mathcal{X}} \min_{\theta \in \mathcal{U}_\eta} \theta x = \{-1, 1\}.$$

Thus, in the extreme case where $\eta = 0$, the optimal pure strategy for the reduced robust optimization problem is equal in optimal value to the optimal mixed strategy for the original robust optimization problem. $\square$

In Example 1, the set of optimal solutions for the reduced robust optimization problem is the same for all η , in the sense that $\arg \max_{x \in \mathcal{X}} \min_{\theta \in \mathcal{U}_\eta} \theta x = \{-1, 1\}$ for all $\eta \in [0, \infty)$. However, it is possible for the set of optimal solutions for (4) to change with η ; see Figure 1 in §4.3. The following Example 2 shows that Corollary 2 can fail to hold if Assumption 2 is violated.

Example 2. Let $\mathcal{X} = \mathcal{U} = \{-1, 1\}$ and $f(x, \theta) = \theta x$. Then

$$\max_{x \in \mathcal{X}} \min_{\theta \in \mathcal{U}} \theta x = -1 \quad \text{and} \quad \min_{\theta \in \mathcal{U}} \max_{x \in \mathcal{X}} \theta x = 1.$$

Assumption 2 is violated because $\mathcal{U}$ is non-convex, and we observe for all $\eta \geq \min_{\theta \in \mathcal{U}} \max_{x \in \mathcal{X}} \theta x = 1$ that

$$\max_{x \in \mathcal{X}} \min_{\theta \in \mathcal{U}_\eta} \theta x = \max_{x \in \{-1, 1\}} \min_{\theta \in \{-1, 1\} : \max_{\hat{x} \in \{-1, 1\}} \theta \hat{x} \leq \eta} \theta x = \max_{x \in \{-1, 1\}} \min_{\theta \in \{-1, 1\}} \theta x = -1,$$

which implies that the equality from Corollary 2 is not satisfied. $\square$

In summary, Corollary 2 gives answers to the questions of *if* and *by how much* a belief that the optimal value of the nominal problem (1) is unlikely to be large can lead to performance guarantees that are less conservative than those obtained by robust optimization. In the following sections, we apply Corollary 2 to practical settings to analyze the value of human expertise and shed light on the usefulness of nominal curves.

4 Identifiability in Data-Driven Assortment Optimization

In §4, we consider a parent company that seeks to identify a new assortment to recommend to a local store using the transactional sales data generated by the store's past assortments. We present an example in which it is impossible for the parent company to identify an assortment that is guaranteed to outperform the store's best past assortment across all of the random utility maximization models that are consistent with the data generated by the store's past assortments. We then show for the example that nominal curves (2) make it possible to identify assortments that in the worst case do not perform much worse than the store's best past assortment, but are guaranteed to strictly outperform the store's best past assortment if the expected revenue of the store's best past assortment happens to be close to optimal.

4.1 Problem Setup

Assortment optimization is a class of problems from revenue management in which the goal is to select a subset of products for a store to offer to its customers that maximizes expected revenue. Let $1, \dots, n$ denote the products that the store may offer, and the revenues of the products are given by $r_1, \dots, r_n > 0$. Let the no-purchase option be denoted by index 0 with $r_0 = 0$, and let $\mathcal{S} \equiv \{S \subseteq \{0, \dots, n\} : 0 \in S\}$ denote the set of all assortments. Given $S \in \mathcal{S}$ and $i \in S$, let $\bar{\mathbb{P}}(i|S) \in [0, 1]$ denote the proportion of the store's customers

that purchase product i when offered assortment S . If the store's true discrete choice model $\bar{\mathbb{P}}$ were known, then an assortment that maximizes the store's expected revenue would be obtained by solving

$$\underset{S \in \mathcal{S}}{\text{maximize}} \quad \sum_{i \in S} r_i \bar{\mathbb{P}}(i | S) \quad (8)$$

We take the role of a parent company that wishes to find a new assortment to recommend to the local brick-and-mortar store to offer to its customers. The true discrete choice model $\bar{\mathbb{P}}$ that captures the local customer preferences at the store is unknown, and the parent company only has historical sales data from the past assortments that were chosen by the store. Moreover, the store manager is risk-averse and reluctant to experiment with new assortments that might lead to a decline in expected revenue. The goal of the parent company is to identify a new assortment that can be trusted to outperform the store's best past assortment across all of the stochastically rational (that is, random utility maximization) discrete choice models that are consistent with the historical sales data generated by the store's past assortments.

The local store in our example has four products that it may offer, denoted by 1, 2, 3, 4, and the per-unit revenues of the products are

$$r_1 = \$2, \quad r_2 = \$10, \quad r_3 = \$31, \quad r_4 = \$40.$$

We assume that the true discrete choice model $\bar{\mathbb{P}}$ of the local store is unknown, but the parent company has information about the true discrete choice model from aggregated transactional sales data generated from the past assortments offered by the store to its customers. Specifically, the store has previously offered two different assortments,

$$S_1 = \{0, 2, 3\}, \quad S_2 = \{0, 1, 3, 4\},$$

and from these assortments the parent company observed from the transactional sales data that

$$\begin{aligned} \bar{\mathbb{P}}(0|S_1) &= 0, & \bar{\mathbb{P}}(2|S_1) &= \bar{\mathbb{P}}(3|S_1) = 1/2, \\ \bar{\mathbb{P}}(0|S_2) &= \bar{\mathbb{P}}(3|S_2) = 0, & \bar{\mathbb{P}}(1|S_2) &= \bar{\mathbb{P}}(4|S_2) = 1/2. \end{aligned}$$

We assume that the two past assortments were offered by the store manager for a sufficiently long time such that there is little statistical uncertainty in these estimates. As is typical in the revenue management literature, we also assume that $\bar{\mathbb{P}}$ is a random utility maximization (RUM) model, which is a general class of discrete choice models that is consistent with stochastic rationality (Block and Marschak 1959) and subsumes most parametric families of discrete choice models studied in revenue management. Define the uncertainty set of all RUM models that are consistent with the transactional sales data generated by the past assortments

$$\mathcal{U} \triangleq \left\{ \mathbb{P} \in \mathcal{P} : \begin{array}{l} \mathbb{P}(0|S_1) = 0, \quad \mathbb{P}(2|S_1) = \mathbb{P}(3|S_1) = 1/2 \\ \mathbb{P}(0|S_2) = \mathbb{P}(3|S_2) = 0, \quad \mathbb{P}(1|S_2) = \mathbb{P}(4|S_2) = 1/2 \end{array} \right\}, \quad (9)$$

where $\mathcal{P}$ denotes the set of all RUM models over the universe of products. The uncertainty set of the form (9), which was introduced by Farias et al. (2013), can be viewed as very general, as it is guaranteed to contain the true discrete choice model $\bar{\mathbb{P}}$ if there is no noise in the transactional sales data and if $\bar{\mathbb{P}}$ is a RUM model.⁶

In view of the above, the identification problem faced by the parent company can be formally stated as follows. First, we observe that the expected revenues of the store's past assortments are

$$\sum_{i \in S_1} r_i \bar{\mathbb{P}}(i|S_1) = r_2 \frac{1}{2} + r_3 \frac{1}{2} = \$20.5, \quad \sum_{i \in S_2} r_i \bar{\mathbb{P}}(i|S_2) = r_1 \frac{1}{2} + r_4 \frac{1}{2} = \$21.$$

The problem of identifying a new assortment that is guaranteed to outperform the store's best past assortment across all of the RUM models that are consistent with the transactional sales data generated by the store's

⁶Both of these assumptions can be relaxed by modifying the set from line (9) to deviate from the historical choice probabilities by at most a given radius; see, for example, Farias et al. (2013, §§3.3 and 5.1.3).

past assortments thus can be cast as the problem of finding an assortment $S \in \mathcal{S}$ that satisfies

$$\sum_{i \in S} r_i \mathbb{P}(i|S) > \$21 \quad \forall \mathbb{P} \in \mathcal{U} \quad (10)$$

where $\$21 = \max\{\$20.5, \$21\}$ is the expected revenue of the store's best past assortment.

4.2 The Identification Problem Has No Solution

There are examples in which there exist assortments that outperform the store's best past assortment across all RUM models that are consistent with transactional sales data generated by the past assortments (Sturt 2025, §3). However, the example from §4.1 is not one of them: we show in §4.3 and Appendix D for each assortment $S \in \mathcal{S}$ that

$$\min_{\mathbb{P} \in \mathcal{U}} \sum_{i \in S} r_i \mathbb{P}(i|S) \leq \$21.$$

As such, there is no solution to the identification problem in §4.1, in the sense that there does not exist an assortment that satisfies (10). This implies that there is no new assortment that the parent company can suggest to the store that can be trusted in the worst case to increase the store's expected revenue.

What should the parent company do in situations like that described above? The most common approach in the literature is to impose parametric assumptions on the structure of RUM models. Over the past two decades, an extensive literature in revenue management has studied parametric subclasses of RUM models such as the multinomial logit and nested logit models. By replacing the uncertainty set of all RUM models that fit the historical sales data $\mathcal{U}$ with only RUM models from a specific parametric family (see §1.2), it may be possible to identify an assortment that satisfies (10) over the restricted uncertainty set.

We take a different approach: rather than making parametric assumptions on the structure of the true discrete choice model, we instead use a belief about the process that generated the past assortments that the store offered to its customers. Indeed, recall that in the setting described above, we assumed that the two past assortments S_1 and S_2 were offered by the local store long enough to have accurate estimates of the choice probabilities. This suggests that the two assortments may not have been chosen randomly or adversarially; rather, these assortments were chosen by store managers informed by their private information about local preferences from observing and interacting with typical store customers (Kök et al. 2008, Farias et al. 2017). As such, the parent company may believe it is reasonable to assume that these past assortments are not highly suboptimal. Motivated by this, we will use nominal curves to search for assortments that perform nearly as well as S_2 in the worst case, but are guaranteed to outperform S_2 if the expected revenue from S_2 happens to be close to the optimal value of the assortment optimization problem (8) with the true but unknown discrete choice model $\mathbb{P}$.

Our approach using nominal curves described above, similarly to the more common approach of imposing parametric assumptions on the set of RUM models, involves making assumptions about the true assortment optimization problem (8). However, our approach of using nominal curves is potentially attractive in practice, since it may be easier to justify the assumption that the local store used private information to select their past assortments than justifying a parametric assumption on the structure of RUM models, e.g., that $\mathbb{P}$ follows a multinomial logit model. Moreover, our approach provides explicit guarantees that hold if the belief about the near-optimality of the best past assortment S_2 is incorrect, whereas the misspecification error from a parametric RUM model in assortment optimization is difficult to analyze in general.

4.3 Numerical Results

In Figure 1, we present the nominal curves corresponding to a subset of feasible assortments in the example from §4.1. Specifically, the figure plots the nominal curves $v_S(\eta) \triangleq \min_{\mathbb{P} \in \mathcal{U}_\eta} \sum_{i \in S} r_i \mathbb{P}(i|S)$ for the two past

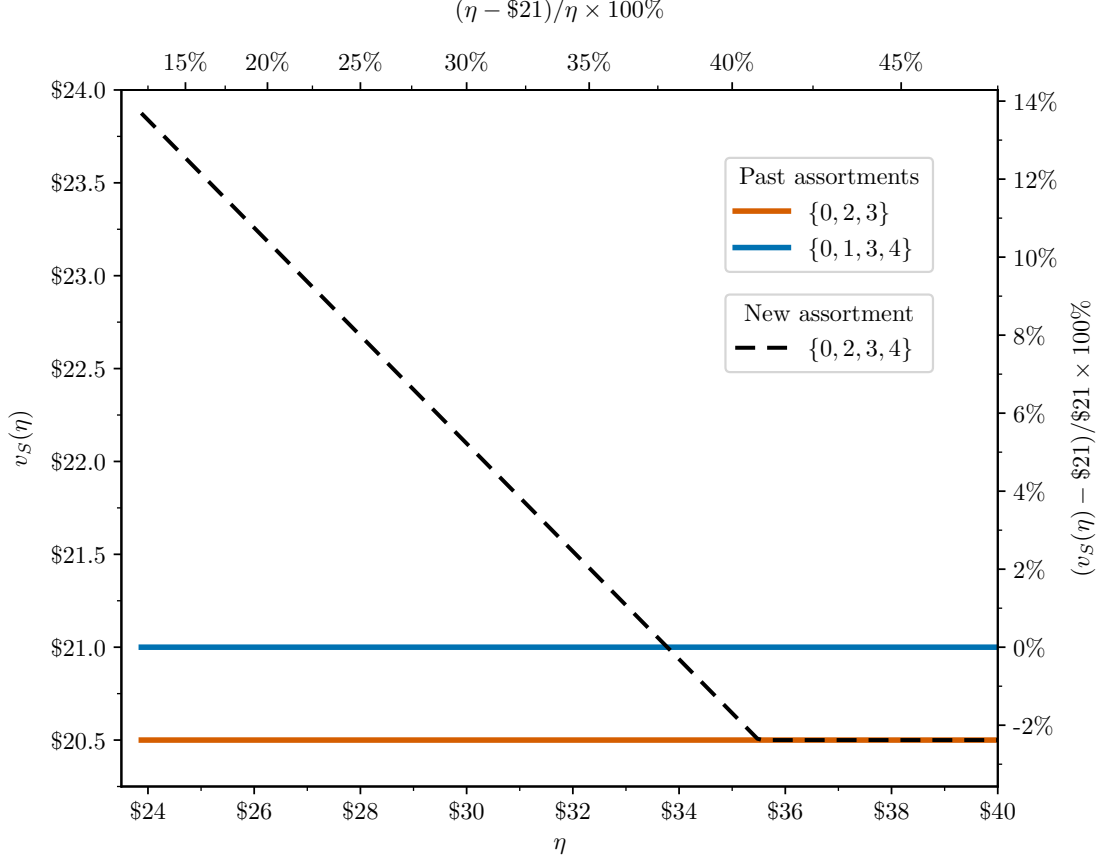


Figure 1: Nominal curves from numerical example in §4.3.

assortments $\{0, 2, 3\}$ and $\{0, 1, 3, 4\}$ as well as a new assortment $\{0, 2, 3, 4\}$, where the reduced uncertainty set

$$\mathcal{U}_\eta = \left\{ \mathbb{P} \in \mathcal{U} : \max_{S \in \mathcal{S}} \sum_{i \in S} r_i \mathbb{P}(i | S) \leq \eta \right\}$$

is the set of all the discrete choice models from the uncertainty set (9) that, if true, would lead to an assortment optimization problem (8) with an optimal expected revenue of at most η . It is shown in Appendix D that the reduced robust optimization problem

$$\max_{S \in \mathcal{S}} \min_{\mathbb{P} \in \mathcal{U}_\eta} \sum_{i \in S} r_i \mathbb{P}(i | S)$$

has S_2 as an optimal solution when $\eta \geq \tilde{\eta}$ and has $\{0, 2, 3, 4\}$ as an optimal solution when $\eta \leq \tilde{\eta}$ and $\mathcal{U}_\eta \neq \emptyset$, where $\tilde{\eta} \approx \$33.778$. The quantities $v_S(\eta)$ are calculated by solving the linear program (2) from Farias et al. (2013, §2.4) with the additional constraints that $\sum_{i \in S'} r_i \mathbb{P}(i | S') \leq \eta$ for all $S' \in \mathcal{S}$.

To make sense of Figure 1, we begin by interpreting the labels. The values of η at the bottom of the plot are possible upper bounds on the optimal value of the nominal problem (8), and the values $(\eta - \$21)/\eta \times 100\%$ at the top of the plot show possible upper bounds on the optimality gap of the store's best past assortment. The values of $v_S(\eta)$ on the left of the plot are the worst-case expected revenue of each assortment S if the optimal value of the nominal problem (8) happens to be less than or equal to η . The values of $(v_S(\eta) - \$21)/\$21 \times 100\%$ on the right of the plot show the worst-case percentage increase in expected revenue from switching the store's best past assortment $S_2 = \{0, 1, 3, 4\}$ to assortment S if the optimal value of the nominal

problem (8) happens to be less than or equal to η .

For example, consider the line from Figure 1 corresponding to the new assortment $\{0, 2, 3, 4\}$ at the point where $\eta = \$30$. At that point, we observe that $(\eta - \$21)/\eta \times 100\% = 30\%$, and it is shown in Figure 1 that $v_{\{0,2,3,4\}}(\$30) \approx \22.097 . We can interpret this point of the nominal curve as saying that if the expected revenue of the store's best past assortment happened to be within 30% of optimal, then the assortment $\{0, 2, 3, 4\}$ is guaranteed to generate an expected revenue of at least \$22.097, or equivalently, guaranteed to increase the store's expected revenue by at least $(v_{\{0,2,3,4\}}(\$30) - \$21)/\$21 \times 100\% = 5.223\%$. By similar reasoning, Figure 1 shows that $\{0, 2, 3, 4\}$ is guaranteed to increase the store's expected revenue by at least 7.988% if the store's best past assortment happened to be within 25% of optimal, and guaranteed to increase the store's expected revenue by at least 10.407% if the store's best past assortment happened to be within 20% of optimal.

From the perspective of the parent company, the main takeaways from Figure 1 are the following. Recall that the expected revenue of the best past assortment S_2 is \$21, which must be within $(\$40 - \$21)/\$40 \times 100\% = 47.5\%$ of optimal since $r_4 = \$40$ is the revenue of the most expensive product. We observe from the nominal curve for the assortment $\{0, 2, 3, 4\}$ that if the optimal value of the assortment optimization problem (8) happens to be less than approximately \$33.778—equivalently, if the best past assortment happens to be within $(\$33.778 - \$21)/\$33.778 \times 100\% \approx 37.8\%$ of optimal—then the assortment $\{0, 2, 3, 4\}$ is guaranteed to generate an expected revenue that is higher than the expected revenue of the store's best past assortment S_2 . If the parent company's belief that the best past assortment is not highly suboptimal is correct, then the new assortment $\{0, 2, 3, 4\}$ has a worst-case expected revenue that can be as large as $(\$23.875 - \$21)/\$21 \times 100\% = 13.7\%$ higher than the expected revenue of the best past assortment; this occurs at $\eta = \$23.875$. If the parent company's belief is incorrect, then the decrease in expected revenue from switching from the best past assortment to $\{0, 2, 3, 4\}$ is at most $-(\$20.5 - \$21)/\$21 \times 100\% = 2.38\%$.

Should the parent company recommend the store to switch to the new assortment $\{0, 2, 3, 4\}$? The answer to this question is ultimately a judgment call for the parent company. That said, we find it reasonable to expect in practice that the parent company would recommend the assortment $\{0, 2, 3, 4\}$ to the store. Indeed, the nominal curve for the new assortment shows that the worst-case decrease in expected revenue for the store is at most 2.38%, and that decrease can occur only if the store's best past assortment was highly suboptimal, which the parent company was assumed to believe is unlikely. Given that the downside risk is mild and believed to be unlikely, and given that the nominal curve shows that the upside for switching to $\{0, 2, 3, 4\}$ may be considerable if the store's best past assortment was not highly suboptimal, the parent company and the store may view the new assortment as attractive.

We conclude with a few additional remarks. First, we observe from Figure 1 that the nominal curves corresponding to the two past assortments S_1 and S_2 remain constant regardless of η . This follows from the construction of the uncertainty set (9), which implies that $v_{S_1}(\eta) = \sum_{i \in S_1} r_i \bar{\mathbb{P}}(i|S_1)$ and $v_{S_2}(\eta) = \sum_{i \in S_2} r_i \bar{\mathbb{P}}(i|S_2)$ for all η such that $\mathcal{U}_\eta \neq \emptyset$. Second, the range of values for η shown in Figure 1 begins at $\eta = \$23.875$ because $\mathcal{U}_\eta$ was found to be empty for $\eta < \$23.875$. Figure 1 also shows that $v_{\{0,2,3,4\}}(\$23.875) = \23.875 . These results are consistent with Proposition 2 and Corollary 2 and imply that $\min_{\mathbb{P} \in \mathcal{U}} \max_{S \in \mathcal{S}} \sum_{i \in S} r_i \mathbb{P}(i|S) = \23.875 . Third, the x -axis ends at $\eta = \$40$ because the optimal value of the assortment optimization problem (8) is at most the revenue of the most expensive product, $r_4 = \$40$.

Nominal curves for assortments not shown in Figure 1 can be found in Appendix D.

5 Combinatorial Optimization with Budget Uncertainty Sets

In §4, we considered an application where the decision maker's belief about the optimal value of the nominal problem derived from a belief that it is unlikely that the best past policy was highly suboptimal. In this section, we consider a different application in which the decision maker's belief may derive from domain expertise and intuition. Specifically, we consider combinatorial optimization problems with uncertain cost coefficients, where the decision maker believes that it is unlikely that the optimal value of the true problem

is small.⁷ We begin in §5.1 by formalizing the application setting. In §5.2, we establish theoretical conditions for this application under which relatively loose beliefs about the optimal value of the nominal problem lead to improved performance guarantees compared to those that can be obtained from robust optimization. In §5.3, we showcase the value of human expertise and the structural results through numerical experiments on shortest path problems.

5.1 Problem Setting

We consider combinatorial optimization programs of the form

$$\min_{I \in \mathcal{I}} \sum_{j \in I} \bar{c}_j \quad (11)$$

where $\bar{\mathbf{c}} \geq \mathbf{0}$ and $\mathcal{I} \subset 2^{\{1, \dots, n\}}$ is a collection of subsets of $\{1, \dots, n\}$ that satisfies $\emptyset \notin \mathcal{I}$. For example, in the context of network problems such as bipartite matching, shortest path, and traveling salesman problems, each $I \in \mathcal{I}$ refers to the edges that are activated, and the cost coefficient $\bar{c}_j \geq 0$ denotes the cost of activating edge j .

We focus on combinatorial optimization problems where the true cost coefficients $\bar{\mathbf{c}}$ are unknown and represented by an uncertainty set $\mathcal{U}$. The study of uncertainty sets in the context of combinatorial optimization with unknown cost coefficients has a rich history dating back to Kouvelis and Yu (1996). Numerous approaches to designing uncertainty sets have been proposed for the context of combinatorial optimization motivated by domain knowledge, risk aversion, and probabilistic guarantees; see Aolaritei et al. (2026). One of the most widely studied choices of the uncertainty set in combinatorial optimization with uncertain cost coefficients is the so-called *budget uncertainty set* (Bertsimas and Sim 2003, §3). It is defined as

$$\begin{aligned} \mathcal{U} &= \left\{ \mathbf{c} \in \mathbb{R}^n : \text{there exists } \mathbf{z} \in [0, 1]^n \text{ such that } \sum_{j=1}^n z_j \leq \Gamma \text{ and } c_j = \hat{c}_j + d_j z_j \text{ for all } j \right\} \\ &= \left\{ \hat{\mathbf{c}} + \mathbf{d} \odot \mathbf{z} : \mathbf{z} \in [0, 1]^n \text{ and } \sum_{j=1}^n z_j \leq \Gamma \right\} \end{aligned} \quad (12)$$

where $\hat{\mathbf{c}} \geq \mathbf{0}$ is a lower bound estimate of the true cost coefficients, $\mathbf{d} > \mathbf{0}$ is an upper bound on the deviations of the cost coefficients, and the integer $\Gamma \geq 1$ controls the number of cost coefficients that can achieve their maximum deviation. As an example, consider a shortest path or traveling salesman problem corresponding to emergency vehicle routing during a flash flood; in that example, $\hat{c}_j$ equals the travel time through two locations denoted by edge j during normal weather conditions, and $\bar{c}_j$ equals the increased travel time.

In this section, we consider applications where the decision maker believes that the optimal value of (11)—that is, the total cost of the optimal solution that we would have chosen if the true cost coefficients were known—is unlikely to be small. In the earlier example of emergency routing in the flash flood, for instance, the exact locations of the flood may be uncertain. Nonetheless, the dispatcher may believe, based on years of experience dispatching during storms and knowledge of the local terrain, that the flooding is likely to be sufficiently widespread that the minimum total travel times will be increased compared to what can normally be achieved in the absence of a storm. A precise statement of such beliefs will be made in §5.2.

For the problem setting described above, the reduced robust optimization problem takes the form

$$\min_{I \in \mathcal{I}} \max_{\mathbf{c} \in \mathcal{U}_\eta} \sum_{j \in I} c_j \text{ where } \mathcal{U}_\eta = \left\{ \mathbf{c} \in \mathcal{U} : \min_{I \in \mathcal{I}} \sum_{j \in I} c_j \geq \eta \right\}.$$

⁷§5 focuses on nominal problems that are minimization problems, and so the belief that the optimal value of the nominal problem is unlikely to be large is reversed to a belief that the optimal value is unlikely to be small. The results from §§2-3 apply to minimization problems by negating the objective function.

In the case of the budget uncertainty set, the above reduced robust optimization problem can be rewritten as

$$\min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in I} (\hat{c}_j + d_j z_j) \quad (13)$$

where the uncertainty set and reduced uncertainty set are rewritten as

$$\mathcal{Z} := \left\{ \mathbf{z} \in [0, 1]^n : \sum_{j=1}^n z_j \leq \Gamma \right\} \text{ and } \mathcal{Z}_\eta = \left\{ \mathbf{z} \in \mathcal{Z} : \min_{I \in \mathcal{I}} \sum_{j \in I} (\hat{c}_j + d_j z_j) \geq \eta \right\}. \quad (14)$$

It is straightforward to apply our results from §3 to the above problem setting. Indeed, the budget uncertainty set $\mathcal{U}$ is convex and compact, the feasible set $\mathcal{I}$ is finite, and the function $f(I, \mathbf{c}) = \sum_{j \in I} c_j$ is linear in $\mathbf{c}$ and bounded on $\mathcal{I}$ and $\mathcal{U}$, and so Assumptions 1 and 2 are satisfied. Since Assumptions 1 and 2 are satisfied, and since (11) is a minimization problem, it follows from Proposition 3 in §3.1 that the optimal value of the reduced robust optimization problem (13) is lower bounded by

$$\max_{\mathbf{z} \in \mathcal{Z}} \min_{I \in \mathcal{I}} \sum_{j \in I} (\hat{c}_j + d_j z_j) \quad (15)$$

and it follows from Corollary 2 from §3.2 that the value of human expertise is equal to the gap between the optimal value of (15) and the optimal value of the robust optimization problem

$$\min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I} (\hat{c}_j + d_j z_j). \quad (16)$$

5.2 Structural Results for Budget Uncertainty Sets

In this subsection, we develop structural results for the reduced robust optimization problem (13) with the budget uncertainty set $\mathcal{Z}$ from (14). The main results of this subsection are theoretical conditions under which the optimal value of the reduced robust optimization problem (13) is strictly less than the optimal value of the robust optimization problem (16). In particular, these theoretical results reveal that surprisingly loose beliefs about the optimal value of the nominal problem can lead to improved performance guarantees compared to those from robust optimization, and these theoretical findings are corroborated numerically in §5.3. All omitted proofs from the present subsection can be found in Appendix C.

Our analysis will make use of the *non-robust* optimization problem, defined as

$$\min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j \quad (17)$$

The relationships between the non-robust problem (17) and the problems from §5.1 are summarized as follows.

Lemma 1. *The optimal value of the non-robust problem (17) is a lower bound on the optimal value of (15). Moreover, if $\eta \leq \min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j$, then $\mathcal{Z}_\eta = \mathcal{Z}$.*

Proof. It follows from the definition of the budget uncertainty set (12) that $\hat{\mathbf{c}} \in \mathcal{U}$, which implies that the optimal value of (17) is a lower bound on the optimal value of (15). It additionally follows from $\mathbf{d} > \mathbf{0}$ and $\mathcal{Z} \subseteq [0, 1]^n$ that $\sum_{j \in I} (\hat{c}_j + d_j z_j) \geq \sum_{j \in I} \hat{c}_j$ for all $I \in \mathcal{I}$ and $\mathbf{z} \in \mathcal{Z}$. This implies that $\mathcal{Z} = \mathcal{Z}_\eta$ for all $\eta \leq \min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j$. $\square$

The first part of Lemma 1, combined with Proposition 3, implies that the optimal value of the non-robust problem (17) is a lower bound on the optimal value of the reduced robust optimization problem (13). The second part of Lemma 1, combined with Proposition 2, shows that the reduced robust optimization problem (13) is well defined and can have an optimal value different from that of the original robust optimization

problem (16) if and only if

$$\eta \in \mathbb{H} := \left(\min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j, \max_{\mathbf{z} \in \mathcal{Z}} \min_{I \in \mathcal{I}} \sum_{j \in I} (\hat{c}_j + d_j z_j) \right]. \quad (18)$$

Equipped with the above, we are ready to develop our main results of §5.2. To motivate these results, recall that Theorem 1 shows that the optimal value of the reduced robust optimization problem (13) will be equal to the optimal value of the max-min problem (15) when η is equal to the optimal value of (15). This implies that if there is a gap between the optimal values of the min-max problem (16) and the max-min problem (15), there must exist some ‘threshold’ for $\eta \in \mathbb{H}$ after which the optimal value of the reduced robust optimization problem (13) becomes strictly less than the optimal value of the min-max problem (16). Our main results of §5.2 characterize the location of this threshold by exploiting the structure of the budget uncertainty set.

The first main result of this subsection, stated below as Corollary 3, establishes simple sufficient conditions for the optimal value of the reduced robust optimization problem (13) to be strictly less than the optimal value of the robust optimization problem (16) for all $\eta \in \mathbb{H}$. Corollary 3 is a special case of a more general result about the worst-case performance of fixed solutions under the reduced budget uncertainty set, stated below as Theorem 2. In the following, we define $\text{supp}(\mathbf{z}) := \{j : z_j > 0\}$.

Theorem 2. *Let $\tilde{I} \in \mathcal{I}$. Then*

$$\max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j) < \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j) \quad \forall \eta \in \mathbb{H} \quad (19)$$

if and only if for every optimal solution $\tilde{\mathbf{z}}$ of $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$, there exists an optimal solution $I^ = I^*(\tilde{\mathbf{z}})$ for the non-robust problem (17) that satisfies $\text{supp}(\tilde{\mathbf{z}}) \cap I^* = \emptyset$.*

In words, Theorem 2 says that a fixed solution $\tilde{I} \in \mathcal{I}$ satisfies (19) if and only if every worst-case realization from the original uncertainty set $\tilde{\mathbf{z}} \in \arg \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$ does not affect the objective value of at least one optimal solution of the non-robust problem (17). Corollary 3 extends Theorem 2 to the optimal value of the reduced robust optimization problem (13) and replaces the ‘if and only if’ claim from Theorem 2 with simpler sufficient conditions.

Corollary 3. *We have*

$$\min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in I} (\hat{c}_j + d_j z_j) < \min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I} (\hat{c}_j + d_j z_j) \quad \forall \eta \in \mathbb{H}$$

if there exists an optimal solution I^{RO} of the robust optimization problem (16) that satisfies either of the following two conditions:

- (a) $|I^{RO}| \geq \Gamma$ and there exists an optimal solution I^* for the non-robust problem (17) that satisfies $I^{RO} \cap I^* = \emptyset$.
- (b) The problem $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{RO}} (\hat{c}_j + d_j z_j)$ has a unique optimal solution $\tilde{\mathbf{z}}$, and there is an optimal solution I^* for the non-robust problem (17) that satisfies $\text{supp}(\tilde{\mathbf{z}}) \cap I^* = \emptyset$.

Informally, the above corollary reveals that the conservatism of the reduced robust optimization problem (13) is driven by the similarity of the optimal solutions of the original robust optimization problem (16) and the non-robust problem (17). More formally, Corollary 3 says that if either of the sufficient conditions from Corollary 3 is satisfied, then any $\eta \in \mathbb{H}$ will make the optimal value of the reduced robust optimization problem (13) strictly less than the optimal value of the original robust optimization problem (16). The first sufficient condition (a) is that there exists an optimal solution of the original robust optimization problem (16) that has a sufficiently large support and is disjoint from an optimal solution for the non-robust

problem (17). The second sufficient condition (b) is that the optimal solution of the original robust optimization problem (16) has a unique worst-case realization from the uncertainty set $\mathcal{Z}$, and that this worst-case realization does not affect the cost coefficients of an optimal solution for the non-robust problem (17). We remark that the uniqueness of an optimal solution of $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j)$ is guaranteed under the assumptions of Corollary 3 if $|I^{\text{RO}}| \geq \Gamma$ and the values of d_j for $j \in I^{\text{RO}}$ are distinct.

It is worthwhile to make a couple of remarks about the practical significance of Theorem 2 and Corollary 3. First, we believe that the sufficient conditions of Corollary 3, or the more general condition from Theorem 2, are relatively mild and may be satisfied in realistic applications. To illustrate this, we show in §5.3 that these conditions can be satisfied in a numerical experiment from Bertsimas and Sim (2003, §6.3). A second takeaway from Corollary 3 is a new insight about the conservatism of robust combinatorial optimization with budget uncertainty sets. Specifically, our analysis shows that if the optimal solution for the robust optimization problem (16) is much different than the optimal solution of the non-robust problem (17) (e.g., if either condition (a) or (b) is satisfied), then Corollary 3 reveals that the worst-case realizations of $\mathbf{z} \in \mathcal{Z}$ for the optimal solution of the robust optimization problem (16) are those that would allow for high-quality best-case performance. This insight is formalized by the following proposition.

Proposition 4. *If I^{RO} is an optimal solution of the robust optimization problem (16) that satisfies conditions (a) or (b), then every worst-case realization*

$$\tilde{\mathbf{z}} \in \arg \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j)$$

satisfies

$$\min_{I \in \mathcal{I}} \sum_{j \in I} (\hat{c}_j + d_j \tilde{z}_j) = \min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j$$

As far as we can tell, Proposition 4 offers a new insight within the robust optimization literature. Specifically, it shows that the parameters that are worst for the optimal solution of the robust optimization problem are parameters under which a decision maker with full information could still achieve the smallest possible optimal cost. Therefore, if the decision maker believes the optimal value of the nominal problem (11) is likely to be strictly greater than the optimal value of the non-robust problem (17), then the parameters that are worst for the optimal solution of the robust optimization problem are ruled out.

This proposition is particularly relevant in the context of the present paper, as it suggests that nominal curves for optimal solutions for (16) may be practically informative. Indeed, suppose that a decision maker computes an optimal solution I^{RO} for the original robust optimization problem (16) but is hesitant about deploying that decision out of concern that the performance of I^{RO} on non-worst-case realizations of $\mathbf{z} \in \mathcal{Z}$ may be nearly as poor as the worst-case performance of I^{RO} . Proposition 4 suggests that this concern can potentially be alleviated through the nominal curve of I^{RO} , if the decision maker believes that the optimal value of the nominal problem (11) is likely to be strictly higher than the optimal value of the non-robust problem (17). We elaborate on this takeaway from Proposition 4 in numerical experiments in §5.3.

We conclude §5.2 by developing results about the optimal value of the reduced robust optimization problem (13) in settings where the conditions from Theorem 2 and Corollary 3 are not satisfied. We develop these results by drawing connections between the reduced uncertainty set $\mathcal{Z}_\eta$ and the *hitting set problem*. To begin, let the collection of near-optimal solutions for the non-robust problem (17) be denoted by

$$\mathcal{I}_\eta := \left\{ I \in \mathcal{I} : \eta - \sum_{j \in I} \hat{c}_j > 0 \right\}. \quad (20)$$

It follows from the above discussion that the above set is nonempty for all $\eta \in \mathbb{H}$. In view of the above notation, the following lemma shows that the constraints in the reduced uncertainty set $\mathcal{Z}_\eta$ are driven by the elements of $\mathcal{I}_\eta$.

Lemma 2. *The reduced uncertainty set satisfies*

$$\mathcal{Z}_\eta = \left\{ \mathbf{z} \in \mathcal{Z} : \sum_{j \in I} d_j z_j \geq \eta - \sum_{j \in I} \hat{c}_j \ \forall I \in \mathcal{I}_\eta \right\}.$$

The above lemma shows that as η increases, the *number* as well as the *tightness* of the constraints in the reduced uncertainty set $\mathcal{Z}_\eta$ increase. In particular, this makes it possible to derive sufficient conditions for when the optimal value of (13) is less than that of (16) based on the combinatorial structure of $\mathcal{I}_\eta$. To that end, let the hitting set number of $\mathcal{I}_\eta$ be defined as the minimum cardinality of a set that intersects each element of $\mathcal{I}_\eta$:

$$\text{HITTINGSETNUM}(\mathcal{I}_\eta) := \min_{\mathbf{z} \in \{0,1\}^n} \left\{ \sum_{j=1}^n z_j : \sum_{j \in I} z_j \geq 1 \text{ for all } I \in \mathcal{I}_\eta \right\}. \quad (21)$$

Equipped with the above terminology, we obtain the following sufficient condition for the optimal value of the reduced robust optimization problem (13) to be strictly less than the optimal value of the robust optimization problem (16) that can hold when the conditions from Theorem 2 and Corollary 3 are not satisfied.

Proposition 5. *Suppose there exists an optimal solution I^{RO} of the robust optimization problem (16) for which $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{RO}} (\hat{c}_j + d_j z_j)$ has a unique solution. If $\eta \in \mathbb{H}$ and $\text{HITTINGSETNUM}(\mathcal{I}_\eta) > \Gamma$, then*

$$\min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in I} (\hat{c}_j + d_j z_j) < \min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I} (\hat{c}_j + d_j z_j)$$

The above proposition gives a condition different from Corollary 3 for when the optimal value of the reduced robust optimization problem (13) will be strictly less than the optimal value of the robust optimization problem (16). The main idea is that if $\text{HITTINGSETNUM}(\mathcal{I}_\eta) > \Gamma$, then it must be the case that $|\text{supp}(\mathbf{z})| > \Gamma$ for all $\mathbf{z} \in \mathcal{Z}_\eta$. This implies that none of the extreme points of $\mathcal{Z}$ are contained in $\mathcal{Z}_\eta$, and thus the unique optimal solution of the linear program $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{RO}} (\hat{c}_j + d_j z_j)$ is not contained in $\mathcal{Z}_\eta$. Proposition 5 thus shows that a strictly positive value of human expertise can arise from the non-robust problem (17) having a diversity of near-optimal solutions.

5.3 Numerical Experiment

Equipped with the results from §5.2, we illustrate the value of human expertise and nominal curves through numerical experiments on a shortest path problem with a budget uncertainty set. Our experiments closely follow the setup from Bertsimas and Sim (2003, §6.3). Each instance of the experiment involves generating a random graph with nodes in $[0, 1]^2$, estimating edge costs $\hat{\mathbf{c}}$ as the Euclidean distance between each pair of nodes in the graph, and using a budget uncertainty set anchored at $\hat{\mathbf{c}}$ of the form given in (12).

In greater detail, each instance of the experiment consists of a randomly generated graph with $|\mathcal{N}| = 60$ nodes and $|\mathcal{E}| = 295$ edges in a two-dimensional Euclidean space. The start node is assigned to coordinate $(0, 0)$, the target node is placed at coordinate $(1, 1)$, and the remaining nodes are assigned uniformly at random over $[0, 1]^2$. The edges are chosen uniformly at random from the set of all pairs of nodes. The set of feasible decisions $\mathcal{I} \subset 2^{\mathcal{E}}$ is the set of all paths from the start node to the target node. The edge cost $\hat{c}_{i,j}$ is calculated as the Euclidean distance between nodes i and j . Each $d_{i,j} \geq 0$ denotes the maximum deviation of the cost on edge (i, j) from its prediction $\hat{c}_{i,j}$, and $d_{i,j}$ is set to be equal to $\gamma_{i,j} \hat{c}_{i,j}$, where $\gamma_{i,j}$ is uniformly distributed in $[0, 8]$. The budget parameter is $\Gamma = 3$.

Our numerical experiment consists of 20 randomly generated instances. In each instance, we solve the robust optimization problem (16) and the non-robust problem (17) once, and the reduced robust optimization problem (13) once for each of 101 different values of η equally spaced over the closure of $\mathbb{H}$. We solve the reduced robust optimization problem (13) for each such η using the mixed-integer linear programming formulation from Proposition 6 in §6.1.1. The results of the experiments are shown in Figure 2 and Table 1.

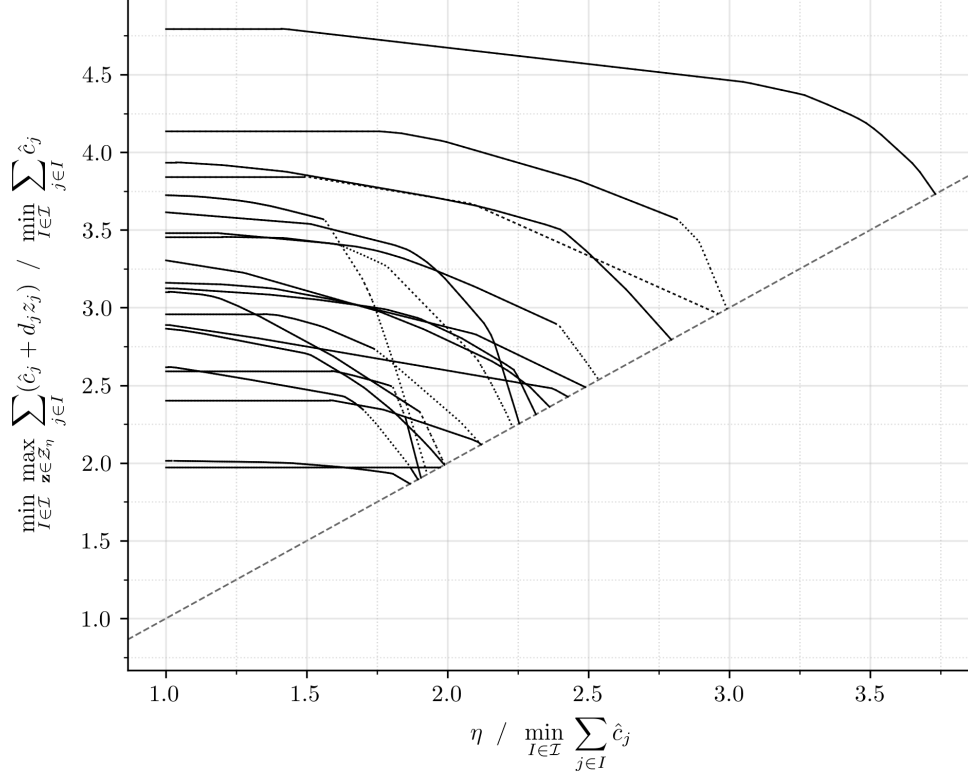


Figure 2: Results of numerical experiments from §5.3.

Figure 2 plots a line for each of the 20 randomly generated instances of the shortest path problem. For each instance, the corresponding line in Figure 2 shows the optimal value of the reduced robust optimization problem (13) as a function of η , where both quantities are normalized by the optimal value of the non-robust problem (17). In other words, Figure 2 shows the minimization equivalent of the upper bound curves (3) from §2.3. The line corresponding to an instance is solid if the optimal solution for the robust optimization problem (16) is an optimal solution for the reduced robust optimization problem (13), and the line is dotted otherwise. The diagonal dashed line shows $y = x$. Consistent with Corollary 2, we observe from Figure 2 for each instance that when η reaches its maximum value, the optimal value of (13) is equal to η .

Table 1 shows the optimal solution for the non-robust problem (17) and the optimal solution for the robust optimization problem (16) for each of the random instances. For the optimal solution I^{RO} of the robust optimization problem (16), a star is indicated above each edge in Table 1 that is in the support of the optimal solution of $\max_{z \in Z} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j)$. Table 1 uses checkmarks to indicate whether Conditions (a) and (b) from Corollary 3 are satisfied.

There are several takeaways from Figure 2 and Table 1. First, we observe that there is a gap between the optimal values of (16) and (15)—and thus a value of human expertise that is strictly positive—in 19 out of the 20 random instances. Moreover, when there is a gap between the optimal values of (16) and (15), the gap can be considerable, with an average ratio of 1.389 between the optimal values of (16) and (15) over the 19 instances with a strict gap. Second, Table 1 shows that six of the 20 random instances satisfied at least one of the sufficient conditions from Corollary 3, thereby implying that the optimal value of the reduced robust optimization problem (13) is strictly less than the optimal value of the robust optimization problem (16) for all $\eta \in \mathbb{H}$. Third, in many of the curves in Figure 2, the optimal solution for the robust optimization problem (16) remained optimal for the reduced robust optimization problem (13) for many values of η , including those for which the optimal value of (13) is strictly less than that of (16). This is consistent with Proposition 4 and demonstrates that nominal curves of optimal solutions for traditional robust combinatorial

Table 1: Optimal solutions for numerical experiments from §5.3.

	Optimal Solution of (17)	Optimal Solution of (16)	Cond (a)	Cond (b)
1	$1 \rightarrow 31 \rightarrow 41 \rightarrow 33 \rightarrow 60$	$1 \xrightarrow{*} 31 \rightarrow 41 \xrightarrow{*} 12 \rightarrow 34 \rightarrow 43 \xrightarrow{*} 60$		
2	$1 \rightarrow 12 \rightarrow 34 \rightarrow 60$	$1 \xrightarrow{*} 58 \rightarrow 2 \xrightarrow{*} 27 \xrightarrow{*} 34 \rightarrow 60$		✓
3	$1 \rightarrow 23 \rightarrow 49 \rightarrow 17 \rightarrow 34 \rightarrow 14 \rightarrow 60$	$1 \xrightarrow{*} 8 \rightarrow 2 \xrightarrow{*} 26 \rightarrow 56 \rightarrow 37 \rightarrow 21 \rightarrow 14 \xrightarrow{*} 60$		
4	$1 \rightarrow 16 \rightarrow 40 \rightarrow 60$	$1 \rightarrow 16 \rightarrow 8 \xrightarrow{*} 48 \xrightarrow{*} 38 \rightarrow 28 \rightarrow 17 \xrightarrow{*} 60$		✓
5	$1 \rightarrow 36 \rightarrow 44 \rightarrow 60$	$1 \xrightarrow{*} 36 \xrightarrow{*} 44 \xrightarrow{*} 60$		
6	$1 \rightarrow 52 \rightarrow 60$	$1 \xrightarrow{*} 52 \xrightarrow{*} 46 \rightarrow 33 \rightarrow 58 \xrightarrow{*} 49 \rightarrow 10 \rightarrow 60$		
7	$1 \rightarrow 47 \rightarrow 44 \rightarrow 60$	$1 \rightarrow 38 \xrightarrow{*} 3 \rightarrow 39 \rightarrow 59 \xrightarrow{*} 44 \xrightarrow{*} 60$		
8	$1 \rightarrow 2 \rightarrow 60$	$1 \xrightarrow{*} 45 \rightarrow 46 \xrightarrow{*} 2 \xrightarrow{*} 60$		
9	$1 \rightarrow 25 \rightarrow 35 \rightarrow 38 \rightarrow 60$	$1 \xrightarrow{*} 16 \rightarrow 43 \xrightarrow{*} 30 \rightarrow 41 \rightarrow 38 \xrightarrow{*} 60$		
10	$1 \rightarrow 60$	$1 \xrightarrow{*} 45 \xrightarrow{*} 40 \xrightarrow{*} 60$	✓	✓
11	$1 \rightarrow 44 \rightarrow 36 \rightarrow 41 \rightarrow 60$	$1 \rightarrow 44 \xrightarrow{*} 36 \rightarrow 41 \xrightarrow{*} 9 \xrightarrow{*} 2 \rightarrow 60$		
12	$1 \rightarrow 56 \rightarrow 59 \rightarrow 7 \rightarrow 41 \rightarrow 60$	$1 \xrightarrow{*} 56 \rightarrow 59 \rightarrow 7 \xrightarrow{*} 13 \xrightarrow{*} 60$		
13	$1 \rightarrow 41 \rightarrow 37 \rightarrow 60$	$1 \xrightarrow{*} 31 \xrightarrow{*} 60$		
14	$1 \rightarrow 51 \rightarrow 22 \rightarrow 23 \rightarrow 60$	$1 \xrightarrow{*} 26 \rightarrow 38 \rightarrow 36 \xrightarrow{*} 14 \rightarrow 23 \xrightarrow{*} 60$		
15	$1 \rightarrow 23 \rightarrow 60$	$1 \xrightarrow{*} 26 \xrightarrow{*} 60$		
16	$1 \rightarrow 46 \rightarrow 14 \rightarrow 32 \rightarrow 60$	$1 \rightarrow 34 \xrightarrow{*} 46 \xrightarrow{*} 9 \xrightarrow{*} 60$	✓	✓
17	$1 \rightarrow 36 \rightarrow 58 \rightarrow 60$	$1 \xrightarrow{*} 54 \xrightarrow{*} 23 \xrightarrow{*} 60$	✓	✓
18	$1 \rightarrow 38 \rightarrow 8 \rightarrow 24 \rightarrow 54 \rightarrow 18 \rightarrow 60$	$1 \xrightarrow{*} 38 \rightarrow 8 \rightarrow 24 \xrightarrow{*} 54 \rightarrow 18 \xrightarrow{*} 60$		
19	$1 \rightarrow 60$	$1 \xrightarrow{*} 60$		
20	$1 \rightarrow 60$	$1 \xrightarrow{*} 51 \xrightarrow{*} 55 \rightarrow 50 \xrightarrow{*} 60$	✓	✓

optimization problems (16) with budget uncertainty sets can reveal non-trivial performance guarantees.

6 Finding Policies and Controlling Disappointment

To use the developments from this paper in real-world applications, we recommend using the approach described in §2.3 for generating a menu of policies to offer to a decision maker. That is, we recommend generating a menu of policies by solving the reduced robust optimization problem (4) over a range of values of η and comparing those policies to one another through their nominal curves (2). By comparing the nominal curves of the policies, the decision maker can apply their own judgment about which scenarios of the nominal problem they believe are plausible and then select one of the policies.

In this section, we facilitate the practical deployment of the approach from §2.3 in several ways. In §6.1, we propose two computational methods that can be used to solve the reduced robust optimization problem (4) and generate the menu of policies. In §6.2, we propose two alternative approaches to generating a menu of policies that may be useful when the decision maker does not find the nominal curves of the policies obtained from the approach in §2.3 to be desirable.

6.1 Algorithms

In this section, we propose two computational methods for solving the reduced robust optimization problem (4). The first is an exact reformulation for a class of mixed-integer linear-fractional programs. The second is more general and based on the cutting plane method.

6.1.1 Exact Reformulations

In some applications, the reduced robust optimization problem (4) can be solved by reformulating it as a compact mixed-integer linear program. Below, we present such reformulations for the following applications when the uncertainty set is a nonempty bounded polyhedron.

Example 3. Combinatorial optimization problems of the form

$$\max_{\mathbf{x} \in \{0,1\}^n : \mathbf{Ax} \leq \mathbf{b}} \sum_{j=1}^n \bar{\theta}_j x_j$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is totally unimodular, $\mathbf{b}$ is integral, and the constraints of the form $\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}$ have without loss of generality been embedded within $\mathbf{Ax} \leq \mathbf{b}$. Examples include shortest path, bipartite matching, and linear assignment problems.

Example 4. Cardinality-constrained assortment optimization problems under the multinomial logit model,

$$\max_{S \subseteq \{1, \dots, n\} : |S| \leq k} \sum_{i \in S} \frac{r_i \bar{\theta}_i}{1 + \sum_{j \in S} \bar{\theta}_j} \quad (22)$$

where $r_1, \dots, r_n > 0$ and $\bar{\theta}_1, \dots, \bar{\theta}_n \geq 0$. Note that (22) does not satisfy Assumption 2 because $\boldsymbol{\theta} \mapsto \sum_{i \in S} r_i \theta_i / (1 + \sum_{j \in S} \theta_j)$ is a nonconvex function in general. Even though the conditions of Corollary 2 are not satisfied in this setting, the value of human expertise can still be strictly positive; we show an example of this at the end of §6.1.1.

To develop a mixed-integer linear programming reformulation of (4) in the above examples, we focus on the more general nominal problem

$$\max_{\mathbf{x} \in \mathcal{X}} \frac{\mathbf{x}^\top \mathbf{U} \bar{\boldsymbol{\theta}} + \mu}{\mathbf{x}^\top \mathbf{L} \bar{\boldsymbol{\theta}} + \lambda}$$

where $\mathcal{X} = \{\mathbf{x} \in \{0,1\}^n : \mathbf{Ax} \leq \mathbf{b}\} \neq \emptyset$ satisfies $\text{conv}(\mathcal{X}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}\}$, the uncertainty set is a bounded polyhedron $\mathcal{U} = \{\boldsymbol{\theta} \in \mathbb{R}^p : \mathbf{D}\boldsymbol{\theta} \geq \mathbf{g}\} \neq \emptyset$, and the dimensions are $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{D} \in \mathbb{R}^{q \times p}$, $\mathbf{g} \in \mathbb{R}^q$, $\mathbf{U}, \mathbf{L} \in \mathbb{R}^{n \times p}$, and $\mu, \lambda \in \mathbb{R}$. We observe that Examples 3 and 4 are special cases of this setting.

In what follows, we show that the reduced robust optimization problem (4) corresponding to the above setting, denoted by

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \frac{\mathbf{x}^\top \mathbf{U} \boldsymbol{\theta} + \mu}{\mathbf{x}^\top \mathbf{L} \boldsymbol{\theta} + \lambda}, \quad (23)$$

can be reformulated as a mixed-integer linear program of polynomial size. The key idea behind the reformulation of (23) is applying strong duality twice: once to construct a polynomial-size extended formulation of the reduced uncertainty set $\mathcal{U}_\eta$, and once to dualize the inner problem of the reduced robust optimization problem (23). We first state our reformulation of (23) for the special case of Example 3, followed by the reformulation for the general case; the proofs of both of these propositions are found in Appendix E.

Proposition 6. *If $\mathcal{U}_\eta$ is nonempty, then*

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \boldsymbol{\theta}^\top \mathbf{x} = \left[\begin{array}{ll} \text{maximize} & \mathbf{g}^\top \boldsymbol{\psi} - \eta \sigma \\ \text{subject to} & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{A} (\mathbf{D}^\top \boldsymbol{\psi} - \mathbf{x}) \leq \mathbf{b} \sigma \\ & \boldsymbol{\psi} \geq \mathbf{0}, \sigma \geq 0 \end{array} \right]$$

Proposition 7. If $\mathcal{U}_\eta$ is nonempty, $\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda > 0$ for all $\mathbf{x} \in \mathcal{X}$ and $\boldsymbol{\theta} \in \mathcal{U}$, and the optimal value of (23) lies in $[\underline{t}, \bar{t}]$ for $0 \leq \underline{t} \leq \bar{t}$, then

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \frac{\mathbf{x}^\top \mathbf{U}\boldsymbol{\theta} + \mu}{\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda} = \left[\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & \begin{array}{l} \mathbf{x} \in \{0,1\}^n, t \in \mathbb{R}, \boldsymbol{\psi} \in \mathbb{R}^q, \\ \mathbf{y} \in \mathbb{R}^n, \sigma \in \mathbb{R}, \mathbf{z} \in \mathbb{R}^n \\ t\lambda - \mu \leq \mathbf{g}^\top \boldsymbol{\psi} - (\eta\lambda - \mu)\sigma \\ \mathbf{D}^\top \boldsymbol{\psi} - (\mathbf{U} - \eta\mathbf{L})^\top \mathbf{y} = \mathbf{U}^\top \mathbf{x} - \mathbf{L}^\top \mathbf{z} \\ \underline{t} \leq t \leq \bar{t} \\ \underline{t}x_i \leq z_i \leq \bar{t}x_i \quad \forall i \in \{1, \dots, n\} \\ t - \bar{t}(1 - x_i) \leq z_i \leq t - \underline{t}(1 - x_i) \quad \forall i \in \{1, \dots, n\} \\ \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ \mathbf{A}\mathbf{y} \leq \mathbf{b}\sigma \\ \boldsymbol{\psi} \geq \mathbf{0}, \sigma \geq 0 \end{array} \end{array} \right].$$

The mixed-integer programming formulation from Proposition 6 is the reformulation of (23) for the special case of $n = p$, $\mathbf{L} = \mathbf{0}$, $\mathbf{U}$ is the identity matrix, $\lambda = 1$, and $\mu = 0$. As such, the formulation from Proposition 6 corresponds to Example 3, and this formulation is used in the numerical experiments from §5.3.⁸ The mixed-integer programming formulation from Proposition 7 is general and requires that we have a known nonnegative lower bound $\underline{t}$ and upper bound $\bar{t}$ on the optimal value of (23). In the context of Example 4, the lower bound can be chosen as \$0 or as the worst-case expected revenue of a heuristic assortment, and the upper bound can be chosen as the revenue of the most expensive product.

We conclude §6.1.1 by providing an instance of Example 4 in which the value of human expertise is strictly positive. The example is useful for two reasons. First, it shows that the exact reformulation from Proposition 7 is not vacuous, in the sense that the reduced robust optimization problem (4) can differ from the robust optimization problem (5) and is thus worthwhile to solve. Second, it demonstrates that the value of human expertise can sometimes be strictly positive in settings where the conditions of Corollary 2 are not satisfied, although this is not the case in general as shown by Example 2 in §3.2.

Example 5. Let $n = 2$, $k = 1$, $r_1 = r_2 = 1$, and $\mathcal{U} = \{\boldsymbol{\theta} \in \mathbb{R}^2 : \sum_{i=1}^2 \theta_i = 3, \boldsymbol{\theta} \geq \mathbf{1}\}$. Then

$$\begin{aligned} \max_{S: |S| \leq 1} \min_{\boldsymbol{\theta} \in \mathcal{U}} \sum_{i \in S} \frac{r_i \theta_i}{1 + \sum_{j \in S} \theta_j} &= \min_{\boldsymbol{\theta} \geq \mathbf{1}: \mathbf{1}^\top \boldsymbol{\theta} = 3} \frac{\theta_1}{1 + \theta_1} = \frac{1}{2}, \text{ and} \\ \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{S: |S| \leq 1} \sum_{i \in S} \frac{r_i \theta_i}{1 + \sum_{j \in S} \theta_j} &= \frac{3/2}{1 + 3/2} = \frac{3}{5}. \end{aligned}$$

We observe that if $\eta = 3/5$, then

$$\begin{aligned} \max_{S: |S| \leq 1} \sum_{i \in S} \frac{r_i \theta_i}{1 + \sum_{j \in S} \theta_j} \leq \eta &\implies \frac{\theta_i}{1 + \theta_i} \leq \frac{3}{5} \quad \forall i \in \{1, 2\} \\ &\implies \theta_i \leq \frac{3}{2} \quad \forall i \in \{1, 2\}, \end{aligned}$$

and so it follows from the definition of $\mathcal{U}$ that $\mathcal{U}_{3/5} = \{(3/2, 3/2)\}$. We thus conclude for the case of $\eta = 3/5$ that

$$\max_{S: |S| \leq 1} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \sum_{i \in S} \frac{r_i \theta_i}{1 + \sum_{j \in S} \theta_j} = \max_{S: |S| \leq 1} \sum_{i \in S} \frac{3/2}{1 + 3/2} = \frac{3}{5},$$

and hence the value of human expertise is $\Delta = 3/5 - 1/2 = 1/10$. $\square$

⁸The numerical experiments in §5.3 focus on the min-max formulation of the problem, which can be obtained from Proposition 6 by negating the objective function.

6.1.2 Cutting Plane Method

In applications where the reformulation techniques from §6.1.1 do not apply, we propose a two-level cutting plane method for solving the reduced robust optimization problem (4). Let Assumption 1 hold. The two-level cutting plane method is presented in Algorithm 1.

In greater detail, the two-level cutting plane method in Algorithm 1 is motivated by applications where the policy space $\mathcal{X}$ is finite but has many elements and the constraint $\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) \leq \eta$ in the reduced uncertainty set $\mathcal{U}_\eta$ does not have a compactly representable dual. Examples include traveling salesman problems or assortment optimization problems of the form studied in §4 with large numbers of products. Algorithm 1 addresses the fact that $\mathcal{U}_\eta$ cannot be represented compactly by using a cutting plane method to approximate the reduced uncertainty set; this corresponds to the loop within Step 2. Indeed, we observe that Step 2 in Algorithm 1 begins with an $\mathbf{x}^* \in \mathcal{X}$ and outputs an optimal solution $\boldsymbol{\theta}^* \in \arg \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}^*, \boldsymbol{\theta})$. The outer cutting plane method solves $\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}, \boldsymbol{\theta})$ by solving $\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \tilde{\mathcal{U}}} f(\mathbf{x}, \boldsymbol{\theta})$ (Step 1) and adding parameters to $\tilde{\mathcal{U}}$ (Step 3) until convergence. The initial $\boldsymbol{\theta} \in \mathcal{U}_\eta$ in Step 0 can be found, for example, by running Step 2 once before the method begins with any initial $\mathbf{x} \in \mathcal{X}$.

The inner cutting plane method (Step 2 in Algorithm 1) maintains its set of cuts $\tilde{\mathcal{X}}$ across iterations of the outer cutting plane method (Steps 1-3 in Algorithm 1). This is intended to warm start the problem of solving $\min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} f(\mathbf{x}^*, \boldsymbol{\theta})$ in each iteration of the outer cutting plane method. For example, suppose one is considering a reduced robust optimization problem $\max_{\mathbf{x} \in \mathcal{X} \subseteq \{0,1\}^n} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \boldsymbol{\theta}^\top \mathbf{x}$ where the set $\mathcal{X}$ is the set of feasible tours in a traveling salesman problem and $\mathcal{U}$ is a compactly representable polytope. In this case, the inner cutting plane method (Step 2) requires iteratively solving the optimization problem $\max_{\mathbf{x} \in \mathcal{X} \subseteq \{0,1\}^n} f(\mathbf{x}, \boldsymbol{\theta})$, which may be computationally expensive. By maintaining the set $\tilde{\mathcal{X}}$ of cuts on the reduced uncertainty set $\mathcal{U}_\eta$ across iterations, the problem $\max_{\mathbf{x} \in \mathcal{X} \subseteq \{0,1\}^n} f(\mathbf{x}, \boldsymbol{\theta})$ may be solved fewer times in total across the iterations of the outer cutting plane method (Steps 1-3).

6.2 Alternative Approaches to Generating Policies

If none of the nominal curves (2) in the menu of policies obtained by the approach from §2.3 are deemed satisfactory to a decision maker, then we suggest two alternative approaches for generating policies.

The first alternative approach, which is based on a modeling technique from the literature on globalized robust optimization (Ben-Tal et al. 2006), is motivated by settings where the decision maker has a concrete guess of an upper bound η on the optimal value of the nominal problem (1) but wants to control the worst-case performance of the policy if their guess is incorrect. This approach obtains a policy by solving

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}} \left\{ f(\mathbf{x}, \boldsymbol{\theta}) + \lambda \max \left\{ \max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \boldsymbol{\theta}) - \eta, 0 \right\} \right\} \quad (24)$$

where $\lambda \in [0, \infty)$ is a parameter selected by a decision maker that controls the disappointment from selecting an η that is not an upper bound on the optimal value of (1).

To make sense of the above alternative approach to obtaining a policy, let us make some observations. First, we observe that the optimal value of (24) is nondecreasing in λ . In the extreme case where $\lambda = 0$, we observe that (24) equals the standard robust optimization problem (5), and it is straightforward to see that (24) simplifies to the reduced robust optimization problem (4) in the case where $\lambda \rightarrow \infty$. In the other cases of λ , the interpretation of λ follows from the following Proposition 8, which specifies the relationship between the reduced uncertainty set $\mathcal{U}_\eta$ and (24).

Proposition 8. *If $\mathbf{x}^*$ is an optimal solution for (24) and $v(\eta, \lambda)$ is the optimal value of (24), then*

$$\begin{aligned} f(\mathbf{x}^*, \boldsymbol{\theta}) &\geq v(\eta, \lambda) & \forall \boldsymbol{\theta} \in \mathcal{U}_\eta, \\ f(\mathbf{x}^*, \boldsymbol{\theta}) &\geq v(\eta, \lambda) - \lambda \left(\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) - \eta \right) & \forall \boldsymbol{\theta} \in \mathcal{U} \setminus \mathcal{U}_\eta. \end{aligned}$$

Two-Level Cutting Plane Method

Step 0 Choose an initial $\boldsymbol{\theta} \in \mathcal{U}_\eta$ and let $\tilde{\mathcal{U}} \leftarrow \{\boldsymbol{\theta}\}$ and $\tilde{\mathcal{X}} \leftarrow \emptyset$.

Step 1 Find an optimal solution $(\mathbf{x}^*, t^*)$ for the optimization problem

$$\begin{aligned} & \underset{\mathbf{x} \in \mathcal{X}, t \in \mathbb{R}}{\text{maximize}} && t \\ & \text{subject to} && t \leq f(\mathbf{x}, \boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \tilde{\mathcal{U}} \end{aligned}$$

The existence of an optimal solution follows from Assumption 1 and the fact that $\tilde{\mathcal{U}}$ is a finite set.

Step 2 Do the following steps.

Step 2a Find an optimal solution $\boldsymbol{\theta}^*$ for the optimization problem

$$\begin{aligned} & \underset{\boldsymbol{\theta} \in \tilde{\mathcal{U}}}{\text{minimize}} && f(\mathbf{x}^*, \boldsymbol{\theta}) \\ & \text{subject to} && f(\mathbf{x}, \boldsymbol{\theta}) \leq \eta \quad \forall \mathbf{x} \in \tilde{\mathcal{X}} \end{aligned}$$

The existence of an optimal solution follows from Assumption 1 and the fact that $\tilde{\mathcal{X}}$ is a finite set.

Step 2b Find an optimal solution $\hat{\mathbf{x}}$ for the optimization problem

$$\hat{v} = \max_{\mathbf{x} \in \tilde{\mathcal{X}}} f(\mathbf{x}, \boldsymbol{\theta}^*)$$

Step 2c If $\hat{v} > \eta$, then let $\tilde{\mathcal{X}} \leftarrow \tilde{\mathcal{X}} \cup \{\hat{\mathbf{x}}\}$ and return to Step 2a. If $\hat{v} \leq \eta$, then proceed with $\boldsymbol{\theta}^*$ to Step 3.

Step 3 If $t^* > f(\mathbf{x}^*, \boldsymbol{\theta}^*)$, then let $\tilde{\mathcal{U}} \leftarrow \tilde{\mathcal{U}} \cup \{\boldsymbol{\theta}^*\}$ and return to Step 1. If $t^* \leq f(\mathbf{x}^*, \boldsymbol{\theta}^*)$, then $\mathbf{x}^*$ is an optimal solution for the reduced robust optimization problem (4)

Algorithm 1: The cutting plane method from §6.1.2 for the reduced robust optimization problem (4).

Proof. If $\mathbf{x}^*$ is an optimal solution for (24), then for all $\boldsymbol{\theta} \in \mathcal{U}$, we have

$$f(\mathbf{x}^*, \boldsymbol{\theta}) + \lambda \max \left\{ \max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \boldsymbol{\theta}) - \eta, 0 \right\} \geq v(\eta, \lambda).$$

Rearranging the above inequality and applying the definition of $\mathcal{U}_\eta$ yields the desired result. $\square$

The above proposition shows that larger values of λ imply that the performance of optimal solutions for (24) is potentially less conservative, but the performance of those policies is degraded if η is not an upper bound on (1). As such, we can solve (24) with varying choices of $\lambda \in [0, \infty)$ to obtain tradeoffs between performance guarantees and confidence about the accuracy of the estimate η . Similarly to the reduced robust optimization problem (4), the inner problem of (24) is a convex optimization problem for a fixed policy under the assumptions from §3, as shown by the following proposition.

Proposition 9. *If Assumptions 1 and 2 hold and $\lambda \in [0, \infty)$, then for each $\mathbf{x} \in \mathcal{X}$, the inner problem $\min_{\boldsymbol{\theta} \in \mathcal{U}} \{f(\mathbf{x}, \boldsymbol{\theta}) + \lambda \max \{\max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \boldsymbol{\theta}) - \eta, 0\}\}$ is a convex optimization problem.*

Proof. The uncertainty set $\mathcal{U}$ is convex by Assumption 2. Given a fixed policy $\mathbf{x} \in \mathcal{X}$, the objective function of the inner problem is the sum of a convex function $\boldsymbol{\theta} \mapsto f(\mathbf{x}, \boldsymbol{\theta})$ and the maximum of 0 and a function of $\boldsymbol{\theta} \mapsto \lambda(\max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \boldsymbol{\theta}) - \eta)$ that is convex since $\lambda \in [0, \infty)$ and $\boldsymbol{\theta} \mapsto f(\mathbf{y}, \boldsymbol{\theta})$ is convex for each $\mathbf{y}$. Therefore, the inner problem of (24) is a convex optimization problem. $\square$

Our second alternative approach is to add hard constraints into the reduced robust optimization problem (4) to restrict to policies that have acceptable worst-case performance guarantees under different possible upper bounds on the optimal value of (1). Given an $\eta_1 \geq \eta$ and an acceptable set $\mathcal{A} \equiv \{(\eta_2, \nu_2), \dots, (\eta_m, \nu_m)\} \subseteq [\eta, \infty) \times \mathbb{R}$ of pairs of upper bounds and worst-case performance guarantees, one can solve

$$\begin{aligned} & \underset{\mathbf{x} \in \mathcal{X}}{\text{maximize}} && \min_{\boldsymbol{\theta} \in \mathcal{U}_{\eta_1}} f(\mathbf{x}, \boldsymbol{\theta}) \\ & \text{subject to} && \min_{\boldsymbol{\theta} \in \mathcal{U}_{\eta_i}} f(\mathbf{x}, \boldsymbol{\theta}) \geq \nu_i \quad \forall i \in \{2, \dots, m\} \end{aligned} \tag{25}$$

The problem (25) gives the option to have fine-grained control over the performance of a policy under different bounds $\eta_1, \dots, \eta_m$. Similarly to the first alternative approach (24), a menu of policies can be obtained by solving (25) with varying choices of acceptable sets.

7 Conclusion and Open Questions

In this work, we studied optimization applications with unknown parameters where the decision maker believes that the optimal value of the nominal problem is unlikely to be large. We proposed nominal curves for translating such beliefs into performance guarantees while retaining guarantees when the belief is incorrect. We introduced a quantity called the *value of human expertise* that captures the maximum improvement in performance guarantees from incorporating such beliefs, and our main result showed under mild assumptions that this quantity is equal to the gap between the optimal values of a min-max and max-min problem. We showed that this gap can be substantial in applications such as assortment optimization and shortest path problems, and that improved performance guarantees can be obtained even under relatively loose beliefs about the nominal problem. We also proposed several approaches to generating menus of policies, as well as computational methods for finding policies that are pointwise optimal.

There are many important directions for future work, including empirical testing, case studies, and connections to fields such as decision theory. One question that could be interesting to investigate is whether the *optimizer's curse* phenomenon (Smith and Winkler 2006, Van Parys et al. 2021, Gupta et al. 2024, Xu et al. 2025, Bastani et al. 2025) can be exploited to construct statistical upper bounds on the optimal value of the nominal problem that lead to improved performance guarantees. Another direction would be to study optimization with human expertise in specific application classes, such as mechanism design and stochastic programming with chance constraints. It would also be interesting to study the use of nominal curves in human-AI collaboration, such as in settings where agents are deployed to solve high-stakes operations management problems.

References

- Shabbir Ahmed and Yongpei Guan. The inverse optimal value problem. *Mathematical Programming*, 102(1):91–110, 2005.
- Ravindra K Ahuja and James B Orlin. Inverse optimization. *Operations Research*, 49(5):771–783, 2001.
- Jerry Anunrojwong, Santiago R Balseiro, and Omar Besbes. The best of many robustness criteria in decision making: Formulation and application to robust pricing. *arXiv preprint arXiv:2403.12260*, 2024.
- Liviu Aolaritei, Ricky Huang, Michael I Jordan, and Paul Grigas. Diffusion-robust optimization over graphs. *arXiv preprint arXiv:2605.30853*, 2026.
- Anil Aswani, Zuo-Jun Shen, and Auyon Siddiq. Inverse optimization with noisy data. *Operations Research*, 66(3):870–892, 2018.
- Achraf Bahamou, Omar Besbes, and Omar Mouchtaki. Fast revenue maximization. *arXiv preprint arXiv:2407.07316*, 2024.
- Maya Balakrishnan, Kris Johnson Ferreira, and Jordan Tong. Human-algorithm collaboration with private

- information: Naïve advice-weighting behavior and mitigation. *Management Science*, 72(1):265–284, 2026.
- Hamsa Bastani, Osbert Bastani, and Bryce McLaughlin. Beating the winner’s curse via inference-aware policy optimization. *arXiv preprint arXiv:2510.18161*, 2025.
- Aharon Ben-Tal, Stephen Boyd, and Arkadi Nemirovski. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming*, 107(1):63–89, 2006.
- Omar Bennouna, Amine Bennouna, Saurabh Amin, and Asuman Ozdaglar. What data enables optimal decisions? an exact characterization for linear optimization. *Advances in Neural Information Processing Systems*, 38:166020–166049, 2025.
- Omar Bennouna, Amine Bennouna, Saurabh Amin, and Asuman Ozdaglar. Data informativeness in linear optimization under uncertainty. *arXiv preprint arXiv:2602.15365*, 2026.
- Dirk Bergemann and Karl Schlag. Robust monopoly pricing. *Journal of Economic Theory*, 146(6):2527–2543, 2011.
- Dimitris Bertsimas and Velibor V Mišić. Robust product line design. *Operations Research*, 65(1):19–37, 2017.
- Dimitris Bertsimas and Melvyn Sim. Robust discrete optimization and network flows. *Mathematical Programming*, 98(1):49–71, 2003.
- H.D. Block and Jacob Marschak. Random orderings and stochastic theories of response. Cowles Foundation Discussion Papers 66, Cowles Foundation for Research in Economics, Yale University, 1959. URL <https://EconPapers.repec.org/RePEc:cwl:cwldpp:66>.
- Timothy CY Chan, Rafid Mahmood, and Ian Yihang Zhu. Inverse optimization: Theory and applications. *Operations Research*, 73(2):1046–1074, 2025.
- Erick Delage, Daniel Kuhn, and Wolfram Wiesemann. “dice”-sion-making under uncertainty: when can a random decision reduce risk? *Management Science*, 65(7):3282–3301, 2019.
- Antoine Désir, Vineet Goyal, Bo Jiang, Tian Xie, and Jiawei Zhang. Robust assortment optimization under the markov chain choice model. *Operations Research*, 72(4):1595–1614, 2024.
- Ky Fan. Minimax theorems. *Proceedings of the National Academy of Sciences*, 39(1):42–47, 1953.
- Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. A nonparametric approach to modeling choice with limited data. *Management Science*, 59(2):305–322, 2013.
- Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. Building optimized and hyperlocal product assortments: A nonparametric choice approach. *Available at SSRN 2905381*, 2017.
- Xinyi Guan and Velibor V Mišić. Randomized robust price optimization. *Management Science*, 2026.
- Vishal Gupta, Michael Huang, and Paat Rusmevichientong. Debiasing in-sample policy performance for small-data, large-scale optimization. *Operations Research*, 72(2):848–870, 2024.
- Dan A Iancu and Nikolaos Trichakis. Pareto efficiency in robust optimization. *Management Science*, 60(1):130–147, 2014.
- Garud N Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.
- Qingwei Jin, Daniel Zhuoyu Long, Yu Sun, and Bin Hu. Distributionally robust discrete choice model and assortment optimization. *Available at SSRN 4045001*, 2022.
- Saravanan Kesavan and Tarun Kushwaha. Field experiment on the profit implications of merchants’ discretionary power to override data-driven decision-making tools. *Management Science*, 66(11):5182–5190, 2020.
- Çağrı Koçyiğit, Garud Iyengar, Daniel Kuhn, and Wolfram Wiesemann. Distributionally robust mechanism design. *Management Science*, 66(1):159–189, 2020.
- A Gürhan Kök, Marshall L Fisher, and Ramnath Vaidyanathan. Assortment planning: Review of literature and industry practice. *Retail supply chain management: Quantitative models and empirical studies*, pages 99–153, 2008.
- Panos Kouvelis and Gang Yu. *Robust Discrete Optimization and Its Applications*, volume 14. Springer Science & Business Media, 1996.

- Daniel Kuhn, Soroosh Shafiee, and Wolfram Wiesemann. Distributionally robust optimization. *Acta Numerica*, 34:579–804, 2025.
- Daniel Zhuoyu Long, Melvyn Sim, and Minglong Zhou. Robust satisficing. *Operations Research*, 71(1):61–82, 2023.
- Zhiyuan Lou, Zhi Chen, Melvyn Sim, Jingui Xie, and Peng Xiong. Data-driven uncertainty sets. *Available at SSRN 4890089*, 2026.
- Peyman Mohajerin Esfahani, Soroosh Shafieezadeh-Abadeh, Grani A Hanasusanto, and Daniel Kuhn. Data-driven inverse optimization with imperfect information. *Mathematical Programming*, 167(1):191–234, 2018.
- Arnab Nilim and Laurent El Ghaoui. Robust control of markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.
- Yanqiu Ruan, Xiaobo Li, Karthyek Murthy, and Karthik Natarajan. A nonparametric approach with marginals for modeling consumer choice. *Management Science*, 2026.
- Walter Rudin. *Real and complex analysis*. McGraw-Hill, Inc., 1987.
- Paat Rusmevichientong and Huseyin Topaloglu. Robust assortment optimization in revenue management under the multinomial logit choice model. *Operations Research*, 60(4):865–882, 2012.
- Soroosh Shafiee, Liviu Aolaritei, Florian Dörfler, and Daniel Kuhn. Nash equilibria, regularization, and computation in optimal transport-based distributionally robust optimization. *Operations Research*, (3):1689–1709, 2026.
- Melvyn Sim, Qinshen Tang, Minglong Zhou, and Taozeng Zhu. The analytics of robust satisficing: predict, optimize, satisfy, then fortify. *Operations Research*, 73(5):2708–2728, 2025.
- Maurice Sion. On general minimax theorems. *Pacific J. Math.*, 8(4):171–176, 1958.
- James E Smith and Robert L Winkler. The optimizer’s curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322, 2006.
- Bradley Sturt. The value of robust assortment optimization under ranking-based choice models. *Management Science*, 71(5):4246–4265, 2025.
- Frode Terkelsen. Some minimax theorems. *Mathematica Scandinavica*, 31(2):405–413, 1972.
- Bart PG Van Parys, Peyman Mohajerin Esfahani, and Daniel Kuhn. From data to decisions: Distributionally robust optimization is optimal. *Management Science*, 67(6):3387–3402, 2021.
- Irina Wang, Bart Van Parys, and Bartolomeo Stellato. Learning decision-focused uncertainty sets in robust optimization. *arXiv preprint arXiv:2305.19225*, 2023.
- Zhengchao Wang, Heikki Peura, and Wolfram Wiesemann. Randomized assortment optimization. *Operations Research*, 72(5):2042–2060, 2024.
- Ningji Wei and Peter Zhang. Adjustability in robust linear optimization. *Mathematical Programming*, 208(1):581–628, 2024.
- Sikun Xu, Raphael Thomadsen, and Dennis Zhang. The winner’s curse in data-driven decision making: Evidence and solutions. *Available at SSRN 5930537*, 2025.
- Jianzhe Zhen, Daniel Kuhn, and Wolfram Wiesemann. A unified theory of robust and distributionally robust optimization via the primal-worst-equals-dual-best principle. *Operations Research*, 73(2):862–878, 2025.

A Review of Topology

Our analysis in §§2 and 3 utilizes basic facts about topology, which we review here.

A.1 Compact sets and semicontinuity

Recall from §2.1 that $\mathcal{X}$ and $\mathcal{U}$ are assumed throughout the paper to be compact nonempty sets (this is stated formally as Assumption 1 in §2.4). It follows from the definition of lower and upper semicontinuity (Rudin 1987, Definition 2.8) that if $\psi : \mathcal{U} \rightarrow \mathbb{R}$ is lower semicontinuous and $\phi : \mathcal{X} \rightarrow \mathbb{R}$ is upper semicontinuous, then the sets

$$\{\boldsymbol{\theta} \in \mathcal{U} : \psi(\boldsymbol{\theta}) \leq \eta\} \text{ and } \{\mathbf{x} \in \mathcal{X} : \phi(\mathbf{x}) \geq \alpha\}$$

are closed for all $\eta, \alpha \in \mathbb{R}$. Because the above sets are closed subsets of compact sets $\mathcal{U}$ and $\mathcal{X}$, we also have that the above sets are compact sets for all $\eta, \alpha \in \mathbb{R}$ (Rudin 1987, Theorem 2.4).

We will utilize several additional properties of lower and upper semicontinuous functions. First, we will use the fact that if (ψ_i) are lower semicontinuous functions, then $\max_i \psi_i$ is a lower semicontinuous function (similarly, $\min_i \phi_i$ is upper semicontinuous if (ϕ_i) are upper semicontinuous functions). This, for example, implies that the reduced uncertainty set from §2, defined as

$$\mathcal{U}_\eta \triangleq \left\{ \boldsymbol{\theta} \in \mathcal{U} : \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) \leq \eta \right\},$$

is a compact set for all $\eta \in \mathbb{R}$. Second, we will utilize semicontinuity in the context of the Weierstrass extreme value theorem, which says that the maximum of a bounded upper semicontinuous function over a compact set is attained, and similarly for the minimum of a bounded lower semicontinuous function over a compact set. This implies that the inner and outer problems in (4), (5) and (6) attain their optimums.

Finally, for a general topological space, a set is compact if every open cover of the set has a finite subcover (Rudin 1987, Definition 2.3). From this definition of compact sets, we obtain the following proposition, which we will use in Appendix A.2.

Proposition 10. *Let $\{\mathcal{X}_i\}_{i \in I}$ be a collection of closed sets that satisfy $\mathcal{X}_i \subseteq \mathcal{X}$ for all $i \in I$. If $\bigcap_{i \in I} \mathcal{X}_i = \emptyset$, then there exists a finite collection $\{i_1, \dots, i_K\} \subseteq I$ that satisfies $\mathcal{X}_{i_1} \cap \dots \cap \mathcal{X}_{i_K} = \emptyset$.*

Proof. Define $\mathcal{Y}_i \triangleq \mathcal{X} \setminus \mathcal{X}_i$ for all $i \in I$. It follows from the fact that each $\mathcal{X}_i$ is a closed set that each $\mathcal{Y}_i$ is an open set, and it follows from the fact that $\bigcap_{i \in I} \mathcal{X}_i = \emptyset$ that $\bigcup_{i \in I} \mathcal{Y}_i = \mathcal{X}$. Because $(\mathcal{Y}_i : i \in I)$ is a collection of open sets that forms a cover of $\mathcal{X}$, and because $\mathcal{X}$ is a compact set, there must exist a finite collection $\{i_1, \dots, i_K\}$ that satisfies $\mathcal{Y}_{i_1} \cup \dots \cup \mathcal{Y}_{i_K} = \mathcal{X}$, which implies that $\mathcal{X}_{i_1} \cap \dots \cap \mathcal{X}_{i_K} = \emptyset$. $\square$

A.2 Omitted details from the proof of Theorem 1

Let $\eta = \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta})$ and $\mathcal{U}^* = \arg \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \boldsymbol{\theta}) = \mathcal{U}_\eta$. For each $\boldsymbol{\theta} \in \mathcal{U}^*$ define

$$\mathcal{X}^\boldsymbol{\theta} \triangleq \{\mathbf{x} \in \mathcal{X} : f(\mathbf{x}, \boldsymbol{\theta}) \geq \eta\} \tag{26}$$

It follows from the fact that $\mathbf{x} \mapsto f(\mathbf{x}, \boldsymbol{\theta})$ is upper semicontinuous and from Appendix A that $\mathcal{X}^\boldsymbol{\theta}$ is a closed set for all $\boldsymbol{\theta} \in \mathcal{U}^*$. Furthermore, it follows from the definition of η that for all $\boldsymbol{\theta} \in \mathcal{U}^*$ and all $\mathbf{x} \in \mathcal{X}$, we have

$$f(\mathbf{x}, \boldsymbol{\theta}) \leq \max_{\hat{\mathbf{x}} \in \mathcal{X}} f(\hat{\mathbf{x}}, \boldsymbol{\theta}) = \eta, \tag{27}$$

where the inequality follows from algebra and the equality follows from the fact that $\boldsymbol{\theta} \in \mathcal{U}^*$. Combining lines (26) and (27), we have shown for all $\boldsymbol{\theta} \in \mathcal{U}^*$ that

$$\mathcal{X}^\theta = \{\mathbf{x} \in \mathcal{X} : f(\mathbf{x}, \boldsymbol{\theta}) = \eta\}.$$

It is supposed at the beginning of Case 2 in the proof of Theorem 1 from §3.2 that there does not exist an $\hat{\mathbf{x}} \in \mathcal{X}$ that satisfies $f(\hat{\mathbf{x}}, \boldsymbol{\theta}) = \eta$ for all $\boldsymbol{\theta} \in \mathcal{U}^*$. Using the above notation, this is equivalent to supposing that

$$\bigcap_{\boldsymbol{\theta} \in \mathcal{U}^*} \mathcal{X}^\theta = \emptyset.$$

We observe that $\{\mathcal{X}^\theta\}_{\boldsymbol{\theta} \in \mathcal{U}^*}$ is a collection of closed sets that satisfy $\mathcal{X}^\theta \subseteq \mathcal{X}$ for all $\boldsymbol{\theta} \in \mathcal{U}^*$. Therefore, it follows from Proposition 10 from Appendix A that there exists a finite collection $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K\} \subseteq \mathcal{U}^*$ that satisfies

$$\bigcap_{k=1}^K \mathcal{X}^{\boldsymbol{\theta}_k} = \emptyset,$$

which completes the omitted details from the beginning of Case 2 in the proof of Theorem 1.

B Proof of Proposition 1

Let $\eta = \min_{\boldsymbol{\theta} \in \mathcal{U}} \max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \boldsymbol{\theta})$. The proof that $v_{\mathbf{x}}(\eta)$ is nonincreasing follows from the fact that $\mathcal{U}_{\eta_1} \subseteq \mathcal{U}_{\eta_2}$ for all $\eta_1 \leq \eta_2$.

To prove that $v_{\mathbf{x}}(\cdot)$ is convex, consider any $\eta_1 < \eta_2$ that satisfy $\eta_1, \eta_2 \in [\eta, \infty)$ and $\lambda \in (0, 1)$. It follows from Assumption 2 and Proposition 2 that the uncertainty set $\mathcal{U}_\eta$ is convex and nonempty for all $\eta \in [\eta, \infty)$. Consider any optimal solutions $\boldsymbol{\theta}_i \in \arg \min_{\boldsymbol{\theta} \in \mathcal{U}_{\eta_i}} f(\mathbf{x}, \boldsymbol{\theta})$ for $i \in \{1, 2\}$, and define $\hat{\eta} = \lambda\eta_1 + (1 - \lambda)\eta_2$ and $\hat{\boldsymbol{\theta}} := \lambda\boldsymbol{\theta}_1 + (1 - \lambda)\boldsymbol{\theta}_2$. It follows from Assumption 2 that $\hat{\boldsymbol{\theta}} \in \mathcal{U}$. Moreover, we have

$$\max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \hat{\boldsymbol{\theta}}) \leq \lambda \max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \boldsymbol{\theta}_1) + (1 - \lambda) \max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \boldsymbol{\theta}_2) \leq \lambda\eta_1 + (1 - \lambda)\eta_2 = \hat{\eta}$$

where the first inequality follows from Assumption 2 and the second follows from the fact that $\boldsymbol{\theta}_i \in \mathcal{U}_{\eta_i}$ for $i \in \{1, 2\}$. We thus conclude that $\hat{\boldsymbol{\theta}} \in \mathcal{U}_{\hat{\eta}}$. Therefore,

$$v_{\mathbf{x}}(\hat{\eta}) \leq f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \leq \lambda f(\mathbf{x}, \boldsymbol{\theta}_1) + (1 - \lambda) f(\mathbf{x}, \boldsymbol{\theta}_2) \leq \lambda v_{\mathbf{x}}(\eta_1) + (1 - \lambda) v_{\mathbf{x}}(\eta_2)$$

where the inequality follows from the fact that $\hat{\boldsymbol{\theta}} \in \mathcal{U}_{\hat{\eta}}$, the second inequality follows from Assumption 2, and the third inequality follows from the definition of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$. We have thus shown that $v_{\mathbf{x}}(\cdot)$ is convex.

We conclude by showing that $v_{\mathbf{x}}(\cdot)$ is continuous on $[\eta, \infty)$. It follows from the fact that $v_{\mathbf{x}}(\cdot)$ is convex that $v_{\mathbf{x}}(\cdot)$ is continuous on (η, ∞) . To show that the continuity extends to $\eta = \eta$, choose any sequence $\eta_k \downarrow \eta$ and let $\boldsymbol{\theta}_k \in \arg \min_{\boldsymbol{\theta} \in \mathcal{U}_{\eta_k}} f(\mathbf{x}, \boldsymbol{\theta})$ for all k . It follows from compactness of $\mathcal{U}$ (Assumption 1) that the net $(\boldsymbol{\theta}_k)$ has a convergent subnet $(\boldsymbol{\theta}_{k_\alpha})_{\alpha \in A}$ with $\boldsymbol{\theta}_{k_\alpha} \rightarrow \hat{\boldsymbol{\theta}}$ for some $\hat{\boldsymbol{\theta}} \in \mathcal{U}$. It thus follows from the lower semicontinuity of $\max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \cdot)$ (Assumption 1) that

$$\max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \hat{\boldsymbol{\theta}}) \leq \liminf_{\alpha} \max_{\mathbf{y} \in \mathcal{X}} f(\mathbf{y}, \boldsymbol{\theta}_{k_\alpha}) \leq \lim_{\alpha} \eta_{k_\alpha} = \eta \quad (28)$$

where the first inequality follows from lower semicontinuity, the second inequality follows from the fact that $\boldsymbol{\theta}_{k_\alpha} \in \mathcal{U}_{\eta_{k_\alpha}}$ for all α , and the equality holds because every subnet of (η_k) converges to η . Moreover, we have

$$f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \leq \liminf_{\alpha} f(\mathbf{x}, \boldsymbol{\theta}_{k_\alpha}) = \lim_{\alpha} v_{\mathbf{x}}(\eta_{k_\alpha}) \quad (29)$$

where the first inequality follows from the fact that $f(\mathbf{x}, \cdot)$ is lower semicontinuous (Assumption 1), and the equality follows from the definition of θ_{k_α} and the fact that every subnet of $(v_{\mathbf{x}}(\eta_k))$ converges to the same limit (which exists because $v_{\mathbf{x}}(\cdot)$ is non-increasing, is bounded because of Assumption 1, and $\eta_k \downarrow \eta$). We thus have

$$v_{\mathbf{x}}(\eta) \leq f(\mathbf{x}, \hat{\theta}) \leq \lim_{\alpha} v_{\mathbf{x}}(\eta_{k_\alpha}) \quad (30)$$

where the first inequality follows from the fact that $\hat{\theta} \in \mathcal{U}_\eta$ as shown in (28) and the second inequality is (29). Since we established previously that $v_{\mathbf{x}}(\cdot)$ is nonincreasing, it follows from the fact that $f(\mathbf{x}, \cdot)$ is bounded (Assumption 1) that $v_{\mathbf{x}}(\eta) \geq \lim_{k \rightarrow \infty} v_{\mathbf{x}}(\eta_k)$. Combining that inequality with (30), we conclude that $v_{\mathbf{x}}(\eta) = \lim_{k \rightarrow \infty} v_{\mathbf{x}}(\eta_k)$, which completes the proof that $v_{\mathbf{x}}(\cdot)$ is continuous on $[\eta, \infty)$.

C Omitted Proofs From §5.2

Proof of Theorem 2. To show the first direction, suppose there exists a vector $\tilde{\mathbf{z}}$ that is an optimal solution of $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$ and satisfies $\text{supp}(\tilde{\mathbf{z}}) \cap I^* \neq \emptyset$ for every optimal solution I^* of the non-robust problem (17). In that case, we observe for every optimal solution I^* of the non-robust problem (17) that

$$\sum_{j \in I^*} (\hat{c}_j + d_j \tilde{z}_j) \geq \sum_{j \in I^*} \hat{c}_j + \sum_{j \in \text{supp}(\tilde{\mathbf{z}}) \cap I^*} d_j \tilde{z}_j > \sum_{j \in I^*} \hat{c}_j = \min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j,$$

Indeed, the first inequality follows from algebra. The strict inequality follows from the fact that $\text{supp}(\tilde{\mathbf{z}}) \cap I^* \neq \emptyset$, the fact that $\mathbf{d} > \mathbf{0}$, and the definition of $\text{supp}(\tilde{\mathbf{z}})$. The equality follows from the definition of I^* . Moreover, we observe for every non-optimal solution I' of the non-robust problem (17) that

$$\sum_{j \in I'} (\hat{c}_j + d_j \tilde{z}_j) \geq \sum_{j \in I'} \hat{c}_j > \min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j$$

where the inequality follows from the fact that $\mathbf{d} > \mathbf{0}$ and the strict inequality follows from the fact that I' is not an optimal solution for (17). We have thus shown in all cases that

$$\min_{I \in \mathcal{I}} \sum_{j \in I} (\hat{c}_j + d_j \tilde{z}_j) > \min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j.$$

We conclude that $\tilde{\mathbf{z}} \in \mathcal{Z}_\eta$ for all $\eta \in \mathbb{H} \cap (\min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j, \min_{I \in \mathcal{I}} \sum_{j \in I} (\hat{c}_j + d_j \tilde{z}_j)) \neq \emptyset$, and so it follows from the fact that $\tilde{\mathbf{z}}$ is an optimal solution of $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$ that

$$\max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j) = \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$$

To show the other direction, consider any arbitrary vector $\tilde{\mathbf{z}}$ that is an optimal solution of $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$, and suppose that there exists an optimal solution I^* of the non-robust problem (17) that satisfies $\text{supp}(\tilde{\mathbf{z}}) \cap I^* = \emptyset$. We observe for each $\eta \in \mathbb{H}$ that

$$\mathcal{Z}_\eta \subseteq \left\{ \mathbf{z} \in \mathcal{Z} : \sum_{j \in I^*} (\hat{c}_j + d_j z_j) \geq \eta \right\} \subseteq \left\{ \mathbf{z} \in \mathcal{Z} : \sum_{j \in I^*} (\hat{c}_j + d_j z_j) > \sum_{j \in I^*} \hat{c}_j \right\} = \left\{ \mathbf{z} \in \mathcal{Z} : \sum_{j \in I^*} d_j z_j > 0 \right\},$$

where the first inclusion follows from the definition of $\mathcal{Z}_\eta$, the second inclusion follows from the fact that $\eta \in \mathbb{H}$ and that I^* is an optimal solution for (17), and the equality follows from the fact that $\mathbf{d} > \mathbf{0}$ and $\mathcal{Z} \subseteq [0, 1]^n$. Moreover, it follows from the fact that $\text{supp}(\tilde{\mathbf{z}}) \cap I^* = \emptyset$ that $\sum_{j \in I^*} d_j z_j = 0$, which implies that $\tilde{\mathbf{z}} \notin \mathcal{Z}_\eta$. Since the optimal solution $\tilde{\mathbf{z}}$ of $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$ was chosen arbitrarily, we conclude that if for every optimal solution $\tilde{\mathbf{z}}$ of $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$ there exists an optimal solution I^* of the non-robust

problem (17) that satisfies $\text{supp}(\tilde{\mathbf{z}}) \cap I^* = \emptyset$, then

$$\max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j) < \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in \tilde{I}} (\hat{c}_j + d_j z_j)$$

□

Proof of Corollary 3. Let I^{RO} denote an optimal solution of the robust optimization problem (16).

Suppose that I^{RO} satisfies condition (a), meaning that $|I^{\text{RO}}| \geq \Gamma$ and there exists an optimal solution I^* for the non-robust problem (17) that satisfies $I^{\text{RO}} \cap I^* = \emptyset$. Consider any arbitrary $\tilde{\mathbf{z}} \in \arg \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j)$. It follows from the fact that $\mathbf{d} > \mathbf{0}$, the fact that Γ is an integer, the fact that $|I^{\text{RO}}| \geq \Gamma$, and the definition of the budget uncertainty set $\mathcal{Z} = \{\mathbf{z} \in [0, 1]^n : \sum_{j=1}^n z_j \leq \Gamma\}$ that $\tilde{\mathbf{z}}$ can satisfy $\tilde{z}_j > 0$ only if $j \in I^{\text{RO}}$. This implies that $\text{supp}(\tilde{\mathbf{z}}) \subseteq I^{\text{RO}}$, and so it follows from the fact that $I^{\text{RO}} \cap I^* = \emptyset$ that $\text{supp}(\tilde{\mathbf{z}}) \cap I^* = \emptyset$. Since $\tilde{\mathbf{z}}$ was chosen arbitrarily, we conclude for all $\tilde{\mathbf{z}} \in \arg \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j)$ that $\text{supp}(\tilde{\mathbf{z}}) \cap I^* = \emptyset$, and so Theorem 2 implies that I^{RO} satisfies (19).

Alternatively, suppose that I^{RO} satisfies condition (b). Let $\tilde{\mathbf{z}}$ denote the unique optimal solution for $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j)$, and let I^* denote the optimal solution of the non-robust problem (17) that satisfies $\text{supp}(\tilde{\mathbf{z}}) \cap I^* = \emptyset$. Since $\tilde{\mathbf{z}}$ is the unique optimal solution, it follows immediately from Theorem 2 that (19) is satisfied.

In summary, we have shown that if condition (a) or (b) is satisfied, then I^{RO} satisfies (19). We thus conclude for each $\eta \in \mathbb{H}$ that

$$\min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I} (\hat{c}_j + d_j z_j) = \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) > \max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) \geq \min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in I} (\hat{c}_j + d_j z_j)$$

where the equality holds because I^{RO} is an optimal solution for the robust optimization problem (16), the first inequality follows from the fact that I^{RO} satisfies (19) and the fact that $\eta \in \mathbb{H}$, and the second inequality holds because I^{RO} is a feasible but possibly suboptimal solution for the reduced robust optimization problem (13). □

Proof of Proposition 4. It follows from Lemma 1 that if $\eta = \min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j$, then $\mathcal{Z} = \mathcal{Z}_\eta$. Moreover, if either of the sufficient conditions (a) or (b) from Corollary 3 is satisfied, then it follows from Theorem 2 that

$$\max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) < \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) \quad \forall \eta \in \mathbb{H}$$

Combining the above reasoning with the definitions of $\mathcal{Z}_\eta$ from (14) and $\mathbb{H}$ from (18), we conclude that

$$\begin{aligned} \arg \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) &= \arg \max_{\mathbf{z} \in \mathcal{Z} \setminus \mathcal{Z}_\eta} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) \quad \forall \eta \in \mathbb{H} \\ &= \bigcap_{\eta \in \mathbb{H}} \arg \max_{\mathbf{z} \in \mathcal{Z} \setminus \mathcal{Z}_\eta} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) \\ &\subseteq \left\{ \mathbf{z} \in \mathcal{Z} : \min_{I \in \mathcal{I}} \sum_{j \in I} (\hat{c}_j + d_j z_j) = \min_{I \in \mathcal{I}} \sum_{j \in I} \hat{c}_j \right\} \end{aligned}$$

which completes the proof of Proposition 4. □

Proof of Lemma 2. We observe that the reduced uncertainty set can be rewritten as

$$\begin{aligned}\mathcal{Z}_\eta &= \left\{ \mathbf{z} \in \mathcal{Z} : \sum_{j \in I} (\hat{c}_j + d_j z_j) \geq \eta \ \forall I \in \mathcal{I} \right\} = \left\{ \mathbf{z} \in \mathcal{Z} : \sum_{j \in I} d_j z_j \geq \eta - \sum_{j \in I} \hat{c}_j \ \forall I \in \mathcal{I} \right\} \\ &= \left\{ \mathbf{z} \in \mathcal{Z} : \sum_{j \in I} d_j z_j \geq \eta - \sum_{j \in I} \hat{c}_j \ \forall I \in \mathcal{I}_\eta \right\}.\end{aligned}$$

The first equality follows from the definition of $\mathcal{Z}_\eta$. The second equality follows from algebra. The third equality follows from the fact that $\mathcal{Z} \subseteq [0, 1]^n$ and $\mathbf{d} > \mathbf{0}$, and the fact that $\eta - \sum_{j \in I} \hat{c}_j > 0$ if and only if $I \in \mathcal{I}_\eta$. $\square$

Proof of Proposition 5. Our proof will make use of the following claim.

Claim 1. *If $\eta \in \mathbb{H}$ and $\text{HITTINGSETNUM}(\mathcal{I}_\eta) > \Gamma$, then $\mathcal{Z}_\eta \cap \{0, 1\}^n = \emptyset$.*

Proof of Claim 1. Let $\eta \in \mathbb{H}$. We observe that

$$\mathcal{Z}_\eta \cap \{0, 1\}^n = \left\{ \mathbf{z} \in \mathcal{Z} \cap \{0, 1\}^n : \sum_{j \in I} d_j z_j \geq \eta - \sum_{j \in I} \hat{c}_j \ \forall I \in \mathcal{I}_\eta \right\} \subseteq \left\{ \mathbf{z} \in \{0, 1\}^n : \begin{array}{l} \sum_{j \in I} z_j > 0 \ \forall I \in \mathcal{I}_\eta \\ \sum_{j=1}^n z_j \leq \Gamma \end{array} \right\} \quad (31)$$

where the equality follows from Lemma 2 and the set inclusion follows from the definition of $\mathcal{I}_\eta$, the fact that $\mathbf{d} > \mathbf{0}$, and the definition of the uncertainty set $\mathcal{Z}$ from line (12). If $\text{HITTINGSETNUM}(\mathcal{I}_\eta) > \Gamma$, then it follows from line (21) that the right-most set in line (31) is empty. This completes the proof of Claim 1. $\square$

Equipped with the above intermediary claim, our proof of Proposition 5 is as follows. Let $\eta \in \mathbb{H}$ satisfy $\text{HITTINGSETNUM}(\mathcal{I}_\eta) > \Gamma$, and suppose that I^{RO} is an optimal solution of the robust optimization problem (16) for which $\max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j)$ has a unique solution. Let $\tilde{\mathbf{z}} \in \arg \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j)$ denote the unique worst-case realization. Because it is unique, and because Γ is an integer, $\tilde{\mathbf{z}}$ must be an extreme point of $\mathcal{Z}$ and thus must satisfy $\tilde{\mathbf{z}} \in \{0, 1\}^n$. It thus follows from Claim 1 that $\tilde{\mathbf{z}} \notin \mathcal{Z}_\eta$, which implies that

$$\max_{\mathbf{z} \in \mathcal{Z}_\eta} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) < \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I^{\text{RO}}} (\hat{c}_j + d_j z_j) = \min_{I \in \mathcal{I}} \max_{\mathbf{z} \in \mathcal{Z}} \sum_{j \in I} (\hat{c}_j + d_j z_j).$$

This completes the proof of Proposition 5. $\square$

D Extended Numerical Results from §4.3

Figure 3 is an expanded version of Figure 1 from §4.3 to show the nominal curves for a larger collection of assortments. Specifically, Figure 3 shows the nominal curves for all assortments $S \in \mathcal{S}$ that satisfy $4 \in S$. It is sufficient to consider only the assortments that satisfy $4 \in S$ because adding the most expensive product to an assortment will not decrease its performance. Figure 3 shows that the assortment $\{0, 2, 3, 4\}$ is optimal for the reduced robust optimization problem for all $\$23.875 \leq \eta \leq \tilde{\eta}$ where $\tilde{\eta} \approx \$33.778$. The y-axis in Figure 3 is clipped at \$15.

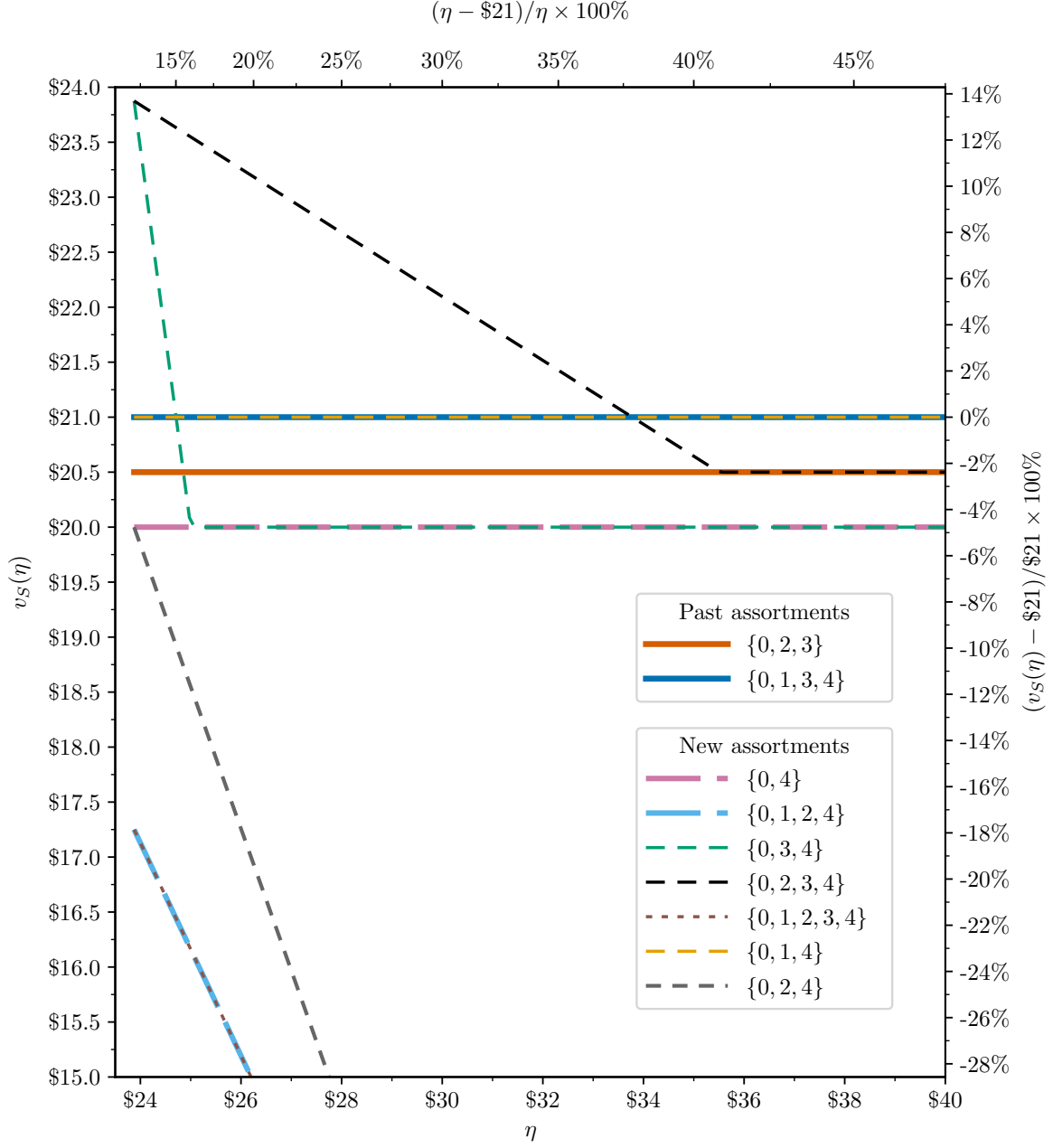


Figure 3: Additional nominal curves from numerical example in §4.3.

E Omitted Proofs from §6.1.1

In this appendix, we consider the setting described in §6.1.1 where $\mathcal{X} = \{\mathbf{x} \in \{0, 1\}^n : \mathbf{Ax} \leq \mathbf{b}\} \neq \emptyset$ satisfies $\text{conv}(\mathcal{X}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}\}$, the uncertainty set is a bounded polyhedron $\mathcal{U} = \{\boldsymbol{\theta} \in \mathbb{R}^p : \mathbf{D}\boldsymbol{\theta} \geq \mathbf{g}\} \neq \emptyset$, and the dimensions are $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{D} \in \mathbb{R}^{q \times p}$, $\mathbf{g} \in \mathbb{R}^q$, $\mathbf{U}, \mathbf{L} \in \mathbb{R}^{n \times p}$, and $\mu, \lambda \in \mathbb{R}$. Our main reformulation steps are contained in the following Lemma 3. After presenting and proving Lemma 3, we present the proofs of Propositions 6 and 7.

Lemma 3. *If $\mathcal{U}_\eta$ is nonempty and $\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda > 0$ for all $\mathbf{x} \in \mathcal{X}$ and $\boldsymbol{\theta} \in \mathcal{U}$, then*

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \frac{\mathbf{x}^\top \mathbf{U}\boldsymbol{\theta} + \mu}{\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda} = \left[\begin{array}{ll} \underset{\mathbf{x} \in \{0,1\}^n, t \in \mathbb{R}, \boldsymbol{\psi} \in \mathbb{R}^q, \mathbf{y} \in \mathbb{R}^n, \sigma \in \mathbb{R}, \mathbf{z} \in \mathbb{R}^n}{\text{maximize}} & t \\ \text{subject to} & t\lambda - \mu \leq \mathbf{g}^\top \boldsymbol{\psi} - (\eta\lambda - \mu)\sigma \\ & \mathbf{D}^\top \boldsymbol{\psi} - (\mathbf{U} - \eta\mathbf{L})^\top \mathbf{y} = \mathbf{U}^\top \mathbf{x} - \mathbf{L}^\top \mathbf{z} \\ & \mathbf{z} = t\mathbf{x} \\ & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{Ay} \leq \mathbf{b}\sigma \\ & \boldsymbol{\psi} \geq \mathbf{0}, \sigma \geq 0 \end{array} \right]$$

Proof. Our proof follows from applying strong duality once to reformulate the reduced uncertainty set $\mathcal{U}_\eta$ as a polyhedron with a polynomial number of constraints, followed by dualizing the inner problem of the reduced robust optimization problem (23). Indeed, we observe for each $\boldsymbol{\theta} \in \mathcal{U}$ that

$$\begin{aligned} \max_{\mathbf{x} \in \mathcal{X}} \frac{\mathbf{x}^\top \mathbf{U}\boldsymbol{\theta} + \mu}{\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda} \leq \eta &\iff \max_{\mathbf{x} \in \mathcal{X}} \{\mathbf{x}^\top (\mathbf{U} - \eta\mathbf{L})\boldsymbol{\theta}\} \leq \eta\lambda - \mu \\ &\iff \max_{\mathbf{x} \in \text{conv}(\mathcal{X})} \{\mathbf{x}^\top (\mathbf{U} - \eta\mathbf{L})\boldsymbol{\theta}\} \leq \eta\lambda - \mu \\ &\iff \max_{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}} \{\mathbf{x}^\top (\mathbf{U} - \eta\mathbf{L})\boldsymbol{\theta}\} \leq \eta\lambda - \mu \\ &\iff \exists \boldsymbol{\gamma} \geq \mathbf{0} \text{ such that } \mathbf{b}^\top \boldsymbol{\gamma} \leq \eta\lambda - \mu \text{ and } \mathbf{A}^\top \boldsymbol{\gamma} = (\mathbf{U} - \eta\mathbf{L})\boldsymbol{\theta} \end{aligned}$$

where the first line follows from algebra and the assumptions that $\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda > 0$ for all $\mathbf{x} \in \mathcal{X}$ and $\boldsymbol{\theta} \in \mathcal{U}$, the second line is the fundamental theorem of linear programming, the third line follows from the earlier assumption that $\text{conv}(\mathcal{X}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}\}$, and the fourth line follows from strong duality. The above implies that the reduced uncertainty set admits a polynomial-size extended formulation, i.e.,

$$\mathcal{U}_\eta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^p : \begin{array}{l} \mathbf{D}\boldsymbol{\theta} \geq \mathbf{g} \\ \max_{\mathbf{x} \in \mathcal{X}} \frac{\mathbf{x}^\top \mathbf{U}\boldsymbol{\theta} + \mu}{\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda} \leq \eta \end{array} \right\} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^p : \begin{array}{l} \mathbf{D}\boldsymbol{\theta} \geq \mathbf{g} \\ \exists \boldsymbol{\gamma} \geq \mathbf{0} \text{ such that } \mathbf{b}^\top \boldsymbol{\gamma} \leq \eta\lambda - \mu \text{ and } \mathbf{A}^\top \boldsymbol{\gamma} = (\mathbf{U} - \eta\mathbf{L})\boldsymbol{\theta} \end{array} \right\}.$$

We observe that

$$\begin{aligned} \max_{\mathbf{x} \in \mathcal{X}} \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \frac{\mathbf{x}^\top \mathbf{U}\boldsymbol{\theta} + \mu}{\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda} &= \left[\begin{array}{ll} \underset{\mathbf{x} \in \mathcal{X}, t \in \mathbb{R}}{\text{maximize}} & t \\ \text{subject to} & t \leq \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \frac{\mathbf{x}^\top \mathbf{U}\boldsymbol{\theta} + \mu}{\mathbf{x}^\top \mathbf{L}\boldsymbol{\theta} + \lambda} \end{array} \right] \\ &= \left[\begin{array}{ll} \underset{\mathbf{x} \in \mathcal{X}, t \in \mathbb{R}}{\text{maximize}} & t \\ \text{subject to} & t\lambda - \mu \leq \min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \mathbf{x}^\top (\mathbf{U} - t\mathbf{L})\boldsymbol{\theta} \end{array} \right] \end{aligned} \quad (32)$$

Using the polyhedral representation of $\mathcal{U}_\eta$ developed above, we observe for each $\mathbf{x} \in \mathcal{X}$ and $t \in \mathbb{R}$ that

$$\min_{\boldsymbol{\theta} \in \mathcal{U}_\eta} \mathbf{x}^\top (\mathbf{U} - t\mathbf{L}) \boldsymbol{\theta} = \begin{bmatrix} \underset{\boldsymbol{\psi} \in \mathbb{R}^q, \mathbf{y} \in \mathbb{R}^n, \sigma \in \mathbb{R}}{\text{maximize}} & \mathbf{g}^\top \boldsymbol{\psi} - (\eta\lambda - \mu)\sigma \\ \text{subject to} & \mathbf{D}^\top \boldsymbol{\psi} - (\mathbf{U} - \eta\mathbf{L})^\top \mathbf{y} = (\mathbf{U} - t\mathbf{L})^\top \mathbf{x}, \\ & \mathbf{A}\mathbf{y} \leq \mathbf{b}\sigma \\ & \boldsymbol{\psi} \geq \mathbf{0}, \sigma \geq 0 \end{bmatrix} \quad (33)$$

Combining (32) and (33) and letting $\mathbf{z} = t\mathbf{x}$ denote the bilinear term, we obtain the desired result. $\square$

Proof of Proposition 6. In the special case of $n = p$, $\mathbf{L} = \mathbf{0}$, $\mathbf{U}$ is the identity matrix, $\lambda = 1$, and $\mu = 0$, the optimization problem from Lemma 3 becomes

$$\begin{bmatrix} \underset{\mathbf{x} \in \{0,1\}^n, \boldsymbol{\psi} \in \mathbb{R}^q, \mathbf{y} \in \mathbb{R}^n, \sigma \in \mathbb{R}}{\text{maximize}} & \mathbf{g}^\top \boldsymbol{\psi} - \eta\sigma \\ \text{subject to} & \mathbf{D}^\top \boldsymbol{\psi} - \mathbf{y} = \mathbf{x} \\ & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{A}\mathbf{y} \leq \mathbf{b}\sigma \\ & \boldsymbol{\psi} \geq \mathbf{0}, \sigma \geq 0 \end{bmatrix}$$

Substituting $\mathbf{y} = \mathbf{D}^\top \boldsymbol{\psi} - \mathbf{x}$ and rearranging terms, we obtain the desired reformulation in Proposition 6. $\square$

Proof of Proposition 7. If the optimal value of (23) is in $[\underline{t}, \bar{t}]$ for $0 \leq t \leq \bar{t}$, then we observe that the bilinear term

$$\mathbf{z} = t\mathbf{x}$$

in the reformulation from Lemma 3 can be replaced with the McCormick envelopes

$$\begin{aligned} \underline{t} &\leq t \leq \bar{t} \\ \underline{t}x_i &\leq z_i \leq \bar{t}x_i & \forall i \in \{1, \dots, n\} \\ t - \bar{t}(1 - x_i) &\leq z_i \leq t - \underline{t}(1 - x_i) & \forall i \in \{1, \dots, n\} \end{aligned}$$

Substituting in those constraints, we obtain the desired reformulation in Proposition 7. $\square$