On the Sparsity of Optimal Linear Decision Rules in Robust Optimization

Haihao Lu
Booth School of Business
University of Chicago, haihao.lu@chicagobooth.edu

Bradley Sturt
Department of Information and Decision Sciences
University of Illinois at Chicago, bsturt@uic.edu

We consider the widely-studied class of production-inventory problems with box uncertainty sets from the seminal work of Ben-Tal et al. (2004) on linear decision rules in robust optimization. We prove that there always exists an optimal linear decision rule for this class of problems in which the number of nonzero parameters in the linear decision rule grows linearly in the number of time periods. This is the first result to prove that optimal linear decision rules are sparse in a widely-studied class of robust optimization problems with many time periods. Harnessing this sparsity guarantee, we introduce a novel reformulation technique that allows robust optimization problems such as production-inventory problems to be solved as a compact linear optimization problem when most of the parameters of the linear decision rules are forced to be equal to zero. We also develop an active set method for identifying the parameters of linear decision rules that are equal to zero at optimality. In numerical experiments on production-inventory problems with hundreds of time periods, we find that our novel reformulation technique coupled with the active set method yield more than a 32x speedup over state-of-the-art linear optimization solvers in computing linear decision rules that are within 1% of optimal. Our proofs and algorithms are based on a principled analysis of extreme points of linear optimization formulations, and we show that our sparsity guarantees extend to other widely-studied classes of robust optimization problems from the literature.

Key words: Linear programming; robust optimization; linear decision rules.

1. Introduction

Over the past two decades, robust optimization has emerged as a leading approach in operations research and management science for sequential decision-making under uncertainty. One of the major reasons for its popularity is computational: multi-period robust optimization problems are often amenable to efficient approximations in complex, real-world operational planning problems. The successful approximation techniques for multi-period robust optimization typically rely on restricting the control policies to a simple functional form, with the most popular restriction being to linear decision rules [4].

The success of linear decision rules in addressing real-world problems is very impressive. On the empirical side, linear decision rules have been found to exhibit strong performance in a myriad of high-stakes robust optimization applications such as disaster response, personalized healthcare, sustainable energy management, transportation routing, and many others [29, 25, 14, 28, 17]. On the theoretical side, a burgeoning literature has established that linear decision rules are provably optimal control policies in many classes of robust optimization problems [7, 9, 24, 21, 1, 29, 15, 20, 32]. The aforementioned papers all build upon the seminal work of Ben-Tal et al. [4], which showed that optimal linear decision rules for robust optimization can be computed in polynomial time even in "fairly complicated models with high-dimensional state spaces and many stages" [4, p. 374].

However, the impressive performance of linear decision rules comes at a price. Compared to simpler classes of control policies such as static decision rules, linear decision rules are represented using a significantly greater number of parameters. Specifically, the number of parameters in linear decision rules grows quadratically in the number of stages of the robust optimization problem, whereas the number of parameters for static decision rules grows linearly in the number of stages. Because the number of parameters for representing linear decision rules can be enormous, computing optimal linear decision rules can require "the solution of monolithic and often dense optimization problems" [19, p.814] which can exceed computer memory in robust optimization problems with as few as seventy-five stages [19, p.827]. This is problematic because real-world applications routinely have many hundreds of stages, particularly when a discrete-time robust optimization problem is approximating a continuous-time operational planning problem.

To get around this, a recent stream of research has advocated for imposing sparsity constraints onto linear decision rules in robust optimization problems. The driving insight behind this stream of research is that if many of the parameters of linear decision rules are forced to be equal to zero, then the problem of optimizing the remaining parameters can often be solved more efficiently. Numerical studies have shown that applying this insight can yield tremendous improvements in computation times with only small losses in performance in applications such as newsvendor networks in pharmaceutical supply chains [3, §4], unit commitment problems in power systems [26, §4.3], and location-transportation problems [2], and column-generation techniques have been proposed for identifying the subset of parameters of linear decision rules to set to zero [30]. The harnessing of sparsity thus stands as one of the most promising directions for efficiently computing optimal linear decision rules for the sizes of robust optimization problems that arise in industry.

Ultimately, the success of using sparsity to develop faster and more memory-efficient algorithms will hinge on whether robust optimization problems have optimal linear decision rules that are sparse. Indeed, if a robust optimization problem does not have optimal linear decision rules that are sparse, then deploying a sparsity-imposing algorithm can risk leading to control policies with unexpectedly and undesirably suboptimal performance. Moreover, if the number of zero parameters in optimal linear decision rules is not a significant proportion of the total parameters, then the benefits of developing sparsity-imposing algorithms can be limited, and research efforts may be better spent on developing alternative algorithmic techniques such as those based on primal-dual saddle point formulations or online convex optimization [27, 22, 6]. To the best of our knowledge, the fundamental question of whether a significant number of zero parameters can be expected in optimal linear decision rules in any class of robust optimization problems with many time periods has remained open.

In this paper, we resolve this open question by considering the widely-studied class of production-inventory problems from the seminal work of Ben-Tal et al. [4] on linear decision rules in robust optimization. This class of production-inventory problems served as the key illustration in [4] that optimal linear decision rules can be computed in polynomial time and provide excellent performance in realistic and complex robust inventory management problems. It has since become one of the most popular classes of problems in the robust optimization literature, serving as a test bed for the complex real-world applications in which linear decision rules are routinely used; see [12] and references therein. The class of production-inventory problems involves a firm which dynamically determines production quantities at multiple factories over a selling season, a single product with uncertain demand that lies in box uncertainty sets, and complex business constraints that link the inventory levels and production decisions across multiple periods of time.

Our main theoretical result of this paper (Theorem 1) establishes that the minimum number of nonzero parameters for optimal linear decision rules for the class of production-inventory problems grows *subquadratically* in the number of time periods. To state our sparsity guarantee more formally, we recall for any instance of such production-inventory problems with E factories and T time periods that the number of parameters for representing linear decision rules is equal to $\frac{1}{2}ET(T+1) = \mathcal{O}(ET^2)$. For this class of production-inventory problems, we prove in Theorem 1 that there always exists an optimal linear decision rule in which the number of nonzero parameters is at most equal to $2+8E+10T+6ET = \mathcal{O}(ET)$ whenever an optimal linear decision rule for the instance exists. In other words, although the number of parameters for representing linear decision rules grows *quadratically* in the number of time periods, Theorem 1 shows that the minimum number of nonzero parameters for representing optimal linear decision rules grows only *linearly* in the number of time periods. Our proof is based on a principled analysis of extreme points of linear optimization formulations, and we demonstrate via numerical experiments that our bounds are indicative of practice.

Our sparsity guarantee (Theorem 1) contributes to the robust optimization literature in three primary ways. First, our result proves for the first time that sparsity can be a fundamental property of optimal linear decision rules in widely-studied classes of robust optimization problems when the number of time periods is large. Our result thus provides the first theoretical foundation for the validity and potential impact of the growing stream of research that is based on harnessing sparsity to develop faster and more memory-efficient algorithms for computing optimal linear decision rules in real-world applications of robust optimization with huge numbers of time periods. Second, we provide evidence that our proof techniques can be extended to establish sparsity guarantees for other classes of robust optimization problems with many time periods. For example, we show that our proof techniques extend to establish sparsity guarantees for the widely-studied class of dynamic newsvendor problems with box uncertainty sets from [5, 9, 24], and our proof techniques also extend to production-inventory problems with multidimensional or non-box uncertainty sets. We hope the developments in this paper inspire and aid the robust optimization community in identifying more examples of robust optimization problems where the existence of sparse optimal solutions is guaranteed. Third, as we discuss below, our sparsity guarantee has immediate implications from the perspective of designing memory-efficient algorithms for computing linear decision rules in robust optimization problems with many time periods.

Harnessing our sparsity guarantees, we introduce a novel reformulation technique (see §4.2) that allows robust optimization problems such as the production-inventory problem to be solved very efficiently when most of the parameters of the linear decision rules are forced to be equal to zero. The key insight behind our novel reformulation technique is that imposing sparsity constraints onto linear decision rules can have a side effect of inducing redundancy in the constraints of the robust counterpart. By exploiting this redundancy in the constraints of the robust counterpart, our reformulation technique drastically decreases the number of auxiliary decision variables in the robust counterpart so that they match, up to a constant factor, the number of nonzero parameters of the linear decision rules (Proposition 1 in §4.2). It follows from Theorem 1 and Proposition 1 that if the nonzero parameters of an optimal linear decision rule for a production-inventory problem were known, then an optimal linear decision rule for the production-inventory problem can be computed by solving a linear optimization problem with only $\mathcal{O}(TE)$ decision variables and constraints.

While Theorem 1 does not specify which parameters of an optimal linear decision rules for the production-inventory problems will be nonzero, we show that the set of nonzero parameters can be found algorithmically to high or perfect accuracy. Specifically, we develop and have open-sourced a simple algorithm based on the active set method (see §4.3-§4.4) for identifying the parameters of sparse optimal linear decision rules that are nonzero. The algorithm consists of iterating over different 'active sets' of nonzero parameters, finding an linear decision rule that is optimal when the parameters of the linear decision rule that are not in the active set are constrained to be equal to zero, and then using shadow prices to update the active set. The algorithm is guaranteed to converge to an optimal linear decision rule after a finite number of iterations and overcomes the issue that we do not know a priori which parameters of optimal linear decision rules will be nonzero.

From a memory-efficiency standpoint, the algorithm can be terminated early with a suboptimal active set to avoid solving a large-scale linear optimization problem that exceeds computer memory.

We show through numerical experiments that our novel reformulation technique (§4.2) and active set method (§4.3-§4.4) can find high-quality linear decision rules in applications that are significantly larger than could be tackled by extant methods. For example, in production-inventory problems with $T = 240$ time periods and $E = 5$ factories, we show in §5.2 that our approach offers a 32x speedup over state-of-the-art linear optimization solvers in computing linear decision rules that are within 1% of optimal. Moreover, we show that our approach scales to applications with hundreds of time periods and dozens of state and decision variables in each time period, a regime which to our knowledge cannot be addressed by any extant algorithms from the robust optimization literature. Our open source implementation of the novel reformulation technique and active set method for computing linear decision rules in robust optimization can be accessed at `https://github.com/brad-sturt/LDRSolver.git`.

In summary, our main contributions in this paper are:

1. We prove that the minimum number of nonzero parameters for representing optimal linear decision rules can grow subquadratically in the number of time periods in widely-studied classes of robust optimization problems.
2. We introduce a novel reformulation technique that allows applications such as production-inventory problems to be solved very efficiently when sparsity constraints are imposed on linear decision rules.
3. We show that our novel reformulation technique can be combined with an active set method to significantly speed up and scale up the size of robust optimization problems that can be solved in practical computation times.

The rest of this paper is organized as follows. In §2, we present a background on linear decision rules in robust optimization and production-inventory problems. In §3, we state our sparsity guarantee for production-inventory problems and present an overview of our proof. In §4, we present a novel reformulation technique and an active set method for computing linear decision rules in robust optimization problems with huge numbers of time periods. In §5, we illustrate the practical implications of our main results via numerical experiments. In §6, we discuss extensions of our main sparsity guarantee to other classes of robust optimization problems. In §7, we conclude and discuss future directions of research. All technical proofs can be found in the supplemental appendices.

We let $\mathbb{R}$ denote the real numbers, we use boldface letters like $\mathbf{x}$ to non-scalar quantities like vectors and matrices, and we let $\|\mathbf{x}\|_0$ denote the number of nonzeros in $\mathbf{x}$. Given an optimization problem $\min_{\mathbf{x}} f(\mathbf{x})$, we say that a solution $\bar{\mathbf{x}}$ is feasible for the optimization problem if and only if the cost satisfies $f(\bar{\mathbf{x}}) < \infty$. It follows from this notation that $\min_{\mathbf{x}} f(\mathbf{x}) < \infty$ if and only if an optimization problem $\min_{\mathbf{x}} f(\mathbf{x})$ has a nonempty feasible region.

2. Background and Problem Setting

2.1. Robust Optimization and Linear Decision Rules

We consider robust optimization problems faced by firms in which decisions $\mathbf{x}_1,\dots,\mathbf{x}_T \in \mathbb{R}^n$ are made sequentially over a planning horizon of T discrete stages. In the beginning of each stage $t \in [T] \equiv \{1,\dots,T\}$, the firm observes a uncertain variable $\zeta_t \in \mathbb{R}$ that is chosen adversarially from an uncertainty set denoted by $\mathcal{U}_t$. The goal of the firm is to choose decisions sequentially, that is, adapting to the uncertain variables observed in the past stages, to minimize the firm's cost under an adversarial choice of the uncertain variables. Such problems are denoted generically by

$$\max_{\zeta_1 \in \mathcal{U}_1} \min_{\mathbf{x}_1 \in \mathbb{R}^n} \cdots \max_{\zeta_T \in \mathcal{U}_T} \min_{\mathbf{x}_T \in \mathbb{R}^n} C(\mathbf{x}_1,\dots,\mathbf{x}_T,\zeta_1,\dots,\zeta_T), \tag{RO}$$

where we adopt the convention that the cost $C(\mathbf{x}_1,\dots,\mathbf{x}_T,\zeta_1,\dots,\zeta_T) \in \mathbb{R}\cup\{\infty\}$ is equal to infinity if the chosen decisions are infeasible for the given realization of the uncertain variables. We assume without loss of generality throughout the paper that $\zeta_1 = 1$. For comprehensive introductions to the many applications of multi-period robust optimization problems of the form (RO), we refer the reader to excellent survey papers such as [13, 18, 31, 10].

Our work focuses on one of the most popular approximation methods for solving (RO), denoted by the optimization problem (LDR). Proposed in the seminal work of Ben-Tal et al. [4], (LDR) aims to obtain a computationally tractable approximation of (RO) by restricting the decisions in each stage to be a linear function of the uncertain variables observed in the past. Formally, the set of optimal linear decision rules for (RO) is defined as the set of optimal solutions for

$$\min_{\mathbf{y}_{t1},\dots,\mathbf{y}_{tt} \in \mathbb{R}^n:\, \forall t \in [T]} \; \max_{\zeta_1 \in \mathcal{U}_1,\dots,\zeta_T \in \mathcal{U}_T} C\left(\sum_{s=1}^{1} \mathbf{y}_{1s}\zeta_s,\dots,\sum_{s=1}^{T} \mathbf{y}_{Ts}\zeta_s,\, \zeta_1,\dots,\zeta_T\right). \tag{LDR}$$

Speaking intuitively, the goal of (LDR) is to obtain the best parameters for the linear decision rules, denoted by $\mathbf{y}_{t1},\dots,\mathbf{y}_{tt} \in \mathbb{R}^n$ for each $t \in [T]$. Given the parameters of the linear decision rules, the decision to make on each stage $t \in [T]$ is computed as $\mathbf{x}_t = \sum_{s=1}^{t} \mathbf{y}_{ts}\zeta_s$. It follows from the above notation that the number of parameters used for representing linear decision rules is given by $\sum_{t=1}^{T} nt = \frac{1}{2}nT(T+1) = \mathcal{O}(nT^2)$.

2.2. The Production-Inventory Problem

Our work focuses on a widely-studied class of production-inventory problems from [4]. The class of problems considers a firm with a central warehouse and E factories that aim to satisfy uncertain demand for a single product over a selling season. The selling season of the firm's product is discretized into T time periods, which are spaced equally over the selling season. In each time period $t \in [T]$, the firm sequentially performs the following three steps:

1. The firm replenishes the inventory level at the central warehouse by producing additional products at their factories. Let $x_{te} \geq 0$ denote the number of product units that the firm decides to produce at each of the factories $e \in [E] \equiv \{1, \dots, E\}$ at a per-unit cost of c_{te}. Each factory produces the additional units with zero lead time, and the additional units are stored immediately in the central warehouse.
2. The firm observes the customer demand at the central warehouse. The demand at the central warehouse is denoted by $\zeta_{t+1} \in \mathcal{U}_{t+1} \equiv [\underline{D}_{t+1}, \bar{D}_{t+1}]$, which must be satisfied immediately without backlogging from the inventory in the central warehouse. The lower and upper bounds in the uncertainty set, denoted by $\underline{D}_{t+1} < \bar{D}_{t+1}$, capture the minimum and maximum level of customer demand that the firm anticipates receiving in each time period t.
3. The firm verifies that the remaining inventory in the warehouse lies within a pre-specified interval given by $[V_{\min}, V_{\max}]$. Specifically, the remaining inventory level in the central warehouse at the end of each time period $t \in [T]$ must satisfy

$$V_{\min} \leq v_1 + \sum_{\ell=1}^{t} \sum_{e=1}^{E} x_{\ell e} - \sum_{s=2}^{t+1} \zeta_s \leq V_{\max},$$

   where v_1 is the initial inventory level in the central warehouse at the beginning of the selling horizon, $\sum_{\ell=1}^{t} \sum_{e=1}^{E} x_{\ell e}$ is the cumulative number of product units that have been produced at the factories up through time period t, and $\sum_{s=2}^{t+1} \zeta_s$ is the cumulative customer demand that has been observed at the central warehouse up through time period t.

In addition to satisfying the constraints on the inventory level in the central warehouse at the end of each time period, the firm's production decisions must satisfy $x_{te} \leq p_{te}$ in each time period $t \in [T]$ and each factory $e \in [E]$, where p_{te} is the maximum production level for factory e in time period t, and the firm's total production quantity across the selling season for each factory $e \in [E]$ must satisfy $\sum_{t=1}^{T} x_{te} \leq Q_e$, where Q_e is the maximum total production level for factory e. The goal of the firm is to satisfy the customer demand at minimal cost while satisfying production and warehouse constraints.

We observe that the above class of production-inventory problems from Ben-Tal et al. [4] is a special case of (RO) in which the cost function has the form

$$C(\mathbf{x}_1, \dots, \mathbf{x}_T, \zeta_1, \dots, \zeta_{T+1}) = \sum_{t=1}^{T} \sum_{e=1}^{E} c_{te} x_{te} \tag{1a}$$

$$\text{subject to} \quad \sum_{t=1}^{T} x_{te} \leq Q_e \qquad \forall e \in [E] \tag{1b}$$

$$0 \leq x_{te} \leq p_{te} \qquad \forall e \in [E], t \in [T] \tag{1c}$$

$$V_{\min} \leq v_1 + \sum_{\ell=1}^{t} \sum_{e=1}^{E} x_{\ell e} - \sum_{s=2}^{t+1} \zeta_s \leq V_{\max} \qquad \forall t \in [T] \tag{1d}$$

and where the uncertainty sets have the form $\mathcal{U}_1 \equiv [\underline{D}_1, \bar{D}_1], \ldots, \mathcal{U}_{T+1} \equiv [\underline{D}_{T+1}, \bar{D}_{T+1}]$ with $\underline{D}_1 = \bar{D}_1 = 1$ and $\underline{D}_{t+1} < \bar{D}_{t+1}$ for each time period $t \in [T]$. We use the convention that the cost function evaluates to (1a) if the constraints (1b)-(1d) are satisfied and equals infinity otherwise.[1] As a result, optimal linear decision rules for production-inventory problems can be obtained by solving (LDR), which we observe can be written as

$$\begin{aligned}
&\underset{\mathbf{y}_{t,1},\ldots,\mathbf{y}_{t,t} \in \mathbb{R}^E:\, \forall t \in [T]}{\text{minimize}} && \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_{T+1} \in \mathcal{U}_{T+1}} \left\{ \sum_{t=1}^{T} \sum_{e=1}^{E} c_{te} \left(\sum_{s=1}^{t} y_{t,s,e} \zeta_s \right) \right\} \\
&\text{subject to} && \sum_{t=1}^{T} \left(\sum_{s=1}^{t} y_{t,s,e} \zeta_s \right) \le Q_e && \forall e \in [E] \\
& && 0 \le \left(\sum_{s=1}^{t} y_{t,s,e} \zeta_s \right) \le p_{te} && \forall e \in [E],\, t \in [T] \\
& && V_{\min} \le v_1 + \sum_{\ell=1}^{t} \sum_{e=1}^{E} \left(\sum_{s=1}^{\ell} y_{\ell,s,e} \zeta_s \right) - \sum_{s=2}^{t+1} \zeta_s \le V_{\max} && \forall t \in [T] \\
& && \forall \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_{T+1} \in \mathcal{U}_{T+1}.
\end{aligned} \tag{LDR-1}$$

3. Main Theoretical Result

In this section, we present our main theoretical result regarding the sparsity of optimal linear decision rules for the class of production-inventory problems. Throughout this section, we assume that:

ASSUMPTION 1.

a. (LDR) *is feasible and the optimal objective value for* (LDR) *is finite.*

b. The uncertainty sets are intervals $\mathcal{U}_t = [\underline{D}_t, \bar{D}_t]$ *with* $\underline{D}_1 = \bar{D}_1 = 1$ *and* $\underline{D}_t < \bar{D}_t$ *for all* $t \ge 2$.

The first assumption states that there exist optimal linear decision rules for the robust optimization problem. The second assumption imposes that the uncertain variables in the robust optimization problem are chosen from box uncertainty sets. In the production-inventory problem, this second assumption corresponds to the fact that the customer demand is uncertain but bounded in each stage $t \ge 2$. The claim that $\underline{D}_1 = \bar{D}_1 = 1$ ensures without loss of generality that linear decision rules can have a nonzero offset.

In our main theoretical result, presented below as Theorem 1, we establish that there always exists a sparse optimal linear decision rule for the class of production-inventory problems (LDR-1):

THEOREM 1. *Consider a cost function of the form* (1a)-(1d) *and let Assumption 1 hold. Then there exists an optimal solution* $\bar{\mathbf{y}}$ *for* (LDR) *that satisfies* $\|\bar{\mathbf{y}}\|_0 \le 2 + 8E + 10T + 6ET$.

[1] We recall that (RO) and (LDR) involve cost functions in which the number of stages with decisions is equal to the number of stages with uncertain variables. The cost function (1a)-(1d) can be easily modified to match this format by adding a dummy decision variable $\mathbf{x}_{T+1} \in \mathbb{R}^E$ and the constants $p_{T+1,e} = c_{T+1,e} = 0$ for each $e \in [E]$.

Recall that the production-inventory problem with linear decision rules (LDR-1) has $\frac{1}{2}ET(T+1)=\mathcal{O}(ET^2)$ parameters. Theorem 1 shows that if (LDR-1) has an optimal solution, then there always exists optimal linear decision rules with $\mathcal{O}(ET)$ nonzero parameters. Furthermore, we notice that the number of parameters for representing static decision rules in production-inventory problems is equal to TE. Theorem 1 thus reveals that the complexity of optimal linear decision rules is at the same level as the complexity of static decision rules, which provides a new perspective for the success of linear decision rules: namely, the fundamental reason that linear decision rules exhibit superior performance to static decision rules in production-inventory problems is not due to the enormous parameter space of linear decision rules, but rather a very small cardinality of significant parameters that are contained only in the linear decision rules.

Remark 1. We note that Theorem 1 can be directly extended to production-inventory problems with multiple products in each time period. Specifically, if $\mathcal{U}_t$ is a multi-dimensional box and $\mathbf{y}_{t,s,e}$ is a vector in the corresponding dimension, one can show the existence of a sparse optimal solution using the same proof techniques as Theorem 1. Furthermore, our proof techniques behind Theorem 1 are not limited to the class of production-inventory problems from Ben-Tal et al. [4], and we present extensions of Theorem 1 to other classes of robust optimization problems in §6.

In the remainder of this section, we present a high-level roadmap for the proof of Theorem 1. Our proof is based on a new understanding of the extreme points of the feasible regions of linear decision rules for a general class of multi-period robust optimization problems. We postpone the detailed proof of Theorem 1 to Appendix B.

Our proof of Theorem 1 focuses on a class of robust optimization problems that includes production-inventory problems as a special case. This class of robust optimization problems is characterized by cost functions of the form

$$\begin{aligned} C(\mathbf{x}_1,\ldots,\mathbf{x}_T,\zeta_1,\ldots,\zeta_T) = \sum_{t=1}^{T}\mathbf{a}_{0,t}^{\intercal}\mathbf{x}_t - \sum_{t=1}^{T} b_{0,t}\zeta_t \\ \text{subject to} \quad \sum_{t=1}^{T}\mathbf{a}_{i,t}^{\intercal}\mathbf{x}_t - \sum_{t=1}^{T} b_{i,t}\zeta_t \le c_i \quad \forall i\in[m], \end{aligned} \tag{C-G}$$

where the parameters of the cost function are $\mathbf{a}_{0,t},\ldots,\mathbf{a}_{m,t}\in\mathbb{R}^n$ and $b_{0,t},\ldots,b_{m,t}\in\mathbb{R}$ for each stage $t\in[T]$ and $c_i\in\mathbb{R}$ for each constraint $i\in[m]$. For robust optimization problems (RO) with cost functions of the form (C-G), we observe that (LDR) can be reformulated by its epigraph as

$$\begin{aligned} &\underset{\substack{c_0\in\mathbb{R} \\ \mathbf{y}_{t,1},\ldots,\mathbf{y}_{t,t}\in\mathbb{R}^n:\,\forall t\in[T]}}{\text{minimize}} && c_0 \\ &\text{subject to} && \max_{\zeta_1\in\mathcal{U}_1,\ldots,\zeta_T\in\mathcal{U}_T}\left\{\sum_{s=1}^{T}\left(-b_{i,s}+\sum_{t=s}^{T}\mathbf{a}_{i,t}^{\intercal}\mathbf{y}_{t,s}\right)\zeta_s\right\}\le c_i \;\forall i\in\{0,\ldots,m\}. \end{aligned} \tag{LDR-G}$$

The decision variables in the above optimization problem include the parameters of the linear decision rules as well as an epigraph decision variable $c_0 \in \mathbb{R}$. For this general class of problems, we will assume in addition to Assumption 1 that:

ASSUMPTION 2. *If $C(\mathbf{x}_1, \ldots, \mathbf{x}_T, \zeta_1, \ldots, \zeta_T) < \infty$, then the decisions satisfy $\mathbf{x}_1, \ldots, \mathbf{x}_T \geq \mathbf{0}$.*

The above assumption essentially stipulates that the constraints of (C-G) ensure that feasible decisions for the robust optimization problem satisfy $\mathbf{x}_t \geq \mathbf{0}$ for each stage $t \in [T]$. This is a reasonable assumption because, in practice, the decisions usually refer to levers like order quantities or prices of products that must take nonnegative values. For example, in the class of production-inventory problems from Ben-Tal et al. [4], the decisions $\mathbf{x}_t \in \mathbb{R}^E$ refer to the production quantities at the factories at time period t, and the constraint (1c) guarantees that the production levels are nonnegative.

Equipped with the above notation for a general class of multi-period robust optimization problems, our proof of Theorem 1 is split into three major steps, organized below as Lemmas 1, 2, and 3. In our first step, denoted below by Lemma 1, we characterize the feasible region of (LDR-G) and show that the set of feasible solutions of (LDR-G) is a polyhedron with at least one extreme point.

LEMMA 1. *Let Assumptions 1 and 2 hold. Then the set of feasible solutions to* (LDR-G) *is a nonempty polyhedron with at least one extreme point.*

Let us make two observations about the above lemma. First, since (LDR-G) has a linear objective function and a polyhedral feasible region, we observe that (LDR-G) is a linear optimization problem. Therefore, whenever this linear optimization problem has at least one extreme point and has a finite optimal objective value, there must exist an optimal solution for (LDR-G) that is an extreme point of its feasible set (see, e.g., [8, Theorem 2.7]). Second, it follows from routine arguments in linear optimization that if $(\bar{\mathbf{y}}, \bar{c}_0)$ is an extreme point of (LDR-G), then the value of $\bar{c}_0$ is uniquely determined by the value of $\bar{\mathbf{y}}$.[2] Therefore, we will for the sake of simplicity omit $\bar{c}_0$ when referring to an extreme point $(\bar{\mathbf{y}}, \bar{c}_0)$ of the set of feasible solutions to (LDR-G).

In the second step of our proof of Theorem 1, we provide an explicit characterization of the extreme points for the set of feasible solutions of (LDR-G). This key step, which is presented below as Lemma 2, reveals that the extreme points of the set of feasible solutions of (LDR-G) can always be represented as the unique solution of a certain decomposable system of equations.

[2] Suppose that $(\bar{\mathbf{y}}, \bar{c}_0)$ is an extreme point of the set of feasible solutions for (LDR-G). Then we readily observe that the value $\bar{c}_0$ must satisfy $\bar{c}_0 = \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} \left\{ \sum_{t=1}^T \mathbf{a}_{i,t}^\intercal \left(\sum_{s=1}^t \mathbf{y}_{t,s} \zeta_s \right) - \sum_{t=1}^T b_{i,t} \zeta_t \right\}$.

Lemma 2. *Let Assumption 1 hold, and let* $(\bar{\mathbf{y}}, \bar{c}_0)$ *be an extreme point of the feasible set of* (LDR-G). *Then there exists an index set* $\mathcal{I}^{\bar{\mathbf{y}}} \subseteq \{0, \ldots, m\}$, *an index set* $\mathcal{T}_i^{\bar{\mathbf{y}}} \subseteq [T]$ *for each* $i \in \mathcal{I}^{\bar{\mathbf{y}}}$, *and a hyperplane* $(\boldsymbol{\alpha}_i^{\bar{\mathbf{y}}}, \beta_i^{\bar{\mathbf{y}}})$ *for each* $i \in \mathcal{I}^{\bar{\mathbf{y}}}$ *such that* $\bar{\mathbf{y}}$ *is the unique solution of the following system:*

$$\sum_{s=1}^{T}\sum_{t=s}^{T} \boldsymbol{\alpha}_{i,t,s}^{\bar{\mathbf{y}}} \cdot \mathbf{y}_{t,s} = \beta_i^{\bar{\mathbf{y}}} \qquad \forall i \in \mathcal{I}^{\bar{\mathbf{y}}}, \tag{HARD}$$

$$\sum_{t=s}^{T} \mathbf{a}_{i,t} \cdot \mathbf{y}_{t,s} = b_{i,s} \qquad \forall i \in \mathcal{I}^{\bar{\mathbf{y}}}, s \in \mathcal{T}_i^{\bar{\mathbf{y}}}. \tag{EASY}$$

Let us provide an interpretation of this second step. In a nutshell, Lemma 2 establishes that every extreme point of the feasible set of (LDR-G) is the unique solution to a linear system that can be decomposed into two types of equations: a small number of hard equations and a large number of easy equations. Indeed, the first type of equations (HARD) is defined by hyperplanes $(\boldsymbol{\alpha}_i^{\bar{\mathbf{y}}}, \beta_i^{\bar{\mathbf{y}}})$ which are functions of $\bar{\mathbf{y}}$, the extreme point of the set of feasible solutions of (LDR-G). We refer to this first type of equations by the moniker (HARD) because the structure of the hyperplanes $(\boldsymbol{\alpha}_i^{\bar{\mathbf{y}}}, \beta_i^{\bar{\mathbf{y}}})$ cannot be analyzed independently of extreme point $\bar{\mathbf{y}}$. In contrast, the second type of equations (EASY) is defined by hyperplanes that are independent of $\bar{\mathbf{y}}$, and so the structure of the second type of equations can be analyzed statically using the structure of the underlying robust optimization problem. The number of equations in (HARD) is at most equal to $m+1 = \mathcal{O}(m)$, where m is the number of constraints in (C-G), whereas the number of equations in (EASY) is at most equal to $(m+1)T = \mathcal{O}(mT)$. Hence, when the number of stages is large, we observe that there can be significantly more equations of type (EASY) than of type (HARD).

The third and final step in our proof of Theorem 1 is to show that the unique solution to the system of equations (HARD)-(EASY) is sparse when (C-G) is equal to the cost function of the production-inventory problem from lines (1a)-(1d). Specifically, it turns out in many practical problems including the production-inventory problem from Ben-Tal et al. [4] that the system of equations (HARD)-(EASY) can be massaged into an instance of a system of equations denoted below (S-1)-(S-3), in which the number of equations in line (S-1) grows linearly in the number of stages of the robust optimization problem. Through this insight, the following Lemma 3 proves that the number of nonzeros in every extreme point of (LDR-G) in problems such as the production-inventory problem from Ben-Tal et al. [4] grows linearly with respect to the number of stages.

Lemma 3. *Let* $\mathbf{P}_1 \in \mathbb{R}^{m_1 \times n}$, $\mathbf{P}_2 \in \mathbb{R}^{m_2 \times n}$, $\mathbf{q} \in \mathbb{R}^{m_1}$, *and* $\mathcal{N} \subseteq [n]$, *where* $m_1, m_2 \le n$. *Suppose that there is a unique* $\bar{\mathbf{z}} \in \mathbb{R}^n$ *that satisfies the system of equations*

$$\mathbf{P}_1 \mathbf{z} = \mathbf{q} \tag{S-1}$$

$$\mathbf{P}_2 \mathbf{z} = \mathbf{0} \tag{S-2}$$

$$z_j = 0 \quad \forall j \in \mathcal{N}, \tag{S-3}$$

and suppose that each column of $\mathbf{P}_2$ *has at most one nonzero entry, that is,* $\sum_{i=1}^{m_2} \mathbb{I}\{p_{2,i,j} \neq 0\} \leq 1$ *for each* $j \in [n]$*. Then the unique solution* $\bar{\mathbf{z}}$ *has at most* $2m_1$ *nonzero entries, that is,* $\|\bar{\mathbf{z}}\|_0 \leq 2m_1$*.*

In summary, our proof of Theorem 1 considers a more general class of cost functions (C-G) and looks at the epigraph formulation (LDR-G) of the corresponding problem of computing optimal linear decision rules. Under reasonable assumptions, we show in Lemma 1 that (LDR-G) is indeed a linear optimization problem with at least one extreme point in its feasible region, thus establishing the existence of an optimal extreme point. Furthermore, we show in Lemma 2 that any extreme point of the feasible region is the unique solution to a linear system with two types of equations (HARD) and (EASY), where the number of equations in (HARD) is significantly smaller than the number of equations in (EASY). By utilizing the structure of the production-inventory problem, we conclude by showing in Lemma 3 that the number of nonzeros in every extreme point grows at most linearly with respect to the number of stages of the robust optimization problem. Our formal proof of Theorem 1 is found in Appendix B.

4. The Value of Sparsity in Algorithm Design

In this section, we discuss the implications of Theorem 1 on the design of memory-efficient algorithms for computing linear decision rules in multi-period robust optimization problems. Specifically, our main contribution to this section is a novel algorithm for solving (LDR) in applications in which the optimal linear decision rules are sparse and the number of time periods is large. Our algorithm is applicable to the broad class of multi-period robust optimization problems in which the uncertainty sets satisfy Assumption 1 and the cost function has the form (C-G)[3]. In §5.2, we show through numerical experiments that our algorithm from this section can provide *significant* speedups over state-of-the-art linear optimization solvers for computing near-optimal linear decision rules in robust optimization problems with huge numbers of time periods.

4.1. Preliminaries

The overarching goal of our algorithm, which is presented formally in §4.3, is to find an optimal or near-optimal solution to (LDR) without ever needing to solve a large-scale linear optimization problem that exceeds computer memory. To accomplish this goal, each iteration of our algorithm consists of solving a restricted version of (LDR) in which sparsity is imposed onto the linear decision rules.

In greater detail, each iteration of our algorithm begins with a choice of an active set $\mathcal{A} \subseteq \{(t,s,j) : t,s \in [T], j \in [n], s \leq t\}$ corresponding to the parameters of the linear decision rules that will be allowed to take values other than zero. If the uncertainty sets in the robust optimization problem

[3] A discussion of the class of multi-period robust optimization problems in which the uncertainty sets satisfy Assumption 1 and the cost function has the form (C-G) can be found in §3.

satisfy Assumption 1 and the cost function in the robust optimization problem has the form (C-G), then the optimization problem (LDR) with sparsity constraints from the active set $\mathcal{A}$ is given by

$$\begin{aligned} \underset{\substack{c_0 \in \mathbb{R} \\ \mathbf{y}_{t,1},\ldots,\mathbf{y}_{t,t} \in \mathbb{R}^n :\, \forall t \in [T]}}{\text{minimize}} \quad & c_0 \\ \text{subject to} \quad & \max_{\zeta_1 \in \mathcal{U}_1,\ldots,\zeta_T \in \mathcal{U}_T} \left\{ \sum_{s=1}^{T} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) \zeta_s \right\} \le c_i \quad \forall i \in \{0,\ldots,m\} \\ & y_{t,s,i} = 0 \quad \forall (t,s,i) \notin \mathcal{A}, \end{aligned}$$

which can be rewritten equivalently as

$$\begin{aligned} \underset{\substack{c_0 \in \mathbb{R} \\ y_{t,s,j} \in \mathbb{R} :\, \forall (t,s,j) \in \mathcal{A}}}{\text{minimize}} \quad & c_0 \\ \text{subject to} \quad & \sum_{s=1}^{T} \max_{\zeta_s \in \mathcal{U}_s} \left\{ \left(-b_{i,s} + \sum_{t,j:(t,s,j) \in \mathcal{A}} a_{i,t,j} y_{t,s,j} \right) \zeta_s \right\} \le c_i \quad \forall i \in \{0,\ldots,m\}. \end{aligned} \tag{LDR-$\mathcal{A}$}$$

By solving the optimization problem (LDR-$\mathcal{A}$) in each iteration of our algorithm, our algorithm obtains a linear decision rule in which at most $|\mathcal{A}|$ of the parameters of the linear decision rule are nonzero. We readily observe that the optimization problem (LDR-$\mathcal{A}$) is a restricted version of the optimization problem (LDR), in the sense that any linear decision rule that is feasible for (LDR-$\mathcal{A}$) is a feasible linear decision rule for (LDR).

Our motivation for designing an algorithm around the optimization problem (LDR-$\mathcal{A}$) is based on computer memory. Indeed, (LDR-$\mathcal{A}$) is an optimization problem with approximately $|\mathcal{A}|$ decision variables, whereas the optimization problem (LDR) has approximately nT^2 decision variables. Hence, if each iteration of our algorithm chooses an active set that is sparse, meaning that $|\mathcal{A}| \ll nT^2$, then storing and manipulating the decision variables of the optimization problem (LDR-$\mathcal{A}$) will require significantly less computer memory than storing and manipulating the decision variables of (LDR). Moreover, we established in §3 that there can exist optimal linear decision rule for real-world applications of robust optimization that satisfy $\|\mathbf{y}\|_0 = \mathcal{O}(nT)$. Therefore, we observe that it can be possible to choose a sparse active set satisfying $|\mathcal{A}| \ll nT^2$ for which the optimization problem (LDR-$\mathcal{A}$) is equivalent to the optimization problem (LDR).

In view of the above motivation, the rest of this section is organized as follows. In §4.2, we develop a novel reformulation technique that enables us to find an optimal solution to (LDR-$\mathcal{A}$) in reasonable computation time when the active set is sparse. In §4.3, we present an iterative algorithm that uses duality to find an optimal active set $\mathcal{A}$ in the optimization problem (LDR-$\mathcal{A}$). In §4.4, we discuss several improvements and implementation details to the active set method.

4.2. Compact Reformulation For Fixed Active Set

In this subsection, we develop a novel reformulation technique for the optimization problem (LDR-$\mathcal{A}$). When the chosen active set is sparse, we show that our novel linear optimization reformulation can have drastically fewer decision variables and constraints compared to the standard linear optimization reformulation that is obtained by the traditional robust counterpart technique.

To motivate our proposed reformulation technique, we begin by discussing the standard reformulation of (LDR-$\mathcal{A}$) obtained using the robust counterpart technique. Indeed, we observe from strong duality for linear optimization and from Assumption 1 that the following equality holds for each time period $s \in \{1,, \ldots, T\}$ and each constraint $i \in \{0, \ldots, m\}$:

$$\max_{\zeta_s \in \mathcal{U}_s} \left\{ \left(-b_{i,s} + \sum_{t,j:(t,s,j)\in\mathcal{A}} a_{i,t,j} y_{t,s,j} \right) \zeta_s \right\} = \begin{bmatrix} \underset{\bar{\omega}_{i,s}, \underline{\omega}_{i,s} \in \mathbb{R}}{\text{minimize}} & \bar{D}_s \bar{\omega}_{i,s} - \underline{D}_s \underline{\omega}_{i,s} \\ \text{subject to} & \bar{\omega}_{i,s} - \underline{\omega}_{i,s} = -b_{i,s} + \sum_{t,j:(t,s,j)\in\mathcal{A}} a_{i,t,j} y_{t,s,j} \\ & \bar{\omega}_{i,s}, \underline{\omega}_{i,s} \geq 0 \end{bmatrix}.$$

Applying the above equality to each period $s \in \{1, \ldots, T\}$ and constraint $i \in \{0, \ldots, m\}$, the robust counterpart technique yields the following standard reformulation of (LDR-$\mathcal{A}$) as a linear optimization problem:

$$\begin{aligned} \underset{\substack{c_0 \in \mathbb{R} \\ y_{t,s,j} \in \mathbb{R}:\, \forall (t,s,j) \in \mathcal{A} \\ \bar{\boldsymbol{\omega}}_i, \underline{\boldsymbol{\omega}}_i \in \mathbb{R}^T : \forall i \in \{0,\ldots,m\}}}{\text{minimize}} \quad & c_0 \\ \text{subject to} \quad & \sum_{s=1}^{T} \left(\bar{D}_s \bar{\omega}_{i,s} - \underline{D}_s \underline{\omega}_{i,s} \right) \leq c_i && \forall i \in \{0,\ldots,m\} \\ & \bar{\omega}_{i,s} - \underline{\omega}_{i,s} = -b_{i,s} + \sum_{t,j:(t,s,j)\in\mathcal{A}} a_{i,t,j} y_{t,s,j} && \forall i \in \{0,\ldots,m\},\ s \in [T] \\ & \bar{\omega}_{i,s}, \underline{\omega}_{i,s} \geq 0 && \forall i \in \{0,\ldots,m\},\ s \in [T]. \end{aligned} \tag{P-$\mathcal{A}$'}$$

Unfortunately, regardless of our choice of the active set, we observe that the numbers of decision variables and constraints in the linear optimization problem (P-$\mathcal{A}$') still grow *quadratically* with respect to the number of time periods. Indeed, the linear optimization problem (P-$\mathcal{A}$') consists of approximately $|\mathcal{A}| + 2mT$ decision variables and mT constraints. Moreover, the number of constraints m in real-world multi-period robust optimization problems typically satisfies the inequality $m \geq nT$.[4] For this reason, the numbers of decision variables and constraints in the linear optimization problem (P-$\mathcal{A}$') will typically grow quadratically with respect to the number of time periods, even when the active set satisfies $|\mathcal{A}| \ll nT^2$.

[4] For example, in the production inventory problem from §2.2 with E factories and T periods, the number of decisions in each period is equal to the number of factories, $n = E$, and the number of constraints is given by $m = E + E(T + 1) + (T + 1)$. More generally, the number of constraints will satisfy the inequality $m \geq nT$ in real-world applications in which there are nonnegativity constraints on each of the decisions in each period of the planning problem.

Our novel linear optimization reformulation of the optimization problem (LDR-$\mathcal{A}$), which is stated at the end of this subsection as (P-$\mathcal{A}$), can be viewed as a simple modification of the linear optimization problem (P-$\mathcal{A}$') that exploits the fact that imposing sparsity constraints onto linear decision rules can have a side effect of inducing redundancy in the constraints of the optimization problem (LDR-$\mathcal{A}$). Indeed, for any active set $\mathcal{A}$ and each period $s \in \{1,\dots,T\}$, let us define the following set of tuples:

$$\mathcal{K}^{\mathcal{A},s} \triangleq \bigcup_{i=0}^{m} \left\{ \left(b_{i,s}, (a_{i,t,j})_{t,j:(t,s,j)\in\mathcal{A}} \right) \right\}.$$

For each period $s \in \{1,\dots,T\}$, the cardinality of the set of tuples $K^{\mathcal{A},s} \triangleq |\mathcal{K}^{\mathcal{A},s}|$ can be interpreted as the number of unique optimization problems $\max_{\zeta_s \in \mathcal{U}_s}\{(-b_{i,s} + \sum_{t,j:(t,s,j)\in\mathcal{A}} a_{i,t,j} y_{t,s,j})\zeta_s\}$ across all of the constraints $i \in \{0,\dots,m\}$. The key observation that underpins our subsequent reformulation is that the total number of unique tuples $K^{\mathcal{A}} \triangleq \sum_{s=1}^{T} K^{\mathcal{A},s}$ is often proportional to the cardinality of the active set $\mathcal{A}$. In particular, for the production-inventory problem from §2.2, we have the following result:

PROPOSITION 1. *For the production inventory problem,* $K^{\mathcal{A}} \le 4|\mathcal{A}| + ET + 5T + E + 1$.

Hence, the above proposition shows for production-inventory problems that the number of unique tuples satisfies $K^{\mathcal{A}} = \mathcal{O}(ET)$ when the active set satisfies $|\mathcal{A}| = \mathcal{O}(ET)$.

We now complete the development of our proposed reformulation by demonstrating that the linear optimization problem (P-$\mathcal{A}$') can be rewritten using $\mathcal{O}(K^{\mathcal{A}} + |\mathcal{A}|)$ decision variables and $\mathcal{O}(K^{\mathcal{A}} + m)$ constraints. Indeed, let the kth tuple in $\mathcal{K}^{\mathcal{A},s}$ for each $k \in \{1,\dots,|\mathcal{K}^{\mathcal{A},s}|\} \equiv [K^{\mathcal{A},s}]$ and each period $s \in [T]$ be denoted by $(b_{k,s}^{\mathcal{A},s}, (a_{k,t,j}^{\mathcal{A},s})_{t,j:(t,s,j)\in\mathcal{A}})$, and let $\pi^{\mathcal{A},s} : \{0,\dots,m\} \to [K^{\mathcal{A},s}]$ be defined for each period $s \in [T]$ as the mapping that satisfies the following equality for each $i \in \{0,\dots,m\}$:

$$\left(b_{i,s}, (a_{i,t,j})_{t,j:(t,s,j)\in\mathcal{A}} \right) = \left(b^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}, \left(a^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),t,j} \right)_{t,j:(t,s,j)\in\mathcal{A}} \right).$$

With the above notation, we observe that (P-$\mathcal{A}$') can be rewritten equivalently as

$$\begin{aligned}
&\underset{\substack{c_0 \in \mathbb{R} \\ y_{t,s,j} \in \mathbb{R}:\, \forall (t,s,j)\in\mathcal{A} \\ \bar{\boldsymbol{\omega}}_i, \underline{\boldsymbol{\omega}}_i \in \mathbb{R}^T : \forall i \in \{0,\dots,m\}}}{\text{minimize}} && c_0 \\
&\text{subject to} && \sum_{s=1}^{T} \left(\bar{D}_s \bar{\omega}_{i,s} - \underline{D}_s \underline{\omega}_{i,s} \right) \le c_i && \forall i \in \{0,\dots,m\} \\
& && \bar{\omega}_{i,s} - \underline{\omega}_{i,s} = -b^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} + \sum_{t,j:(t,s,j)\in\mathcal{A}} a^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),t,j} y_{t,s,j} && \forall i \in \{0,\dots,m\},\ s \in [T] \\
& && \bar{\omega}_{i,s}, \underline{\omega}_{i,s} \ge 0 && \forall i \in \{0,\dots,m\},\ s \in [T].
\end{aligned}$$

In particular, we observe that there exists an optimal solution for the above optimization problem for which the equality $(\bar{\omega}_{i,s}, \underline{\omega}_{i,s}) = (\bar{\omega}_{i',s}, \underline{\omega}_{i',s})$ holds for all constraints $i, i' \in \{0,\dots,m\}$ and all periods $s \in [T]$ that satisfy the equality $\pi^{\mathcal{A},s}(i) = \pi^{\mathcal{A},s}(i')$. Therefore, the above linear optimization problem can be rewritten as

$$
\begin{aligned}
&\underset{\substack{c_0 \in \mathbb{R} \\ y_{t,s,j} \in \mathbb{R}:\, \forall (t,s,j) \in \mathcal{A} \\ \bar{\omega}^{\mathcal{A},s}_{k,s}, \underline{\omega}^{\mathcal{A},s}_{k,s} \in \mathbb{R}: \forall s \in [T], k \in [K^{\mathcal{A},s}]}}{\text{minimize}} && c_0 \\
&\text{subject to} && \sum_{s=1}^{T} \left(\bar{D}_s \bar{\omega}^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} - \underline{D}_s \underline{\omega}^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} \right) \le c_i && \forall i \in \{0,\ldots,m\} \\
& && \bar{\omega}^{\mathcal{A},s}_{k,s} - \underline{\omega}^{\mathcal{A},s}_{k,s} = -b^{\mathcal{A},s}_{k,s} + \sum_{t,j:(t,s,j) \in \mathcal{A}} a^{\mathcal{A},s}_{k,t,j} y_{t,s,j} && \forall s \in [T], k \in [K^{\mathcal{A},s}] \\
& && \bar{\omega}^{\mathcal{A},s}_{k,s}, \underline{\omega}^{\mathcal{A},s}_{k,s} \ge 0 && \forall s \in [T], k \in [K^{\mathcal{A},s}],
\end{aligned}
\tag{P-$\mathcal{A}$}
$$

with the understanding that any optimal solution of (P-$\mathcal{A}$) can be transformed into an optimal solution for (P-$\mathcal{A}$') by applying the transformation $\left(\bar{\omega}_{i,s}, \underline{\omega}_{i,s}\right) \triangleq \left(\bar{\omega}^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}, \underline{\omega}^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}\right)$ for each period $s \in [T]$ and constraint $i \in \{0,\ldots,m\}$. We observe from inspection that the linear optimization problem (P-$\mathcal{A}$) has $1 + |\mathcal{A}| + 2K^{\mathcal{A}} = \mathcal{O}\left(K^{\mathcal{A}} + |\mathcal{A}|\right)$ decision variables and $1 + m + K^{\mathcal{A}} = \mathcal{O}\left(K^{\mathcal{A}} + m\right)$ constraints. The linear optimization problem (P-$\mathcal{A}$) thus constitutes our novel linear optimization reformulation of the optimization problem (LDR-$\mathcal{A}$).

4.3. Active Set Method

Equipped with our compact linear optimization reformulation of (LDR-$\mathcal{A}$) from the end of §4.2, we now formally present our algorithm for solving the optimization problem (LDR). Our algorithm for finding an optimal solution for (LDR) consists of iteratively solving the optimization problem (LDR-$\mathcal{A}$) with different choices of the active set $\mathcal{A}$. Specifically, each iteration of our algorithm starts with an active set $\mathcal{A}$. We then perform the following steps.

Step 1. Given the active set $\mathcal{A}$, the first step of the current iteration of the active set method consists of solving the dual of the linear optimization problem (P-$\mathcal{A}$), which is given by

$$
\begin{aligned}
&\underset{\substack{\lambda_0,\ldots,\lambda_m \in \mathbb{R}, \\ \zeta^{\mathcal{A},s}_{k,s} \in \mathbb{R}\, \forall s \in [T], k \in [K^{\mathcal{A},s}]}}{\text{maximize}} && -\sum_{i=1}^{m} c_i \lambda_i - \sum_{s=1}^{T} \sum_{k=1}^{K^{\mathcal{A},s}} b^{\mathcal{A},s}_{k,s} \zeta^{\mathcal{A},s}_{k,s} \\
&\text{subject to} && \sum_{k=1}^{K^{\mathcal{A},s}} a^{\mathcal{A},s}_{k,t,j} \zeta^{\mathcal{A},s}_{k,s} = 0 && \forall (t,s,j) \in \mathcal{A} \\
& && \underline{D}_s \left(\sum_{i \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i) = k} \lambda_i \right) \le \zeta^{\mathcal{A},s}_{k,s} \le \bar{D}_s \left(\sum_{i \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i) = k} \lambda_i \right) && \forall s \in [T], k \in [K^{\mathcal{A},s}] \\
& && \lambda_0 = 1 \\
& && \lambda_i \ge 0 && \forall i \in \{1,\ldots,m\}.
\end{aligned}
\tag{D-$\mathcal{A}$}
$$

To simplify the discussion of our algorithm, we assume for now that the optimal objective value of (D-$\mathcal{A}$) is finite.[5] In that case, it follows from strong duality for linear optimization that the optimal

[5] This assumption is relaxed in §4.4.5.

objective value of (D-$\mathcal{A}$) is equal to the optimal objective value of (LDR-$\mathcal{A}$). Moreover, we observe that an optimal linear decision rule for the optimization problem (LDR-$\mathcal{A}$) can be obtained by extracting the KKT solutions of the equality constraints $\sum_{k=1}^{K^{\mathcal{A},s}} a_{k,t,j}^{\mathcal{A},s} \zeta_{k,s}^{\mathcal{A},s} = 0$ for all $(t,s,j) \in \mathcal{A}$. Hence, if the optimization problem (D-$\mathcal{A}$) has a feasible solution, then a feasible solution for the optimization problem (LDR) with objective value equal to the optimal objective value of (LDR-$\mathcal{A}$) can be obtained from solving the linear optimization problem (D-$\mathcal{A}$) using the simplex method.[6]

Step 2. The next step of the current iteration of our algorithm consists of checking whether the linear decision rule extracted from the KKT solution of the linear optimization problem (D-$\mathcal{A}$) is an optimal solution of the optimization problem (LDR). To perform this step, we utilize the following proposition. In the following proposition, and throughout the rest of the paper, we define the fraction 0/0 to be equal to 0.

PROPOSITION 2. *Consider any optimal solution for* (D-$\mathcal{A}$). *If the solution satisfies*

$$\sum_{i=0}^{m} a_{i,t,j} \left(\frac{\lambda_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta_{\pi^{\mathcal{A},s}(i),s}^{\mathcal{A},s} = 0 \quad \forall (t,s,j) \notin \mathcal{A}, \tag{1}$$

then the linear decision rule extracted from the KKT solution of the linear optimization problem (D-$\mathcal{A}$) *is an optimal solution of the optimization problem* (LDR).

The above proposition provides the condition that we will use for terminating the algorithm. Namely, after solving the optimization problem (D-$\mathcal{A}$), our algorithm checks whether the optimal solution that was obtained from solving the optimization problem (D-$\mathcal{A}$) satisfies line (1). If line (1) is satisfied, then the algorithm terminates. Otherwise, our algorithm proceeds to Step 3. We note that if the active set satisfies $|\mathcal{A}| \ll nT^2$, then the equalities in line (1) can be checked efficiently from the perspectives of computation time as well as the perspective of computer memory (see §4.4.1 for implementation details).

Step 3. If the algorithm does not terminate in Step 2, then the last step of the current iteration of our algorithm consists of choosing the active set that will be used in the next iteration. To choose the new active set, we employ a heuristic that uses the optimal solution for the linear optimization problem (D-$\mathcal{A}$) to construct a new active set by adding elements to the current active set. In greater detail, let us define the following sets:

$$\mathcal{A}^{\neq} \triangleq \left\{ (t,s,j) \notin \mathcal{A} : \quad \sum_{i=0}^{m} a_{i,t,j} \left(\frac{\lambda_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta_{\pi^{\mathcal{A},s}(i),s}^{\mathcal{A},s} \neq 0 \right\};$$

$$\mathcal{A}^{=} \triangleq \left\{ (t,s,j) \notin \mathcal{A} : \quad \sum_{i=0}^{m} a_{i,t,j} \left(\frac{\lambda_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta_{\pi^{\mathcal{A},s}(i),s}^{\mathcal{A},s} = 0 \right\}.$$

[6] If the linear optimization problem (D-$\mathcal{A}$) is solved using the simplex algorithm, then the KKT solution can be extracted directly from the optimal basis, where $y_{t,s,j}$ is the KKT solution.

The set $\mathcal{A}^{\neq}$ can be interpreted as the set of all tuples $(t,s,j) \notin \mathcal{A}$ that do not satisfy the equality in line (1), whereas the set $\mathcal{A}^{=}$ can be interpreted as the set of all tuples $(t,s,j) \notin \mathcal{A}$ that do satisfy the equality in line (1). To make sense of the above definitions, we offer the following remarks.

REMARK 2. It follows from construction that $\mathcal{A} \cup \mathcal{A}^{\neq} \cup \mathcal{A}^{=} = \{(t,s,j) : t,s \in [T], j \in [n], s \le t\}$.

REMARK 3. It follows from the fact that the algorithm did not terminate in Step 2 that $\mathcal{A}^{\neq} \neq \emptyset$.

REMARK 4. The sets $\mathcal{A}^{\neq}, \mathcal{A}^{=}$ are computed for free during Step 2.

In view of the above definitions of the sets $\mathcal{A}^{\neq}$ and $\mathcal{A}^{=}$, we now state our heuristic for choosing the active set that will be used in the next iteration. Specifically, our heuristic chooses the active set that will be used in the next iteration as $\mathcal{A} \cup \widetilde{\mathcal{A}}$ for some nonempty subset $\widetilde{\mathcal{A}} \subseteq \mathcal{A}^{\neq}$. The specific procedure that we use for choosing the subset $\widetilde{\mathcal{A}} \subseteq \mathcal{A}^{\neq}$ is described in §4.4.3, and the motivation behind our heuristic is given by the following proposition.

PROPOSITION 3. *Consider any optimal solution of the linear optimization problem* (D-$\mathcal{A}$)*, and let $\mathcal{A}^{\neq}$ and $\mathcal{A}^{=}$ be the subsets of tuples corresponding to that optimal solution. If the active set in the next iteration is chosen to be $\mathcal{A} \cup \widetilde{\mathcal{A}}$ for some subset $\widetilde{\mathcal{A}} \subseteq \mathcal{A}^{=}$, then the optimal objective value of* (LDR-$\mathcal{A} \cup \widetilde{\mathcal{A}}$) *will be equal to the optimal objective value of* (LDR-$\mathcal{A}$).

To make sense of the above proposition, we recall that the goal of our algorithm is to find an active set $\mathcal{A}$ for which the optimal objective value of (LDR-$\mathcal{A}$) is equal to the optimal objective value of (LDR). Moreover, we recall that the optimal objective value of (LDR-$\mathcal{A}$) (which is equal to the optimal objective value of (D-$\mathcal{A}$)) is always greater than or equal to the optimal objective value of (LDR). In view of those recollections, Proposition 3 shows that choosing an active set in the next iteration of algorithm that includes elements from $\mathcal{A}^{\neq}$ is imperative for decreasing the optimal objective value of the linear optimization problem (D-$\mathcal{A}$). This motivates our heuristic for choosing the active set that will be used in the next iteration to be equal to $\mathcal{A} \cup \widetilde{\mathcal{A}}$ for some nonempty subset $\widetilde{\mathcal{A}} \subseteq \mathcal{A}^{\neq}$.

4.4. Improvements

We summarize here several improvements and omitted details for our algorithm from §4.3.

4.4.1. Evaluating the termination condition. In Step 2 of each iteration of our algorithm, we decide whether to terminate the algorithm by determining whether the optimal solution obtained from solving (D-$\mathcal{A}$) satisfies line (1). In the following Lemma 4, we show that line (1) can be evaluated efficiently, both with respect to computation time and computer memory. Note that the following lemma makes use of the quantity $Z \triangleq \sum_{i=0}^{m} \sum_{t=1}^{T} \sum_{s=1}^{t} \sum_{j=1}^{n} \mathbb{I}\{a_{i,t,j} \neq 0\}$, which is defined here as the total number of nonzeros among the vectors $\mathbf{a}_{i,t} \in \mathbb{R}^n$ for $i \in \{0,\ldots,m\}$ and $t \in [T]$ in the robust optimization problem.

LEMMA 4. *The quantities* $\sum_{i=0}^{m} a_{i,t,j}\left(\frac{\lambda_i}{\sum_{i'\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i')=\pi^{\mathcal{A},s}(i)}\lambda_{i'}}\right)\zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}$ *for each* $(t,s,j)\notin\mathcal{A}$ *can be computed in a total of* $\mathcal{O}(T(Z+nK^{\mathcal{A}}))$ *time and* $\mathcal{O}(Z+nT^2)$ *space.*

In the production-inventory problem, we observe that $Z=\Theta(ET^2)$ due to constraint (1d). Therefore, it follows from Proposition 1 that Lemma 4 requires a total of $\mathcal{O}(nT^2)=\mathcal{O}(|\{(t,s,j):(t,s,j)\notin\mathcal{A}\}|)$ space for the production-inventory problem. Although this memory usage is quadratic in the number of periods T, we note that the memory usage can be further reduced from $\mathcal{O}(Z+nT^2)$ to $\mathcal{O}(Z)$ by reusing the allocated memory for representing each of the quantities $\sum_{i=0}^{m} a_{i,t,j}\left(\frac{\lambda_i}{\sum_{i'\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i')=\pi^{\mathcal{A},s}(i)}\lambda_{i'}}\right)\zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}$ for each $(t,s,j)\notin\mathcal{A}$. Hence, we conclude that the computer memory required for performing Step 2 and Step 3 in our algorithm can scale linearly in the computer memory required to encode the constraints of the underlying multi-period robust optimization problem.

4.4.2. Removing constraints from (D-$\mathcal{A}$). In Step 1 of each iteration of our algorithm, we must solve the linear optimization problem (D-$\mathcal{A}$). To improve the practical efficiency of solving this linear optimization problem (D-$\mathcal{A}$), we can remove inequality constraints that are not going to be binding at optimality. In particular, we remove constraints from (D-$\mathcal{A}$) using the following lemma:

LEMMA 5. *Define the following set of tuples:*

$$\mathcal{C}\triangleq\left\{(k,s):s\in[T],k\in[K^{\mathcal{A},s}],\left(b^{\mathcal{A},s}_{k,s},\left(a^{\mathcal{A},s}_{k,t,j}\right)_{t,j:(t,s,j)\in\mathcal{A}}\right)=\left(0,(0)_{t,j:(t,s,j)\in\mathcal{A}}\right)\right\}.$$

Then the following constraints can be simultaneously removed from the linear optimization problem (D-$\mathcal{A}$) *without changing the optimal objective value:*

$$\underline{D}_s\left(\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i\right)\le\zeta^{\mathcal{A},s}_{k,s}\le\bar{D}_s\left(\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i\right)\quad\forall(k,s)\in\mathcal{C}.\tag{2}$$

4.4.3. Updating the active set. In Step 3 of each iteration of our algorithm, we must choose a nonempty subset of tuples $\widetilde{\mathcal{A}}\subseteq\mathcal{A}^{\neq}$ to add to the current active set. In our implementation of the algorithm, we choose the subset of tuples by initializing $\widetilde{\mathcal{A}}\leftarrow\emptyset$. We then iterate over each of the decision rules $t\in[T]$ and $j\in[n]$, draw a random period $s\in\{1,\ldots,s\}$ uniformly over all of the periods s that satisfy $(t,s,j)\in\mathcal{A}^{\neq}$, and append $\widetilde{\mathcal{A}}\leftarrow\widetilde{\mathcal{A}}\cup\{(t,s,j)\}$. By following this procedure, the active set will increase by at most Tn components in each iteration.

To prevent the active set from becoming unnecessarily large, we employ a classical removal strategy from the cutting plane literature (Deletion Rule II from [16]). To employ this removal strategy, we maintain a record $v_{t,s,j}\in\mathbb{R}$ for each tuple $(t,s,j)\in\mathcal{A}$ of the optimal objective value of the optimization problem (LDR-$\mathcal{A}$) in the iteration in which the tuple (t,s,j) was added into the active set. At the end of Step 3 of each iteration of our algorithm, we remove each tuple $(t,s,j) \in \mathcal{A}$ from the active set for which the optimal objective value of the linear optimization problem (D-$\mathcal{A}$) is strictly less than $v_{t,s,j}$ and the dual variable $y_{t,s,j}$ associated with the equality constraint $\sum_{k=1}^{K^{\mathcal{A},s}} a_{k,t,j}^{\mathcal{A},s} \zeta_{k,s}^{\mathcal{A},s} = 0$ is equal to zero. It is established by [16] that the active set method with this removal strategy is guaranteed to converge to an optimal solution for (LDR) after finitely many iterations.

4.4.4. Initial active set. In the first iteration of our algorithm, we use the following initial active set:

$$\mathcal{A} = \{(t,1,j) : t \in [T], j \in [n]\} \cup \{(t,t,j) : t \in [T], j \in [n]\}.$$

We refer to the initial active set defined above as the *Markovian active set*. In the case of the Markovian active set, the optimization problem (LDR-$\mathcal{A}$) will yield linear decision rules where the decisions in each period $t \in [T]$ have the form $\mathbf{x}_t = \mathbf{y}_{t1} + \mathbf{y}_{tt}\zeta_t$.

4.4.5. Addressing infeasibility. In Step 1 of our algorithm in §4.3, we assumed that the optimal objective value of the linear optimization problem (D-$\mathcal{A}$) was finite. Here we propose a simple modification to Step 1 to address settings in which the optimal objective value of (D-$\mathcal{A}$) is not finite.

To motivate our modification, we begin by making the following observation.

Lemma 6. *If Assumption 1 holds, then* (D-$\mathcal{A}$) *always has a feasible solution.*

The above Lemma 6 shows that the optimal objective value of (D-$\mathcal{A}$) is not finite if and only if the optimal objective value of (D-$\mathcal{A}$) is unbounded. Therefore, for every active set $\mathcal{A}$, we observe that there always exists a sufficiently large constant $M \geq 0$ such that the optimal objective value of the following optimization problem is finite:

$$\begin{aligned}
\underset{\substack{\lambda_0,\ldots,\lambda_m \in \mathbb{R}, \\ \zeta_{k,s}^{\mathcal{A},s} \in \mathbb{R}\ \forall s \in [T], k \in [K^{\mathcal{A},s}]}}{\text{maximize}} \quad & -\sum_{i=1}^{m} c_i \lambda_i - \sum_{s=1}^{T} \sum_{k=1}^{K^{\mathcal{A},s}} b_{k,s}^{\mathcal{A},s} \zeta_{k,s}^{\mathcal{A},s} \\
\text{subject to} \quad & \sum_{k=1}^{K^{\mathcal{A},s}} a_{k,t,j}^{\mathcal{A},s} \zeta_{k,s}^{\mathcal{A},s} = 0 && \forall (t,s,j) \in \mathcal{A} \\
& \underline{D}_s \left(\sum_{i \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i) = k} \lambda_i \right) \leq \zeta_{k,s}^{\mathcal{A},s} \leq \bar{D}_s \left(\sum_{i \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i) = k} \lambda_i \right) && \forall s \in [T], k \in [K^{\mathcal{A},s}] \\
& \lambda_0 = 1 \\
& 0 \leq \lambda_i \leq M && \forall i \in \{1,\ldots,m\}.
\end{aligned} \tag{D-$\mathcal{A}$-M}$$

In view of the above observation, we now present our simple modification to Step 1 to address settings in which the optimal objective value of (D-$\mathcal{A}$) is not finite. In our modification of Step 1, we first solve the linear optimization problem (D-$\mathcal{A}$). If the optimal objective value of (D-$\mathcal{A}$) is finite, then we proceed to Step 2. Otherwise, we solve (D-$\mathcal{A}$-M) with greater and greater values of $M \geq 0$ until the linear optimization problem (D-$\mathcal{A}$-M) has a feasible solution, at which point we proceed to Step 2. This modification thus ensures that the algorithm can eventually reach Step 3 (in which new tuples will be added to the active set) regardless of the choice of the active set at the beginning of each iteration.

5. Numerical Experiments

5.1. Practical Implications of Theorem 1

To investigate the practical implications of Theorem 1, we perform numerical experiments that generalize those from [4, 12]. Specifically, our numerical experiments focus on instances of (LDR-1) in which the customer demand and production costs of a new product follow a cyclic pattern due to seasonality over a selling horizon of one year. Given a discretization of the selling season into T stages, the customer demand in each stage $t \in \{2, \ldots, T+1\}$ is given by

$$\phi_t = 1 + 0.5 \sin\left(\frac{2\pi(t-2)}{T}\right), \qquad \theta_t = 0.2, \qquad \mathcal{U}_t = \left[\frac{1000(1-\theta)\phi_t}{T/24}, \frac{1000(1+\theta)\phi_t}{T/24}\right], \tag{3}$$

where parameters ϕ_t and θ_t are interpreted, respectively, as a phase parameter, which captures seasonality, and a demand parameter, which controls the radius of the uncertainty sets. Given E factories available to the firm, the production costs and capacities for each stage $t \in [T]$ and each factory $e \in [E]$ are

$$c_{te} = \left(1 + \frac{e-1}{E-1}\right)\phi_t, \qquad p_{te} = \frac{567}{(T/24)\,(E/3)}, \qquad Q_e = \frac{13600}{E/3}, \tag{4}$$

and the capacities and initial inventory at the central warehouse are

$$V_{\min} = 500, \quad V_{\max} = 2000, \quad v_1 = 500. \tag{5}$$

In the case of $T = 24$ and $E = 3$, our parameters are equivalent to those of [4, §5] and [12, Table 1]. We follow this setup from [4, 12] to capture a realistic setting for the parameters of the problem, and our generalization of the setup from the prior work allows us to explore the impact of T on sparsity. We are particularly interested in settings where T grows large, in which case the robust optimization problem serves as an approximation of a continuous-review ordering system. For each value of T and E, we compute the optimal linear decision rules by solving the linear optimization formulation from [4, Equation (39)] using primal simplex method. Additional computational results are presented in Appendix H.

In Figure 1, we present the results of numerical experiments with $E = 3$ factories and varying numbers of periods T. The results of our numerical experiments provide two key takeaways.

Our first takeaway from Figure 1 is that the level of sparsity of the optimal linear decision rules obtained in the numerical experiments is very significant when the number of periods is large. Indeed, the left plot in Figure 1 shows that the number of nonzero parameters in the optimal linear decision rules decreases to 3% of the total number of parameters when inventory levels and production decisions are made twice per week over a selling horizon of one year. From a practical perspective, such a low density level of nonzeros in the optimal linear decision rules provides a clear motivation for the growing stream of research that harnesses sparsity to develop faster and more memory-efficient algorithms in various applications [3, 26, 2, 30].

Our second takeaway from Figure 1 is that Theorem 1 is found to be predictive of the level of sparsity in the optimal linear decision rules. In the right plot of Figure 1, we present the number of nonzero parameters in the optimal linear decision rules obtained using primal simplex method, the upper bound from Theorem 1, and the number of parameters in static decision rules (which serves as a natural lower bound). Indeed, the multiplicative gap between the upper bound from Theorem 1 and the lower bound from the static rule is very narrow (it is roughly a factor of 11). As we can see, the numbers of nonzero parameters in the optimal linear decision rules stay in this a narrow band and grow linearly in the number of stages. This result not only validates our theory but also showcases that the number of nonzero parameters in optimal linear decision rules does not grow sublinearly in T.

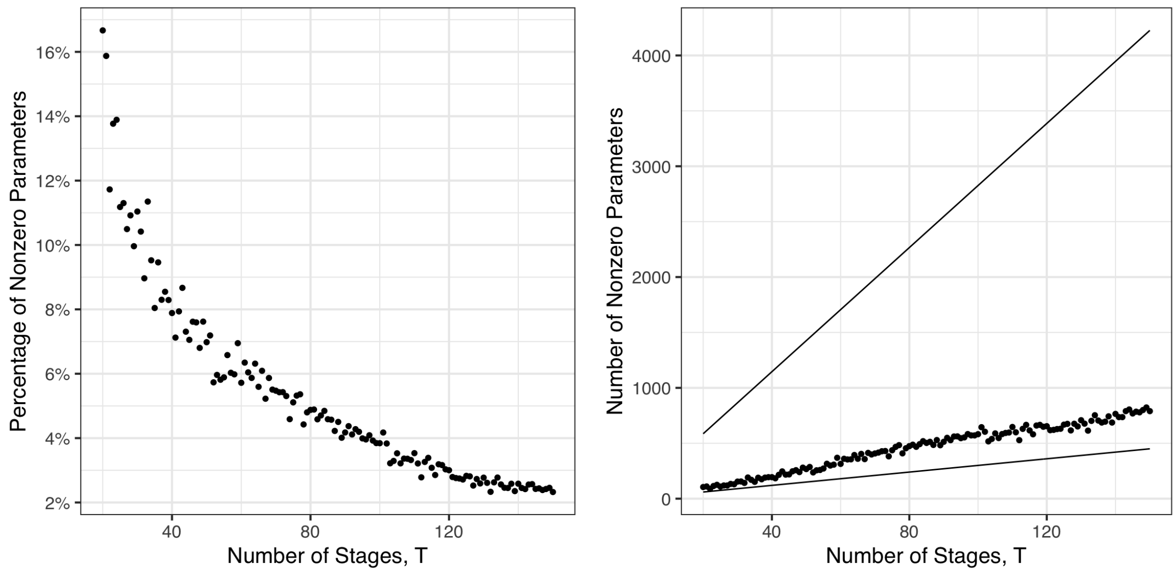
5.2. Value of Sparsity in Large-scale Problems

We next showcase the practical value of our novel reformulation technique and active set method from §4 in computing linear decision rules for robust optimization problems with hundreds of time periods and as many as fifty decision and state variables in each time period.

In greater detail, our numerical experiments in this subsection focus on the same problem setup of production-inventory problems that is described at the beginning of §5.1. That is, our numerical experiments in this subsection focus on instances of (LDR-1) in which the customer demand and production constraints follow the same setup as given in lines (3)-(5). In contrast to the previous subsection, our analysis here focuses on the regime where the number of periods T and the number of factories E grow large.[7] For this regime, we compare the practical efficiency of our novel reformulation technique and active set method from §4 to the practical efficiency of using the robust counterpart technique. In our implementation of the active set method, we apply all of the improvements from §4.4 and solve (D-$\mathcal{A}$) in each iteration with Gurobi using the dual simplex

[7] We note for this application that the dimension of the state and decision spaces in each period scale linearly in the number of factories. Specifically, for the production-inventory problem, the state space is determined by the inventory at the warehouse and the cumulative production quantity at each factory in each time period. The decisions include the production quantity at each factory in each time period.

Figure 1 **Sparsity of optimal linear decision rules for production-inventory problem, $E = 3$.**



Note. Each point represents the optimal linear decision rules computed for the corresponding number of stages T and for $E = 3$ factories. Left figure shows the percentage of parameters of optimal linear decision rules which are nonzero. Right figure shows the number of nonzero parameters in optimal linear decision rules compared to the upper bound from Theorem 1 (top solid black line) and the number of parameters in static decision rules (bottom solid black line).

method. In our implementation of the robust counterpart, we solve the dual linear optimization formulation of the robust counterpart of (LDR-1) (see problem (D) in Appendix F) with Gurobi using the barrier method.[8]

In Figure 2, we compare the computation time of our novel fromulation technique coupled with the active set method to the computation time of the robust counterpart technique for approximately solving production-inventory problems with linear decision rules. These experiments focus on the scalability of these methods with $E = 5$ factories and $T \in \{48, 96, 144, 192, 240\}$ time periods. Figure 2 reports on the computation time for the active set method to obtain feasible linear decision rules that are within 10%, 1%, and 0.1% of optimal. To perform a direct comparison, we terminate the barrier method when the primal-dual gap reaches 10%, 1%, and 0.1%. To make the experimental setup more favorable to the barrier method, we disable the crossover functionality of the barrier method to reduce its computation time.[9] Optimality gaps for the active set method are obtained

[8] We here present the comparison with the Gurobi barrier method because it is the most effective algorithm of Gurobi in numerical performance. Comparisons of computation times for solving the robust counterpart with primal simplex, dual simplex, and barrier method are found in Appendix H.

[9] Since Gurobi utilizes infeasible-interior-point method in their barrier implementation, with such tolerance level, it returns neither a primal feasible solution nor a dual feasible solution to the linear optimization problem obtained by taking the robust counterpart. Rather, the barrier method only produces bounds on the optimal objective value of the robust counterpart and near feasible solutions. In contrast, our approach outputs a primal feasible solution whenever (D-$\mathcal{A}$) has an optimal solution (see §4.4.5). In all numerical experiments shown in this subsection, (D-$\mathcal{A}$) provided an optimal solution in each iteration of the active set method.

by comparing the objective value of the feasible linear decision rules obtained by the active set method to the optimal objective value obtained by solving the robust counterpart to optimality.

The results of these experiments reveal that our novel reformulation technique coupled with the active set method offers significant improvements in computation time compared to the robust counterpart technique. Indeed, for problems with $T = 240$ time periods and $E = 5$ factories, our approach is over 170x faster in computing linear decision rules that are within 10% of optimal, and over 32x faster in computing linear decision rules that are within 1% of optimal. More broadly, the computation times for our novel reformulation technique and active set method are a very small fraction of the computation times for the robust counterpart for all problem sizes, and the gap in computation times becomes increasingly pronounced as the number of time periods increases.

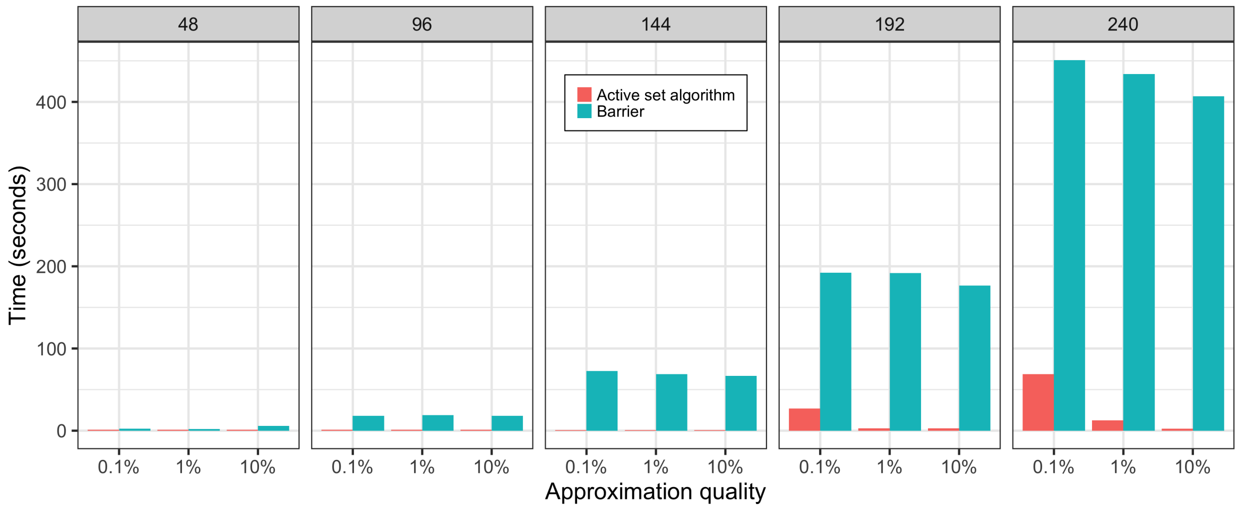
In Figure 3, we show the relationship between computation time and objective value of linear decision rules obtained by our novel reformulation technique and active set method in experiments with $T = 240$ time periods and $E \in \{10, 20, 30, 40, 50\}$ factories. The takeaways from this figure are three-fold. First, we observe from Figure 3 that sparse linear decision rules with the Markovian active set (see §4.4.4) are not optimal in general, as the objective value of the linear decision rules found by the active set method decreases as the number of iterations increases. Second, Figure 3 shows that our novel reformulation technique and active set method scales to applications with high-dimensional state spaces and hundreds of time periods. In particular, the large numbers of factories here leads to a number of decision and state variables that prevent methods such as robust dual dynamic programming [19] from being applied.[10] Third, the practical efficiency of our approach is particularly notable given that the size of the linear optimization problem obtained by the robust counterpart for $T = 240$ time periods and $E = 50$ factories is well beyond the capabilities of classic linear optimization solvers. Indeed, for this problem setting, the (primal) linear optimization problem obtained by the robust counterpart technique would require 13,281,993 decision variables to represent parameters of the linear decision rule and the auxiliary variables, and would require 5,936,502 constraints and 250,895,943 nonzeros.

6. Extensions

In this section, we show that our proof techniques from §3 can be extended to establish sparsity guarantees for other classes of robust optimization problems. These extensions thus demonstrate that the sparsity guarantees and proof techniques developed in this paper are not exclusive to the class of production-inventory problems from Ben-Tal et al. [4], and they provide a starting point for using our proof techniques to establish sparsity guarantees for other applications.

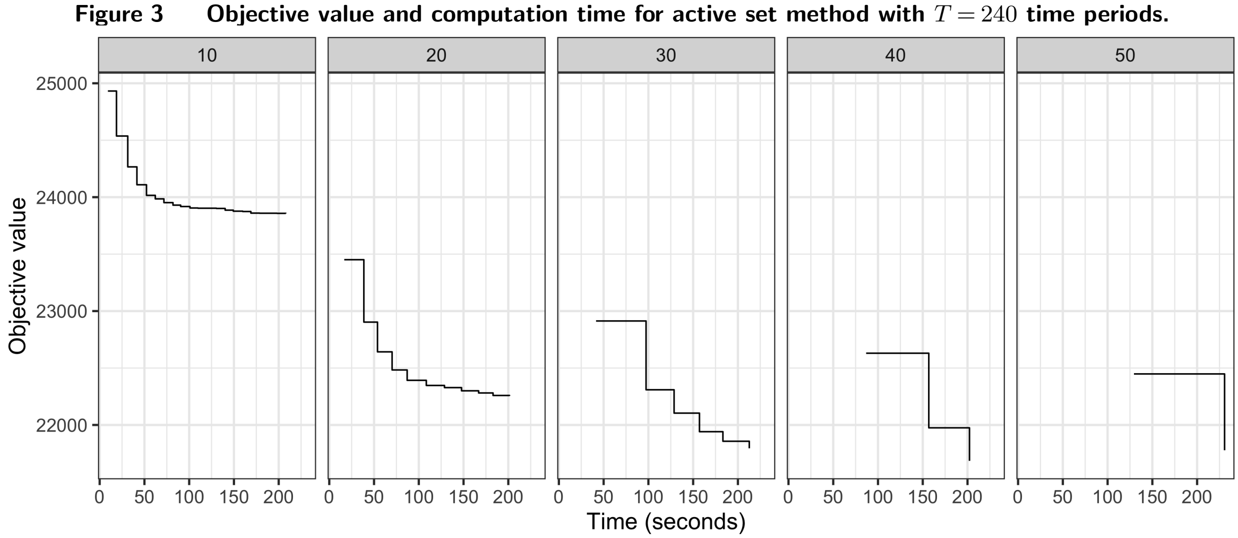
[10] Numerical experiments from [19, p.827-p.828] suggest that the robust dual dynamic programming technique scales gracefully with respect to the number of time periods but scales slowly with respect to the number of decision and state variables in each time period.

Figure 2 Computation times for active set method and robust counterpart with $E = 5$ factories.



Note. Results shown for experiments with $E = 5$ factories and $T \in \{48, 96, 144, 192, 240\}$ time periods. Blue bars show the computation times (in seconds) for the barrier method applied to the robust counterpart to have a primal dual gap that is within 10%, 1%, and 0.1%. Barrier method is run without performing crossover, and so the barrier does not return a primal or dual feasible solution. Red bars show the computation time (in seconds) for the active set method to obtain a feasible linear decision rule that is within 10%, 1%, and 0.1% of optimal. Optimality gaps are obtained by comparing the objective value of the linear decision rules obtained by the active set method to the optimal objective value obtained by solving the robust counterpart to optimality.

Figure 3 Objective value and computation time for active set method with $T = 240$ time periods.



Note. Results shown for experiments with $T = 240$ time periods and $E \in \{10, 20, 30, 40, 50\}$ factories. Black lines show the objective value of the linear decision rule found by the active set method as a function of computation time.

Inventory Management with Lead Times. As discussed in the literature, the class of production-inventory problems from Ben-Tal et al. [4] admits a number of natural variations. For example, a common variation introduces lead-times to the factories, whereby the production quantity $x_{te} \geq 0$

at factory $e \in [E]$ on period $t \in [T]$ will not become available at the central warehouse until stage $t + \delta_e$, with δ_e denoting the lead time for factory $e \in [E]$. The resulting cost function for this generalization of (1a)-(1d) can thus be written as

$$C(\mathbf{x}_1,\dots,\mathbf{x}_T,\zeta_1,\dots,\zeta_{T+1}) = \sum_{e=1}^{E}\sum_{t=1}^{T} c_{te}x_{te} \tag{2a}$$

$$\text{subject to} \quad \sum_{t=1}^{T} x_{te} \le Q_e \qquad \forall e \in [E] \tag{2b}$$

$$0 \le x_{te} \le p_{te} \qquad \forall e \in [E], t \in [T] \tag{2c}$$

$$V_{\min} \le v_1 + \sum_{e=1}^{E}\sum_{\ell=1}^{t-\delta_e} x_{\ell e} - \sum_{s=2}^{t+1}\zeta_s \le V_{\max} \qquad \forall t \in [T]. \tag{2d}$$

By applying the proof techniques developed in §3, it can be shown that the $\mathcal{O}(ET)$ sparsity result for optimal linear decision rules is retained:

THEOREM 2. *Consider a cost function of the form* (2a)-(2d) *and let Assumption 1 hold. Suppose that $c_{te} > 0$ for every $t \in [T]$ and $e \in [E]$, and let $\delta \triangleq \min_{e\in[E]} \delta_e$ denote the minimum lead time. Then there exists an optimal solution $\bar{\mathbf{y}}$ for* (LDR) *that satisfies $\|\bar{\mathbf{y}}\|_0 \le 2 + 8E + 10T + 6E(T-\delta)$.*

We observe from Theorem 2 that if the minimum lead time across the factories $\delta \triangleq \min_{e\in[E]} \delta_e$ converges to T, then the number of nonzero parameters in the optimal linear decision rules converges to $\mathcal{O}(E+T)$. Hence, the upper bound provided by Theorem 2 conforms with the observation that additional inventory should only purchased in the early time periods when the lead times of all of the factories are close to the total number of time periods T.

Dynamic Newsvendor Problem. A significant research effort in the robust optimization literature has been dedicated to dynamic newsvendor problems with box uncertainty sets; see [5, 9, 24]. This class of dynamic newsvendor problems is characterized by a single factory and nonlinear convex cost functions which capture the holding and backorder costs for inventory at the warehouse. Specifically, the cost function for these dynamic newsvendor problems is given by

$$C(x_1,\dots,x_T,\zeta_1,\dots,\zeta_{T+1}) = \sum_{t=1}^{T}\left(c_t x_t + h_t\left[v_1 + \sum_{s=1}^{t} x_s - \sum_{s=2}^{t+1}\zeta_s\right]^+ + b_t\left[-v_1 - \sum_{s=1}^{t} x_s + \sum_{s=2}^{t+1}\zeta_s\right]^+\right) \tag{3a}$$

$$\text{subject to} \quad 0 \le x_t \le p_t \quad \forall t \in [T], \tag{3b}$$

where the firm begins at the start of the selling season with an initial inventory of v_1 units of product, and, in each stage $t \in [T]$, the firm can decide to produce an additional $x_t \in [0, p_t]$ units of product from a single factory at a cost of c_t per unit. The customer demands for the product are denoted by $\zeta_2,\dots,\zeta_{T+1} \in \mathbb{R}$, and the holding and backorder costs for the inventory at the end of each period $t \in [T]$ are given by h_t and b_t.

A fundamental result for this class of dynamic newsvendor problems is that if Assumption 1 holds, then the optimal objective values for (RO) and (LDR) with cost function (3a)-(3b) are equal, that is, linear decision rules are optimal control policies [9, Theorem 3.1]. However, with the exception of guarantees on whether of the parameters of optimal linear decision rules are positive or negative ([9, Proposition 5.1], [24, Lemma 5]), the structure of optimal linear decision rules for the production quantities in dynamic newsvendor problems has been unknown. Using our proof techniques from §3, we establish for the first time that optimal linear decision rules for the class of dynamic newsvendor problems are not only optimal for (RO); they are also sparse.

THEOREM 3. *Consider a cost function of the form* (3a)-(3b) *and let Assumption 1 hold. Then there exists an optimal solution* $\bar{\mathbf{y}}$ *for* (LDR) *that satisfies* $\|\bar{\mathbf{y}}\|_0 \le 10 + 12T$.

We emphasize that Theorem 3 establishes that $10+12T$ is an upper bound on the *total* number of nonzero parameters for optimal linear decision rules for the production quantities across *all* of the time periods of the dynamic newsvendor problem. Theorem 3 thus implies that there always exists an optimal linear decision rule for the production quantities in the dynamic newsvendor problem in which the *average* number of nonzero parameters for an optimal linear decision rule in each time period is upper bounded by $(10+12T)/T = 12 + \frac{10}{T}$. Since optimal linear decision rules are optimal control policies for the dynamic newsvendor problem [9, Theorem 3.1], Theorem 3 thus establishes for the first time that the optimal production quantity decision in each time period can be made using an average of $12 + \frac{10}{T} = \mathcal{O}(1)$ of the demand realizations observed in the past time periods.

Inventory Management with Non-Box Uncertainty Sets. The previous results show that there exist sparse optimal linear decision rules in production-inventory problems with box uncertainty sets. Here we showcase how our proof techniques can be extended to production-inventory problems with non-box uncertainty sets. Specifically, our results here focus on the class of production-inventory problems from Ben-Tal et al. [4] in which the demand is chosen from a non-separable budget uncertainty set of the form

$$\mathcal{U} \triangleq \left\{ \boldsymbol{\zeta} \equiv (\zeta_1, \ldots, \zeta_{T+1}) : \begin{array}{ll} \zeta_1 = 1 & \\ \underline{D}_s \le \zeta_s \le \bar{D}_s & \forall s \in \{2, \ldots, T+1\} \\ \displaystyle\sum_{s=2}^{T+1} \frac{\zeta_s - \underline{D}_s}{\bar{D}_s - \underline{D}_s} \le T - k & \end{array} \right\}. \tag{6}$$

The difference compared with the original model is that the budget uncertainty is no longer separable over time period t due to the budget constraint $\sum_{s=2}^{T+1} \frac{\zeta_s - \underline{D}_s}{\bar{D}_s - \underline{D}_s} \le T - k$. This budget constraint means that among the T demand variables ζ_t, there are at most $T-k$ demands that choose their maximum values. Now we consider the linear decision rule approximation of the class of production-inventory problems with budget uncertainty set:

$$
\begin{aligned}
&\underset{\mathbf{y}_{t,1},\dots,\mathbf{y}_{t,t}\in\mathbb{R}^E:\,\forall t\in[T]}{\text{minimize}} && \max_{\boldsymbol{\zeta}\equiv(\zeta_1,\dots,\zeta_{T+1})\in\mathcal{U}} \left\{ \sum_{t=1}^{T}\sum_{e=1}^{E} c_{te}\left(\sum_{s=1}^{t} y_{t,s,e}\zeta_s\right)\right\} && \\
&\text{subject to} && \sum_{t=1}^{T}\left(\sum_{s=1}^{t} y_{t,s,e}\zeta_s\right) \le Q_e && \forall e\in[E] \\
& && 0 \le \left(\sum_{s=1}^{t} y_{t,s,e}\zeta_s\right) \le p_{te} && \forall e\in[E],\, t\in[T] \\
& && V_{\min} \le v_1 + \sum_{\ell=1}^{t}\sum_{e=1}^{E}\left(\sum_{s=1}^{\ell} y_{\ell,s,e}\zeta_s\right) - \sum_{s=2}^{t+1}\zeta_s \le V_{\max} && \forall t\in[T] \\
& && \forall \boldsymbol{\zeta}\equiv(\zeta_1,\dots,\zeta_{T+1})\in\mathcal{U}.
\end{aligned}
\tag{LDR-4}
$$

In the following theorem, we show that our sparsity guarantees can be extended to optimal linear decision rules for the optimization problem (LDR-4). More broadly, the following theorem establishes that our proof techniques for establishing sparsity guarantees for optimal linear decision rules can be extended to applications in which the uncertainty sets are dependent across time periods.

THEOREM 4. *If* (LDR-4) *is feasible and has a finite optimal objective value, then there exists an optimal solution* $\bar{\mathbf{y}}$ *for* (LDR-4) *that satisfies* $\|\bar{\mathbf{y}}\|_0 \le 2 + 8E + 10T + 6ET + 12kT$.

The proof of Theorem 4 can be found in Appendix D. The basic idea of the proof technique is to reduce the problem with a non-box uncertainty set to one with a box uncertainty set by dualizing the coupling constraint (i.e., the budget constraint for this example). We then utilize the proof framework presented in §3.

We emphasize that Theorem 4 only establishes the existence of a sparse solution of (LDR-4) in the case when k is small. It is an open question whether sparsity guarantees hold when k is large relative to T, as is typically recommended in the literature. Nevertheless, we hope the proof of Theorem 4 can inspire the community to develop more examples of robust optimization problems with many time periods where the existence of sparse optimal linear decision rules is guaranteed.

7. Conclusion and Future Research

In this paper, we proved that there exist sparse optimal linear decision rules for the widely-studied class of production-inventory problems from Ben-Tal et al. [4], and that the number of nonzero parameters in sparse optimal linear decision rules grows linearly in the number of time periods. Apart from their theoretical significance, our sparsity guarantees were shown to give rise to new practical algorithms for computing linear decision rules in robust optimization problems with huge numbers of time periods that were out-of-reach by previous state-of-the-art algorithms. More broadly, we hope that this paper offers new motivations to firms for using linear decision rules in real-world large-scale applications, as well as opens up new research directions at the interface between the practice and theory of robust optimization. These research directions include harnessing the structure of sparse optimal linear decision rules to design effective algorithms for computing optimal linear decision rules in real-world large-scale applications, studying the tradeoffs between interpretability of linear decision rules and Pareto efficiency [23, 11], analyzing the implications of sparsity of linear decision rules on time inconsistency in risk-averse planning problems, and investigating whether sparsity can be combined in Fourier-Motzkin elimination to obtain high-accuracy approximations in robust optimization problems with many recourse decisions [32].

References

[1] Amir Ardestani-Jaafari and Erick Delage. Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research*, 64(2):474–494, 2016.

[2] Amir Ardestani-Jaafari and Erick Delage. The value of flexibility in robust location–transportation problems. *Transportation Science*, 52(1):189–209, 2018.

[3] Chaithanya Bandi, Eojin Han, and Omid Nohadani. Sustainable inventory with robust periodic-affine policies and application to medical supply chains. *Management Science*, 65(10):4636–4655, 2019.

[4] Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[5] Aharon Ben-Tal, Boaz Golany, Arkadi Nemirovski, and Jean-Philippe Vial. Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing & Service Operations Management*, 7(3):248–271, 2005.

[6] Aharon Ben-Tal, Elad Hazan, Tomer Koren, and Shie Mannor. Oracle-based robust optimization via online learning. *Operations Research*, 63(3):628–638, 2015.

[7] Dimitris Bertsimas and Vineet Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming*, 134(2):491–531, 2012.

[8] Dimitris Bertsimas and John N Tsitsiklis. *Introduction to Linear Optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

[9] Dimitris Bertsimas, Dan A Iancu, and Pablo A Parrilo. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010.

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

[11] Dimitris Bertsimas, Stefan ten Eikelder, Dick den Hertog, and Nikolaos Trichakis. Pareto adaptive robust optimality via a fourier-motzkin elimination lens. *arXiv preprint arXiv:2012.04419*, 2020.

[12] Frans JCT de Ruiter, Ruud Brekelmans, and Dick den Hertog. The impact of the existence of multiple adjustable robust solutions. *Mathematical Programming*, 160(1):531–545, 2016.

[13] Erick Delage and Dan A Iancu. Robust multistage decision making. In *The Operations Research Revolution*, pages 20–46. INFORMS, 2015.

[14] Stefan Eikelder, Ali Ajdari, Thomas Bortfeld, Dick den Hertog, et al. Adjustable robust treatment-length optimization in radiation therapy. *arXiv preprint arXiv:1906.12116*, 2019.

[15] Omar El Housni and Vineet Goyal. On the optimality of affine policies for budgeted uncertainty sets. *Mathematics of Operations Research*, 2021.

[16] Jack Elzinga and Thomas G Moore. A central cutting plane algorithm for the convex programming problem. *Mathematical Programming*, 8:134–145, 1975.

[17] Charles Gauvin, Erick Delage, and Michel Gendreau. Decision rule approximations for the risk averse reservoir management problem. *European Journal of Operational Research*, 261(1):317–336, 2017.

[18] Angelos Georghiou, Daniel Kuhn, and Wolfram Wiesemann. The decision rule approach to optimization under uncertainty: methodology and applications. *Computational Management Science*, 16(4):545–576, 2019.

[19] Angelos Georghiou, Angelos Tsoukalas, and Wolfram Wiesemann. Robust dual dynamic programming. *Operations Research*, 67(3):813–830, 2019.

[20] Angelos Georghiou, Angelos Tsoukalas, and Wolfram Wiesemann. On the optimality of affine decision rules in robust and distributionally robust optimization. *Available at Optimization Online*, 2021.

[21] Chrysanthos E Gounaris, Wolfram Wiesemann, and Christodoulos A Floudas. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research*, 61(3):677–693, 2013.

[22] Nam Ho-Nguyen and Fatma Kılınç-Karzan. Online first-order framework for robust convex optimization. *Operations Research*, 66(6):1670–1692, 2018.

[23] Dan A Iancu and Nikolaos Trichakis. Pareto efficiency in robust optimization. *Management Science*, 60(1):130–147, 2014.

[24] Dan A Iancu, Mayank Sharma, and Maxim Sviridenko. Supermodularity and affine policies in dynamic robust optimization. *Operations Research*, 61(4):941–956, 2013.

[25] Hongming Li, Erick Delage, Ning Zhu, Michael Pinedoc, and Shoufeng Ma. Distributional robustness and inequity mitigation in disaster preparedness of humanitarian operations. *Available at Optimization Online*, 2022.

[26] Alvaro Lorca, X Andy Sun, Eugene Litvinov, and Tongxin Zheng. Multistage adaptive robust optimization for the unit commitment problem. *Operations Research*, 64(1):32–51, 2016.

[27] Krzysztof Postek and Shimrit Shtern. First-order algorithms for robust optimization problems via convex-concave saddle-point lagrangian reformulation. *arXiv preprint arXiv:2101.02669*, 2021.

[28] Sajad Aliakbari Sani, Olivier Bahn, and Erick Delage. Affine decision rule approximation to immunize against demand response uncertainty in smart grids' capacity planning. *European Journal of Operational Research*, 2020.

[29] David Simchi-Levi, Nikolaos Trichakis, and Peter Yun Zhang. Designing response supply chain against bioattacks. *Operations Research*, 67(5):1246–1268, 2019.

[30] Shanshan Wang and Erick Delage. A column generation scheme for distributionally robust multi-item newsvendor problems. *INFORMS Journal on Computing*, 2023.

[31] İhsan Yanıkoğlu, Bram L Gorissen, and Dick den Hertog. A survey of adjustable robust optimization. *European Journal of Operational Research*, 277(3):799–813, 2019.

[32] Jianzhe Zhen, Dick Den Hertog, and Melvyn Sim. Adjustable robust optimization via fourier–motzkin elimination. *Operations Research*, 66(4):1086–1100, 2018.

Technical Proofs and Additional Results

Appendix A: Proofs of Lemmas 1, 2, and 3

Throughout the proofs in Appendix A, we will denote the set of feasible solutions to (LDR-G) by

$$\mathcal{Y} \triangleq \left\{ (\mathbf{y}, c_0) : \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} \left\{ \sum_{t=1}^T \mathbf{a}_{i,t}^\intercal \left(\sum_{s=1}^t \mathbf{y}_{t,s} \zeta_s \right) - \sum_{t=1}^T b_{i,t} \zeta_t \right\} \le c_i \;\; \forall i \in \{0, \ldots, m\} \right\}. \tag{EC.1}$$

We remark that it follows readily from Assumption 1 that $\mathcal{Y}$ is a nonempty convex polyhedron.

Proof of Lemma 1. Our proof of Lemma 1 is based on contradiction. Indeed, for the sake of developing a contradiction, suppose that the set $\mathcal{Y}$ does not have an extreme point. Since the set $\mathcal{Y}$ is a polyhedron, it follows from the supposition that $\mathcal{Y}$ has no extreme points that $\mathcal{Y}$ must contain a line. In other words, there must exist a solution $(\mathbf{y}^0, c_0^0) \in \mathcal{Y}$ and a direction $(\mathbf{d}, r) \neq (\mathbf{0}, 0)$ such that $(\mathbf{y}^0, c_0^0) + \alpha(\mathbf{d}, r) \in \mathcal{Y}$ for all $\alpha \in \mathbb{R}$.

We begin by showing under the supposition that $\mathcal{Y}$ has no extreme points that $\mathbf{d} \neq \mathbf{0}$. Indeed, we recall from Assumption 1 that the optimal objective value of (LDR) is finite. Moreover, we recall that $(\mathbf{y}^0, c_0^0) + \alpha(\mathbf{d}, r) \in \mathcal{Y}$ for all $\alpha \in \mathbb{R}$. Since the objective value of $(\mathbf{y}^0, c_0^0) + \alpha(\mathbf{d}, r)$ is equal to $c_0^0 + \alpha r$, it must be the case that $r = 0$. Therefore, it follows from the fact that $(\mathbf{d}, r) \neq (\mathbf{0}, 0)$ that $\mathbf{d} \neq \mathbf{0}$.

Next, it follows from Assumption 2 that $(\mathbf{y}^0, c_0^0) + \alpha(\mathbf{d}, r) \in \mathcal{Y}$ for all $\alpha \in \mathbb{R}$ must satisfy

$$\sum_{s=1}^t \left(\mathbf{y}_{t,s}^0 + \alpha \mathbf{d}_{t,s} \right) \zeta_s \ge \mathbf{0} \qquad \forall \alpha \in \mathbb{R},\, t \in [T],\, \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T.$$

Since the above inequalities hold for all $\alpha \in \mathbb{R}$, it follows from algebra that

$$\sum_{s=1}^t \mathbf{d}_{t,s} \zeta_s = \mathbf{0} \qquad \forall t \in [T],\, \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T. \tag{EC.2}$$

Recall that the uncertainty sets are intervals of the form $\mathcal{U}_1 \triangleq [\underline{D}_1, \bar{D}_1], \ldots, \mathcal{U}_T \triangleq [\underline{D}_T, \bar{D}_T]$, where $\underline{D}_1 = \bar{D}_1 = 1$ and $\underline{D}_t < \bar{D}_t$ for all $t \in \{2, \ldots, T\}$. Therefore, we observe that the equalities on line (EC.2) imply that the equality $\mathbf{d}_{t,s} = \mathbf{0}$ must hold for all $s \in \{2, \ldots, T\}$ and $t \in \{s, \ldots, T\}$. Moreover, it follows from line (EC.2), from the fact that $\underline{D}_1 = \bar{D}_1 = 1$, and from the fact that $\mathbf{d}_{t,s} = \mathbf{0}$ for all $s \in \{2, \ldots, T\}$ and $t \in \{s, \ldots, T\}$ that the equality $\mathbf{d}_{t,1} = -\sum_{s=2}^t \mathbf{d}_{t,s} \zeta_s = \mathbf{0}$ must hold for all $t \in \{1, \ldots, T\}$. We have thus shown that $\mathbf{d} = \mathbf{0}$, which contradicts the supposition that the set of optimal solutions $\mathcal{Y}$ has a line. This concludes our proof that $\mathcal{Y}$ has at least one extreme point. □

Proof of Lemma 2. Let $(\bar{\mathbf{y}}, \bar{c}_0)$ denote an extreme point of the set $\mathcal{Y}$ for (LDR-G). We first observe from the definitions of the uncertainty sets $\mathcal{U}_1 = [\underline{D}_1, \bar{D}_1], \ldots, \mathcal{U}_T = [\underline{D}_T, \bar{D}_T]$ and from algebra that $\mathcal{Y}$ can be written equivalently as

$$\begin{aligned}
\mathcal{Y} &= \left\{ (\mathbf{y}, c_0) : \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} \left\{ \sum_{t=1}^T \mathbf{a}_{i,t}^\intercal \left(\sum_{s=1}^t \mathbf{y}_{t,s} \zeta_s \right) - \sum_{t=1}^T b_{i,t} \zeta_t \right\} \le c_i \;\; \forall i \in \{0, \ldots, m\} \right\} \\
&= \left\{ (\mathbf{y}, c_0) : \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} \left\{ \sum_{s=1}^T \left(-b_{i,s} + \sum_{t=s}^T \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) \zeta_s \right\} \le c_i \;\; \forall i \in \{0, \ldots, m\} \right\} \\
&= \left\{ (\mathbf{y}, c_0) : \sum_{s=1}^T \max \left\{ \left(-b_{i,s} + \sum_{t=s}^T \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) \underline{D}_s,\; \left(-b_{i,s} + \sum_{t=s}^T \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) \bar{D}_s \right\} \le c_i \;\; \forall i \in \{0, \ldots, m\} \right\} \\
&= \left\{ (\mathbf{y}, c_0) : \sum_{s=1}^T \left(-b_{i,s} + \sum_{t=s}^T \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) D_s^* \le c_i \;\; \forall D_s^* \in \left\{ \underline{D}_s, \bar{D}_s \right\},\; s \in [T], \text{ and } i \in \{0, \ldots, m\} \right\}.
\end{aligned}$$

Since $(\bar{\mathbf{y}}, \bar{c}_0)$ is an extreme point of the set $\mathcal{Y}$, and since the set $\mathcal{Y}$ is a polyhedron, we observe that $(\bar{\mathbf{y}}, \bar{c}_0)$ is a basic feasible solution of $\mathcal{Y}$. In other words, $(\bar{\mathbf{y}}, \bar{c}_0)$ must be the unique solution to the system of constraints in $\mathcal{Y}$ which are active constraints at $(\bar{\mathbf{y}}, \bar{c}_0)$. Let $\mathcal{I}$ denote the subset of $\{0, \ldots, m\}$ corresponding to the active constraints at the basic feasible solution $(\bar{\mathbf{y}}, \bar{c}_0)$, that is, let

$$\mathcal{I} \triangleq \{0\} \cup \left\{ i : \text{there exists } D_s^* \in \left\{ \underline{D}_s, \bar{D}_s \right\} \text{ for each stage } s \in [T] \text{ such that } \sum_{s=1}^{T} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} \right) D_s^* = c_i \right\}.$$

Since $(\bar{\mathbf{y}}, \bar{c}_0)$ is an element of $\mathcal{Y}$, we observe for each $i \in \mathcal{I}$ that the equality

$$\sum_{s=1}^{T} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} \right) D_s^* = \begin{cases} c_i, & \text{if } i \neq 0, \\ \bar{c}_0, & \text{if } i = 0 \end{cases}$$

is satisfied whenever the following equality holds for each $s \in [T]$:

$$D_s^* = \begin{cases} \bar{D}_s, & \text{if } -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} > 0, \\ \underline{D}_s, & \text{if } -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} < 0, \\ \underline{D}_s \text{ or } \bar{D}_s & \text{if } -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} = 0. \end{cases}$$

Therefore, we conclude from the above observation that the set of active constraints at the basic feasible solution $(\bar{\mathbf{y}}, \bar{c}_0)$ is given by the system of equalities

$$\sum_{s \in \mathcal{T}_i^>} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) \bar{D}_s + \sum_{s \in \mathcal{T}_i^<} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) \underline{D}_s + \sum_{s \in \mathcal{T}_i^=} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) D_s^* = c_i$$
$$\forall i \in \mathcal{I} \text{ and } D_s^* \in \left\{ \underline{D}_s, \bar{D}_s \right\} \forall s \in \mathcal{T}_i^=, \tag{EC.3}$$

where we define the disjoint index sets $\mathcal{T}_i^>$, $\mathcal{T}_i^=$, $\mathcal{T}_i^<$ for each $i \in \mathcal{I}$ as

$$\mathcal{T}_i^> \triangleq \left\{ s : -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} > 0 \right\}, \quad \mathcal{T}_i^= \triangleq \left\{ s : -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} = 0 \right\},$$
$$\mathcal{T}_i^< \triangleq \left\{ s : -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} < 0 \right\}.$$

We observe that the system of equalities (EC.3) can be rewritten as

$$\sum_{s \in \mathcal{T}_i^>} \left(-b_{0,s} + \sum_{t=s}^{T} \mathbf{a}_{0,t}^\intercal \mathbf{y}_{t,s} \right) \bar{D}_s + \sum_{s \in \mathcal{T}_i^<} \left(-b_{0,s} + \sum_{t=s}^{T} \mathbf{a}_{0,t}^\intercal \mathbf{y}_{t,s} \right) \underline{D}_s = c_i \qquad \forall i \in \mathcal{I}$$
$$\sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} = b_{i,s} \qquad \forall i \in \mathcal{I}, s \in \mathcal{T}_i^=,$$

which concludes our proof of Lemma 2. □

Proof of Lemma 3. The overarching idea of the proof of Lemma 3 is that the equations $\mathbf{P}_2 \mathbf{z} = \mathbf{0}$ from (S-2) can be eliminated by eliminating the corresponding variables.

We begin by making several assumptions without loss of generality. First, we recall that $\mathbf{P}_2$ is a matrix with more columns than rows. We henceforth assume without loss of generality that $\mathbf{P}_2$ has linearly independent rows (otherwise we may remove the rows which are linearly dependent without changing the set of feasible solutions to the system of equations (S-1)-(S-3)). We also assume for each row $i \in [m_2]$ of the matrix $\mathbf{P}_2$ that there exists a column $j \in [n] \setminus \mathcal{N}$ such that $p_{2,i,j} \neq 0$ (otherwise we may remove row i without changing the set of feasible solutions to the system of equations (S-1)-(S-3)).

We will also use the following notation. For each row $i \in [m_2]$ of the matrix $\mathbf{P}_2$, let $j_i \in [n] \setminus \mathcal{N}$ denote the smallest column for which $p_{2,i,j} \neq 0$, and let $\mathcal{S}_i \triangleq \{j \in [n] \setminus \{j_i\} : p_{2,i,j} \neq 0\}$ denote the remaining columns for which the i-th row of $\mathbf{P}_2$ has nonzero entries. Finally, let $\mathcal{S} \triangleq \{j_1, \ldots, j_{m_2}\}$, in which case it follows from the assumption that $\sum_{i=1}^{m_2} \mathbb{I}\{p_{2,i,j} \neq 0\} \leq 1$ for each $j \in [n]$ that $\mathcal{S}, \mathcal{S}_1, \ldots, \mathcal{S}_{m_2}$ are disjoint sets.

We now perform a substitution of variables to eliminate the constraints (S-2) from the system of equations. Specifically, by performing the substitution $z_{j_i} = -\sum_{j \in \mathcal{S}_i} z_j$ for each $i \in [m_2]$, we conclude that $(\bar{z}_j : j \in [n] \setminus \mathcal{S})$ is the unique solution to the following system of equations:

$$\sum_{j \in [n] \setminus \mathcal{S}} \mathbf{P}_{1,j} z_j + \sum_{i=1}^{m_2} \mathbf{P}_{1,j_i} \left(-\sum_{j \in \mathcal{S}_i} z_j \right) = \mathbf{q}$$
$$z_j = 0 \quad \forall j \in \mathcal{N}.$$

Since $(\bar{z}_j : j \in [n] \setminus \mathcal{S})$ is the unique solution to the above system, and since $\mathbf{q} \in \mathbb{R}^{m_1}$, we conclude that $(\bar{z}_j : j \in [n] \setminus \mathcal{S})$ must have at most m_1 nonzero entries. Moreover, it follows from the fact that $\mathcal{S}, \mathcal{S}_1, \ldots, \mathcal{S}_{m_2}$ are disjoint sets and the fact that $z_{j_i} = -\sum_{j \in \mathcal{S}_i} z_j$ for each $i \in [m_2]$ that at most m_1 of the entries $z_{j_1}, \ldots, z_{j_{m_2}}$ are nonzero. This concludes our proof that $\|\bar{\mathbf{z}}\|_0 \leq 2m_1$. $\square$

Appendix B: Proofs of Theorem 1 and Theorem 2

We observe that Theorem 1 is a special case of Theorem 2 from §6 in which $\delta_e = 0$ for all factories $e \in [E]$. Therefore, it suffices to present the proof of Theorem 2.

Proof of Theorem 2. Our proof is split into three steps, corresponding to Lemmas 1, 2, and 3.

Step 1: We begin in the first step of our proof of Theorem 2 by applying Lemma 1 for the epigraph formulation of (LDR) for cost functions given by (2a)-(2d). To do this, we first notice that since the minimal lead time across the factories is $\delta \triangleq \min_{e \in [E]} \delta_e$ and since $c_{te} > 0$, it must be the case that $x_{te} = 0$ for every $t \geq T - \delta + 1$ and $e \in [E]$ at every optimal solution. Next, we rewrite the cost function from lines (2a)-(2d) to match the format used by Lemmas 1 and 2. Indeed, we recall that Lemmas 1 and 2 involve cost functions in which the number of stages with decisions is equal to the number of stages with uncertain variables. Thus, by introducing dummy decision variables $\mathbf{x}_{T+1} = 0 \in \mathbb{R}^E$, we observe that the cost function on lines (2a)-(2d) can be equivalently written as

$$C(\mathbf{x}_1, \ldots, \mathbf{x}_{T+1}, \zeta_1, \ldots, \zeta_{T+1}) = \sum_{e=1}^{E} \sum_{t=1}^{T+1-\delta} c_{te} x_{te} \tag{2a}$$

$$\text{subject to} \quad \sum_{t=1}^{T+1-\delta} x_{te} \leq Q_e \qquad \forall e \in [E] \tag{2b}$$

$$0 \leq x_{te} \leq p_{te} \qquad \forall e \in [E], t \in [T+1-\delta] \tag{2c}$$

$$V_{\min} \leq v_1 + \sum_{e=1}^{E} \sum_{\ell=1}^{t-\delta_e} x_{\ell e} - \sum_{s=2}^{t+1} \zeta_s \leq V_{\max} \qquad \forall t \in [T], \tag{2d}$$

where we define $p_{T+1,e} \equiv 0$ and $c_{T+1,e} \equiv 0$ for each factory $e \in [E]$. After adding the dummy decision variables $\mathbf{x}_{T+1}$, we observe that the above cost function matches the format used by Lemmas 1 and 2.

Before proceeding onward, let us make two brief remarks about the convention used in the above cost function. First, we remark that the constraint (2c) implies that the dummy decision variables $\mathbf{x}_{T+1}$ will always be equal to zero. Second, we remark that the decisions x_{te} for each $t \geq T - \delta_e + 1$ are unnecessary in the above formulation, in the sense that these decisions could always be set to zero without loss of generality. To see why the decisions are unnecessary, we notice that the decision x_{te} for each $t \geq T - \delta_e + 1$ does not appear in constraint (2d); therefore, since $c_{te} \geq 0$, we observe that x_{te} can at optimality take its minimum value that is allowed by constraints (2b) and (2c). Although they are not necessary, the decisions x_{te} for each $t \geq T - \delta_e + 1$ are included in the above cost function to simplify the notation in the rest of the proof.

We next derive the optimization problem (LDR) corresponding to the above cost function (2a)-(2d). For the sake of clarity, we will show each of the intermediary algebra steps in this derivation. Indeed, we first observe from algebra that line (2a) can be written with linear decision rules as

$$\sum_{e=1}^{E} \sum_{t=1}^{T+1-\delta} c_{te} x_{te} = \sum_{e=1}^{E} \sum_{t=1}^{T+1-\delta} c_{te} \left(\sum_{s=1}^{t} y_{t,s,e} \zeta_s \right) = \sum_{s=1}^{T+1-\delta} \left(\sum_{t=s}^{T+1-\delta} \sum_{e=1}^{E} c_{te} y_{t,s,e} \right) \zeta_s,$$

where the first equality comes from using linear decision rules and the second equality follows from algebra. We observe that the left-hand side of line (2b) for each $e \in [E]$ can be written with linear decision rules as

$$\sum_{t=1}^{T+1-\delta} x_{te} = \sum_{t=1}^{T+1-\delta} \left(\sum_{s=1}^{t} y_{t,s,e} \zeta_s \right) = \sum_{s=1}^{T+1-\delta} \left(\sum_{t=s}^{T+1-\delta} y_{t,s,e} \right) \zeta_s,$$

where the first equality comes from using linear decision rules and the second equality follows from algebra. We observe that the decision in line (2c) for each $e \in [E]$ and $t \in [T + 1 - \delta]$ can be written with linear decision rules as

$$x_{te} = \sum_{s=1}^{t} y_{t,s,e} \zeta_s + \sum_{s=t+1}^{T+1} 0 \zeta_s,$$

where the equality comes from using linear decision rules. Finally, we observe that the inventory in line (2d) for each $t \in [T]$ can be written with linear decision rules as

$$\begin{aligned}
&v_1 + \sum_{e=1}^{E} \sum_{\ell=1}^{t-\delta_e} x_{\ell e} - \sum_{s=2}^{t+1} \zeta_s \\
&= v_1 + \sum_{e=1}^{E} \sum_{\ell=1}^{t-\delta_e} \left(\sum_{s=1}^{\ell} y_{\ell,s,e} \zeta_s \right) - \sum_{s=2}^{t+1} \zeta_s \\
&= v_1 + \sum_{s=1}^{t} \left(\sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \leq t-\ell} y_{\ell,s,e} \right) \zeta_s - \sum_{s=2}^{t+1} \zeta_s \\
&= v_1 + \left(\sum_{\ell=1}^{t} \sum_{e \in [E]: \delta_e \leq t-\ell} y_{\ell,1,e} \right) \zeta_1 + \sum_{s=2}^{t} \left(-1 + \sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \leq t-\ell} y_{\ell,s,e} \right) \zeta_s - \zeta_{t+1} + \sum_{s=t+2}^{T+1} 0 \zeta_s,
\end{aligned}$$

where the first equality comes from using linear decision rules and the second and third equalities follow from algebra. Combining the above steps, we conclude that the epigraph form of the optimization problem (LDR) with cost function (2a)-(2d) is equivalent to

$$\begin{aligned}
\underset{\substack{c_0 \in \mathbb{R} \\ \mathbf{y}_{t,1},\ldots,\mathbf{y}_{t,t} \in \mathbb{R}^E :\, \forall t \in [T+1]}}{\text{minimize}} \quad & c_0 \\
\text{subject to} \quad & \sum_{s=1}^{T+1-\delta} \left(\sum_{t=s}^{T+1-\delta} \sum_{e=1}^{E} c_{te} y_{t,s,e} \right) \zeta_s \le c_0 \\
& \sum_{s=1}^{T+1-\delta} \left(\sum_{t=s}^{T+1-\delta} y_{t,s,e} \right) \zeta_s \le Q_e \qquad \forall e \in [E] \\
& 0 \le \sum_{s=1}^{t} y_{t,s,e} \zeta_s + \sum_{s=t+1}^{T+1} 0 \zeta_s \le p_{te} \quad \forall e \in [E],\, t \in [T+1-\delta] \\
& V_{\min} \le v_1 + \left(\sum_{\ell=1}^{t} \sum_{e \in [E]: \delta_e \le t-\ell} y_{\ell,1,e} \right) \zeta_1 \\
& \qquad + \sum_{s=2}^{t} \left(-1 + \sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \le t-\ell} y_{\ell,s,e} \right) \zeta_s - \zeta_{t+1} + \sum_{s=t+2}^{T+1} 0 \zeta_s \le V_{\max} \quad \forall t \in [T] \\
& \forall \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_{T+1} \in \mathcal{U}_{T+1}.
\end{aligned} \tag{LDR-2}$$

With the above notation, we are now ready to invoke Lemma 1. Indeed, we recall from the statement of Theorem 2 that Assumption 1 holds for cost function (2a)-(2d). Furthermore, the constraint in (2c) guarantees Assumption 2 holds. Therefore, it follows from Lemma 1 that the set of feasible solutions to (LDR-2) is a nonempty polyhedron with at least one extreme point.

Step 2: In the second step of our proof of Theorem 2, we use Lemma 2 to characterize the structure of extreme points for the feasible set of (LDR-2). Indeed, let $(\bar{\mathbf{y}}, \bar{c}_0)$ denote an extreme point of the set of feasible solutions of (LDR-2). Since Assumption 1 holds, it follows from Lemma 2 that there exists

- index sets $\mathcal{I}^{2b,\bar{\mathbf{y}}} \subseteq [E]$, $\underline{\mathcal{I}}^{2c,\bar{\mathbf{y}}}, \bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}} \subseteq [T+1-\delta] \times [E]$, and $\underline{\mathcal{I}}^{2d,\bar{\mathbf{y}}}, \bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}} \subseteq [T]$;
- index sets $\mathcal{T}^{2a,\bar{\mathbf{y}}} \subseteq [T+1-\delta]$, $\mathcal{T}_e^{2b,\bar{\mathbf{y}}} \subseteq [T+1-\delta]$ for each $e \in \mathcal{I}^{2b,\bar{\mathbf{y}}}$, $\underline{\mathcal{T}}_{t,e}^{2c,\bar{\mathbf{y}}} \subseteq [T+1-\delta]$ for each $(t,e) \in \underline{\mathcal{I}}^{2c,\bar{\mathbf{y}}}$, $\bar{\mathcal{T}}_{t,e}^{2c,\bar{\mathbf{y}}} \subseteq [T+1]$ for each $(t,e) \in \bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}}$, $\underline{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \subseteq [T+1]$ for each $t \in \underline{\mathcal{I}}^{2d,\bar{\mathbf{y}}}$, and $\bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \subseteq [T+1]$ for each $t \in \bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}}$;
- hyperplanes $(\boldsymbol{\alpha}^{2a,\bar{\mathbf{y}}}, \beta^{2a,\bar{\mathbf{y}}})$, $(\boldsymbol{\alpha}_e^{2b,\bar{\mathbf{y}}}, \beta_e^{2b,\bar{\mathbf{y}}})$ for each $e \in \mathcal{I}^{2b,\bar{\mathbf{y}}}$, $(\underline{\boldsymbol{\alpha}}_{t,e}^{2c,\bar{\mathbf{y}}}, \underline{\beta}_{t,e}^{2c,\bar{\mathbf{y}}})$ for each $(t,e) \in \underline{\mathcal{I}}^{2c,\bar{\mathbf{y}}}$, $(\bar{\boldsymbol{\alpha}}_{t,e}^{2c,\bar{\mathbf{y}}}, \bar{\beta}_{t,e}^{2c,\bar{\mathbf{y}}})$ for each $(t,e) \in \bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}}$, $(\underline{\boldsymbol{\alpha}}_t^{2d,\bar{\mathbf{y}}}, \underline{\beta}_t^{2d,\bar{\mathbf{y}}})$ for each $t \in \underline{\mathcal{I}}^{2d,\bar{\mathbf{y}}}$, and $(\bar{\boldsymbol{\alpha}}_t^{2d,\bar{\mathbf{y}}}, \bar{\beta}_t^{2d,\bar{\mathbf{y}}})$ for each $t \in \bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}}$

such that $\bar{\mathbf{y}}$ is the unique solution to the following system of equations:

$$\boldsymbol{\alpha}^{2a,\bar{\mathbf{y}}} \cdot \mathbf{y} = \beta^{2a,\bar{\mathbf{y}}} \tag{HARD-2a}$$

$$\boldsymbol{\alpha}_e^{2b,\bar{\mathbf{y}}} \cdot \mathbf{y} = \beta_e^{2b,\bar{\mathbf{y}}} \qquad \forall e \in \mathcal{I}^{2b,\bar{\mathbf{y}}} \tag{HARD-2b}$$

$$\underline{\boldsymbol{\alpha}}_{t,e}^{2c,\bar{\mathbf{y}}} \cdot \mathbf{y} = \underline{\beta}_{t,e}^{2c,\bar{\mathbf{y}}} \qquad \forall (t,e) \in \underline{\mathcal{I}}^{2c,\bar{\mathbf{y}}} \tag{HARD-2c-LB}$$

$$\bar{\boldsymbol{\alpha}}_{t,e}^{2c,\bar{\mathbf{y}}} \cdot \mathbf{y} = \bar{\beta}_{t,e}^{2c,\bar{\mathbf{y}}} \qquad \forall (t,e) \in \bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}} \tag{HARD-2c-UB}$$

$$\underline{\boldsymbol{\alpha}}_t^{2d,\bar{\mathbf{y}}} \cdot \mathbf{y} = \underline{\beta}_t^{2d,\bar{\mathbf{y}}} \qquad \forall t \in \underline{\mathcal{I}}^{2d,\bar{\mathbf{y}}} \tag{HARD-2d-LB}$$

$$\bar{\boldsymbol{\alpha}}_t^{2d,\bar{\mathbf{y}}} \cdot \mathbf{y} = \bar{\beta}_t^{2d,\bar{\mathbf{y}}} \qquad \forall t \in \bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}} \tag{HARD-2d-UB}$$

$$\sum_{t=s}^{T+1} \sum_{e=1}^{E} c_{te} y_{t,s,e} = 0 \qquad \forall s \in \mathcal{T}^{2a,\bar{\mathbf{y}}} \tag{EASY-2a}$$

$$\sum_{t=s}^{T+1} y_{t,s,e} = 0 \qquad \forall e \in [E],\ s \in \mathcal{T}_e^{2b,\bar{\mathbf{y}}} \tag{EASY-2b}$$

$$y_{t,s,e} = 0 \qquad \forall e \in [E],\ t \in [T+1],\ s \in \mathcal{T}_{t,e}^{2c,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_{t,e}^{2c,\bar{\mathbf{y}}} \text{ if } s \le t \tag{EASY-2c-i}$$

$$0 = 0 \qquad \forall e \in [E],\ t \in [T+1],\ s \in \mathcal{T}_{t,e}^{2c,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_{t,e}^{2c,\bar{\mathbf{y}}} \text{ if } s \ge t+1 \tag{EASY-2c-ii}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \le t-\ell} y_{\ell,s,e} = 0 \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \text{ if } s = 1 \tag{EASY-2d-i}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \le t-\ell} y_{\ell,s,e} = 1 \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \text{ if } s \in \{2,\ldots,t\} \tag{EASY-2d-ii}$$

$$0 = 1 \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \text{ if } s = t+1 \tag{EASY-2d-iii}$$

$$0 = 0 \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \text{ if } s \ge t+2. \tag{EASY-2d-iv}$$

Let us make two observations about the above system of equations. First, we recall from its construction that there is exactly one solution to the above system, namely, $\bar{\mathbf{y}}$. Therefore, we readily observe that there must be no constraints of the form (EASY-2d-iii), i.e., the inequality $s \le t$ must hold for all $t \in [T]$ and all $s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}}$. Second, we observe that we can remove the constraints of type (EASY-2c-ii) and (EASY-2d-iv) without loss of generality. Hence, we will drop the constraints (EASY-2d-iii), (EASY-2c-ii), and (EASY-2d-iv) in our subsequent analysis.

Step 3: In the third step of our proof of Theorem 2, we use Lemma 3 to show that every extreme point $(\bar{\mathbf{y}}, \bar{c}_0)$ of the feasible set of (LDR-2) is sparse.

To motivate our usage of Lemma 3, let us make several observations about the system of equations corresponding to the extreme point $(\bar{\mathbf{y}}, \bar{c}_0)$. First, we observe that there are $\mathcal{O}(ET)$ equations in each of the lines (HARD-2a), (HARD-2b), (HARD-2c-LB), (HARD-2c-UB), (HARD-2d-LB), (HARD-2d-UB), (EASY-2a), (EASY-2b), and (EASY-2d-i). Second, we recall from our discussion at the end of Step 2 that the equations in lines (EASY-2c-ii), (EASY-2d-iii) and (EASY-2d-iv) can be dropped without loss of generality. Third, we observe that the equations in line (EASY-2c-i) will ultimately impose sparsity into the solution the system of equations. Finally, we observe that there are up to $\mathcal{O}(ET^2)$ equations on line (EASY-2d-ii).

With the goal of applying Lemma 3 to the above system of equations, we perform algebraic manipulations on the equations in (EASY-2d-ii). Indeed, we first define the following index sets:

$$\mathscr{S}^{\bar{\mathbf{y}}} \triangleq \bigcup_{t=1}^{T} \left(\left(\mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \right) \cap \{2,\ldots,t\} \right),$$

$$\mathscr{T}_s^{\bar{\mathbf{y}}} \triangleq \left\{ t \in [T] : s \in \left(\mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \right) \cap \{2,\ldots,t\} \right\} \quad \forall s \in \mathscr{S}^{\bar{\mathbf{y}}}.$$

With the above notation, we readily observe that (EASY-2d-ii) can be written equivalently as

$$\sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \le t-\ell} y_{\ell,s,e} = 1 \quad \forall s \in \mathscr{S}^{\bar{\mathbf{y}}},\ t \in \mathscr{T}_s^{\bar{\mathbf{y}}}. \tag{EASY-2d-ii}$$

For each $s \in \mathscr{S}^{\bar{\mathbf{y}}}$, let the elements of $\mathscr{T}_s^{\bar{\mathbf{y}}}$ be indexed in ascending order by $t_{s,1}^{\bar{\mathbf{y}}} < \ldots < t_{s,|\mathscr{T}_s^{\bar{\mathbf{y}}}|}^{\bar{\mathbf{y}}}$. With this notation, we observe that (EASY-2d-ii) can be written equivalently as the following system of equations:

$$\sum_{\ell=s}^{t_{s,1}^{\bar{\mathbf{y}}}} \sum_{e \in [E]: \delta_e \le t_{s,1}^{\bar{\mathbf{y}}}-\ell} y_{\ell,s,e} = 1 \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}}} \tag{EASY-2d-ii'}$$

$$\sum_{\ell=s}^{t^{\bar{\mathbf{y}}}_{s,k+1}} \sum_{e\in[E]:\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} - \sum_{\ell=s}^{t^{\bar{\mathbf{y}}}_{s,k}} \sum_{e\in[E]:\delta_e\le t^{\bar{\mathbf{y}}}_{s,k}-\ell} y_{\ell,s,e} = 1-1=0 \qquad \forall s\in\mathscr{S}^{\bar{\mathbf{y}}},\ k\in\{1,\ldots,|\mathscr{T}^{\bar{\mathbf{y}}}_s|-1\}. \tag{EASY-2d-ii”}$$

In particular, it follows from algebra that (EASY-2d-ii”) is equivalent to

$$\sum_{\ell=t^{\bar{\mathbf{y}}}_{s,k}+1}^{t^{\bar{\mathbf{y}}}_{s,k+1}} \sum_{e\in[E]:\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} + \sum_{\ell=s}^{t^{\bar{\mathbf{y}}}_{s,k}} \sum_{e\in[E]:t^{\bar{\mathbf{y}}}_{s,k}-\ell+1\le\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} = 0 \quad \forall s\in\mathscr{S}^{\bar{\mathbf{y}}},\ k\in\{1,\ldots,|\mathscr{T}^{\bar{\mathbf{y}}}_s|-1\}. \tag{EASY-2d-ii”}$$

To simplify our notation, we now compactly represent the constraints from lines (HARD-2a), (HARD-2b), (HARD-2c-LB), (HARD-2c-UB) (HARD-2d-LB), (HARD-2d-UB), (EASY-2a), (EASY-2b), (EASY-2d-i) and (EASY-2d-ii’) using the index set $\mathscr{I}^{\bar{\mathbf{y}}}$ and hyperplanes $(\boldsymbol{\alpha}^{\bar{\mathbf{y}}}_i, \beta^{\bar{\mathbf{y}}}_i)$ for each $i\in\mathscr{I}^{\bar{\mathbf{y}}}$, where

$$\begin{aligned}
|\mathscr{I}^{\bar{\mathbf{y}}}| &= \underbrace{1}_{\text{(HARD-2a)}} + \underbrace{\left|\mathcal{I}^{2b,\bar{\mathbf{y}}}\right|}_{\text{(HARD-2b)}} + \underbrace{\left|\mathcal{I}^{2c,\bar{\mathbf{y}}}\right|}_{\text{(HARD-2c-LB)}} + \underbrace{\left|\bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}}\right|}_{\text{(HARD-2c-UB)}} + \underbrace{\left|\mathcal{I}^{2d,\bar{\mathbf{y}}}\right|}_{\text{(HARD-2d-LB)}} + \underbrace{\left|\bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}}\right|}_{\text{(HARD-2d-UB)}} \\
&\quad + \underbrace{\left|\mathcal{T}^{2a,\bar{\mathbf{y}}}\right|}_{\text{(EASY-2a)}} + \underbrace{\sum_{e=1}^{E}\left|\mathcal{T}^{2b,\bar{\mathbf{y}}}_e\right|}_{\text{(EASY-2b)}} + \underbrace{\left|\left\{t\in[T]:\ 1\in\mathcal{I}^{2d,\bar{\mathbf{y}}}_t\cup\bar{\mathcal{T}}^{2d,\bar{\mathbf{y}}}_t\right\}\right|}_{\text{(EASY-2d-i)}} + \underbrace{\left|\mathscr{S}^{\bar{\mathbf{y}}}\right|}_{\text{(EASY-2d-ii')}} \\
&\le \underbrace{1}_{\text{(HARD-2a)}} + \underbrace{E}_{\text{(HARD-2b)}} + \underbrace{(T+1-\delta)E}_{\text{(HARD-2c-LB)}} + \underbrace{(T+1-\delta)E}_{\text{(HARD-2c-UB)}} + \underbrace{T}_{\text{(HARD-2d-LB)}} + \underbrace{T}_{\text{(HARD-2d-UB)}} \\
&\quad + \underbrace{T+1}_{\text{(EASY-2a)}} + \underbrace{(T+1-\delta)E}_{\text{(EASY-2b)}} + \underbrace{T}_{\text{(EASY-2d-i)}} + \underbrace{T-1}_{\text{(EASY-2d-ii')}} \\
&= 1+4E+5T+3ET-3E\delta.
\end{aligned}$$

With the above notation, we have shown that $\bar{\mathbf{y}}$ is the unique solution to the following system of equations:

$$\boldsymbol{\alpha}^{\bar{\mathbf{y}}}_i\cdot\mathbf{y} = \beta^{\bar{\mathbf{y}}}_i \quad \forall i\in\mathscr{I}^{\bar{\mathbf{y}}} \tag{HARD-combined}$$

$$\sum_{\ell=t^{\bar{\mathbf{y}}}_{s,k}+1}^{t^{\bar{\mathbf{y}}}_{s,k+1}} \sum_{e\in[E]:\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} + \sum_{\ell=s}^{t^{\bar{\mathbf{y}}}_{s,k}} \sum_{e\in[E]:t^{\bar{\mathbf{y}}}_{s,k}-\ell+1\le\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} = 0 \quad \forall s\in\mathscr{S}^{\bar{\mathbf{y}}},\ k\in\{1,\ldots,|\mathscr{T}^{\bar{\mathbf{y}}}_s|-1\}. \tag{EASY-2d-ii”}$$

$$y_{t,s,e} = 0 \quad \forall e\in[E],\ t\in[T+1],\ s\in\mathcal{T}^{2c,\bar{\mathbf{y}}}_{t,e}\cup\bar{\mathcal{T}}^{2c,\bar{\mathbf{y}}}_{t,e} \text{ if } s\le t \tag{EASY-2c-i}$$

We next apply Lemma 3 to the above system of equations by noticing (HARD-combined), (EASY-2d-ii”), and (EASY-2c-i) follow the same structure of (S-1), (S-2) and (S-3), respectively. Furthermore, we notice that there is no overlapping index of y in (EASY-2d-ii”), thus each column of the corresponding $\mathbf{P}_2$ has at most one nonzero entry. Since $\bar{\mathbf{y}}$ is the unique solution to the above system of equations, it follows from Lemma 3 that the number of nonzero entries in $\bar{\mathbf{y}}$ satisfies

$$\|\bar{\mathbf{y}}\|_0 \le 2\,|\mathcal{I}| \le 2\,(1+4E+5T+3ET-3E\delta),$$

Since the above inequality holds for every extreme point $(\bar{\mathbf{y}},\bar{c}_0)$, since we have proven that (LDR-2) has at least one extreme point, and since we observe from Assumption 1 that (LDR-2) has an optimal solution which is an extreme point, our proof of Theorem 2 is complete. □

Appendix C: Proofs of Theorem 3

Proof of Theorem 3. Our proof of Theorem 3 follows a similar organization to that of Theorem 2. Specifically, the proof of Theorem 2 is split into three steps which correspond to Lemmas 1, 2, and 3. In contrast to the proof of Theorem 2, the proof of Theorem 3 requires the introduction of auxiliary decision rules to account for nonlinear cost functions.

Step 1: We begin in the first step of our proof of Theorem 3 by applying Lemma 1 to an epigraph formulation of (LDR) for cost function (3a)-(3b). Indeed, we observe that (RO) with cost function (3a)-(3b) can be reformulated as

$$\begin{aligned}
&\underset{\mathbf{x},\mathbf{z}}{\text{minimize}} && \max_{\zeta_1 \in \mathcal{U}_1,\ldots,\zeta_{T+1}\in\mathcal{U}_{T+1}} \left\{ \sum_{t=1}^{T+1} c_t x_t(\zeta_1,\ldots,\zeta_t) + \sum_{t=1}^{T+1} z_t(\zeta_1,\ldots,\zeta_t)\right\} \\
&\text{subject to} && z_{t+1}(\zeta_1,\ldots,\zeta_{t+1}) \ge h_t\left(v_1 + \sum_{s=1}^{t} x_s(\zeta_1,\ldots,\zeta_s) - \sum_{s=2}^{t+1}\zeta_s\right) && \forall t\in[T] \\
&&& z_{t+1}(\zeta_1,\ldots,\zeta_{t+1}) \ge -b_t\left(v_1 + \sum_{s=1}^{t} x_s(\zeta_1,\ldots,\zeta_s) - \sum_{s=2}^{t+1}\zeta_s\right) && \forall t\in[T] \\
&&& 0 \le x_t(\zeta_1,\ldots,\zeta_t) \le p_t && \forall t\in[T+1] \\
&&& z_1(\zeta_1)\ge 0 \\
&&& \quad \forall \zeta_1\in\mathcal{U}_1,\ldots,\zeta_{T+1}\in\mathcal{U}_{T+1},
\end{aligned}$$

where the auxiliary decision rule $z_{t+1}(\zeta_1,\ldots,\zeta_{t+1})$ captures the holding and backorder costs in each period $t\in[T]$, $p_{T+1}=0$, $c_{T+1}=1$, and $x_{T+1}(\zeta_1,\ldots,\zeta_{T+1})$ and $z_1(\zeta_1)$ are dummy decision rules that can always be identically equal to zero at optimality. We note that the dummy decision rules $x_{T+1}(\zeta_1,\ldots,\zeta_{T+1})$ and $z_1(\zeta_1)$ have been introduced into the above optimization problem to match the setting of Lemmas 1 and 2, in which the number of decisions is constant in each stage (in this case, the number of decisions in each stage is $n=2$). We also readily observe from inspection that any feasible solution of the above optimization problem will satisfy $x_t(\zeta_1,\ldots,\zeta_t), z_t(\zeta_1,\ldots,\zeta_t)\ge 0$ for all stages $t\in[T+1]$ and all realizations $\zeta_1\in\mathcal{U}_1,\ldots,\zeta_{T+1}\in\mathcal{U}_{T+1}$.

It follows from [9, Theorem 3.1] and Assumption 1 that the auxiliary decision rules $z_t(\zeta_1,\ldots,\zeta_t)$ and the production decision rules $x_t(\zeta_1,\ldots,\zeta_t)$ for all $t\in[T+1]$ in the above optimization problem can be replaced with linear decision rules without any loss of optimality. Hence, we conclude that the optimization problems (RO) and (LDR) with cost function (3a)-(3b) are equivalent to one another and can be written as

$$\begin{aligned}
&\underset{\substack{y_{t,1},\ldots,y_{t,t}\in\mathbb{R}\ \forall t\in[T+1]\\ w_{t,1},\ldots,w_{t,t}\in\mathbb{R}\ \forall t\in[T+1]}}{\text{minimize}} && \max_{\zeta_1 \in \mathcal{U}_1,\ldots,\zeta_{T+1}\in\mathcal{U}_{T+1}} \left\{ \sum_{t=1}^{T+1} c_t\left(\sum_{s=1}^{t} y_{t,s}\zeta_s\right) + \sum_{t=1}^{T+1}\left(\sum_{s=1}^{t} w_{t,s}\zeta_s\right)\right\} \\
&\text{subject to} && \sum_{s=1}^{t+1} w_{t+1,s}\zeta_s \ge h_t\left(v_1 + \sum_{\ell=1}^{t}\left(\sum_{s=1}^{\ell} y_{\ell,s}\zeta_s\right) - \sum_{s=2}^{t+1}\zeta_s\right) && \forall t\in[T] \\
&&& \sum_{s=1}^{t+1} w_{t+1,s}\zeta_s \ge -b_t\left(v_1 + \sum_{\ell=1}^{t}\left(\sum_{s=1}^{\ell} y_{\ell,s}\zeta_s\right) - \sum_{s=2}^{t+1}\zeta_s\right) && \forall t\in[T] \\
&&& 0 \le \sum_{s=1}^{t} y_{t,s}\zeta_s \le p_t && \forall t\in[T+1] \\
&&& w_{1,1}\zeta_1 \ge 0 \\
&&& \quad \forall \zeta_1\in\mathcal{U}_1,\ldots,\zeta_{T+1}\in\mathcal{U}_{T+1}.
\end{aligned}$$

By rearranging terms in the above optimization problem, by adding an epigraph decision variable $c_0 \in \mathbb{R}$, and by noticing that $\sum_{\ell=t+1}^{t} h_t y_{\ell,s} = -\sum_{\ell=t+1}^{t} b_t y_{\ell,s} = 0$ for each period $t \in [T]$, we observe that the above optimization problem can be written equivalently as

$$\begin{aligned}
\underset{\substack{c_0 \in \mathbb{R} \\ y_{t,1},\ldots,y_{t,t} \in \mathbb{R}\ \forall t \in [T+1] \\ w_{t,1},\ldots,w_{t,t} \in \mathbb{R}\ \forall t \in [T+1]}}{\text{minimize}} \quad & c_0 \\
\text{subject to} \quad & \sum_{s=1}^{T+1} \left(\sum_{t=s}^{T+1} (c_t y_{t,s} + w_{t,s}) \right) \zeta_s \le c_0 && && \text{(3a)} \\
& \sum_{s=1}^{t+1} \left(\sum_{\ell=s}^{t} h_t y_{\ell,s} - h_t - w_{t+1,s} \right) \zeta_s + \sum_{s=t+2}^{T+1} 0\zeta_s \le -h_t v_1 && \forall t \in [T] && \text{(3b)} \\
& \sum_{s=1}^{t+1} \left(\sum_{\ell=s}^{t} -b_t y_{\ell,s} + b_t - w_{t+1,s} \right) \zeta_s + \sum_{s=t+2}^{T+1} 0\zeta_s \le b_t v_1 && \forall t \in [T] && \text{(3c)} \\
& 0 \le \sum_{s=1}^{t} y_{t,s}\zeta_s + \sum_{s=t+1}^{T+1} 0\zeta_s \le p_t && \forall t \in [T+1] && \text{(3d)} \\
& \sum_{s=1}^{t} w_{t,s}\zeta_s + \sum_{s=t+1}^{T+1} 0\zeta_s \ge 0 && \forall t \in [1] && \text{(3e)} \\
& \quad \forall \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_{T+1} \in \mathcal{U}_{T+1}.
\end{aligned}$$

With the above notation, we are now ready to invoke Lemma 1. Indeed, we recall from the statement of Theorem 3 that Assumption 1 holds for cost function (3a)-(3b). Therefore, it follows from the above reasoning, the constraints in (3b)-(3e), and Lemma 1 that the set of feasible solutions to the above optimization problem is a nonempty polyhedron with at least one extreme point.

Step 2: In the second step of our proof of Theorem 3, we use Lemma 2 to characterize the structure of extreme points for the feasible set of the above optimization problem. Indeed, let $(\bar{\mathbf{y}}, \bar{\mathbf{w}}, \bar{c}_0)$ denote an extreme point of the set of feasible solutions of the above optimization problem. Then it follows readily from Lemma 2 that there exists

- index sets $\mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T]$, $\underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$, and $\mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [1]$;
- index sets $\mathcal{T}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$, $\mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $\underline{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $\bar{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, and $\mathcal{T}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}$;
- hyperplanes $(\boldsymbol{\alpha}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}})$, $(\boldsymbol{\alpha}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $(\boldsymbol{\alpha}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $(\underline{\boldsymbol{\alpha}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \underline{\boldsymbol{\gamma}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \underline{\beta}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $(\bar{\boldsymbol{\alpha}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \bar{\boldsymbol{\gamma}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \bar{\beta}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, and $(\boldsymbol{\alpha}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}$

such that $(\bar{\mathbf{y}}, \bar{\mathbf{w}})$ is the unique solution to the following system of equalities:

$$\boldsymbol{\alpha}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \boldsymbol{\gamma}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \beta^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \text{(HARD-3a)}$$

$$\boldsymbol{\alpha}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \boldsymbol{\gamma}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \beta_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \text{(HARD-3b)}$$

$$\boldsymbol{\alpha}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \boldsymbol{\gamma}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \beta_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \text{(HARD-3c)}$$

$$\bar{\boldsymbol{\alpha}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \bar{\boldsymbol{\gamma}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \bar{\beta}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3d-UB}$$

$$\underline{\boldsymbol{\alpha}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \underline{\boldsymbol{\gamma}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \underline{\beta}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3d-LB}$$

$$\boldsymbol{\alpha}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \boldsymbol{\gamma}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \beta_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3e}$$

$$\sum_{t=s}^{T+1} (c_t y_{t,s} + w_{t,s}) = 0 \qquad \forall s \in \mathcal{T}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{EASY-3a}$$

$$\sum_{\ell=s}^{t} h_t y_{\ell,s} - h_t - w_{t+1,s} = 0 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t+1 \tag{EASY-3b-i}$$

$$0 = 0 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \ge t+2 \tag{EASY-3b-ii}$$

$$-\sum_{\ell=s}^{t} b_t y_{\ell,s} + b_t - w_{t+1,s} = 0 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t+1 \tag{EASY-3c-i}$$

$$0 = 0 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \ge t+2 \tag{EASY-3c-ii}$$

$$y_{t,s} = 0 \qquad \forall t \in [T+1],\ s \in \bar{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \underline{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t \tag{EASY-3d-i}$$

$$0 = 0 \qquad \forall t \in [T+1],\ s \in \bar{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \underline{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \ge t+1 \tag{EASY-3d-ii}$$

$$w_{t,s} = 0 \qquad \forall t \in [1], s \in \mathcal{T}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t \tag{EASY-3e-i}$$

$$0 = 0 \qquad \forall t \in [1], s \in \mathcal{T}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \ge t+1. \tag{EASY-3e-ii}$$

Step 3: In the third step of our proof of Theorem 3, we apply Lemma 3 to every extreme point $(\bar{\mathbf{y}}, \bar{\mathbf{w}}, \bar{c}_0)$. With the goal of Lemma 3 in mind, we first observe from (EASY-3b-i) and (EASY-3c-i) that any solution to the above system of equations must satisfy the following equalities:

$$w_{t+1,s} = \sum_{\ell=s}^{t} h_t y_{\ell,s} - h_t \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t+1$$

$$w_{t+1,s} = -\sum_{\ell=s}^{t} b_t y_{\ell,s} + b_t \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t+1.$$

We can thus eliminate the variables $w_{t+1,s}$ for each $t \in [T]$ and $s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}$ with $s \le t+1$ in the above system of equations by the substitution

$$w_{t+1,s} \leftarrow \begin{cases} \sum_{\ell=s}^{t} h_t y_{\ell,s} - h_t, & \text{if } t \in [T] \text{ and } s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \\ -\sum_{\ell=s}^{t} b_t y_{\ell,s} + b_t, & \text{if } t \in [T] \text{ and } s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \setminus \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}. \end{cases}$$

With this substitution, we observe that (EASY-3b-i) and (EASY-3c-i) can be replaced with

$$\sum_{\ell=s}^{t} y_{\ell,s} = 1 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}},\ s \le t+1 \tag{EASY-3bc-i}$$

We next perform algebraic manipulations on the equations in (EASY-3bc-i). Indeed, we first define the following sets:

$$\mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \triangleq \bigcup_{t=1}^{T} \left(\left(\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \right) \cap \{1, \ldots, t+1\} \right),$$

$$\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \triangleq \left\{ t \in [T] : s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \{1, \ldots, t+1\} \right\} \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}.$$

With the above notation, we observe that the constraints (EASY-3bc-i) can be rewritten as

$$\sum_{\ell=s}^{t} y_{\ell,s} = 1 \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}},\ t \in \mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}.$$

Moreover, let the elements of $\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$ be denoted by $t_{s,1}^{\bar{\mathbf{y}},\bar{\mathbf{w}}} < \ldots < t_{s,|\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}|}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$. With this notation, we readily observe using algebra that the constraints (EASY-3bc-i) can be replaced with the following equivalent constraints:

$$\sum_{\ell=s}^{t_{s,1}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}} y_{\ell,s} = 1 \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{EASY-3bc-i'}$$

$$\sum_{\ell=t_{s,k}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}+1}^{t_{s,k+1}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}} y_{\ell,s} = 0 \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}},\ k \in \{1,\ldots,|\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}|-1\}. \tag{EASY-3bc-i''}$$

To simplify our notation, we now compactly represent the constraints from lines (HARD-3a), (HARD-3b), (HARD-3c), (HARD-3d-UB), (HARD-3d-LB), (HARD-3e), (EASY-3a), and (EASY-3bc-i') using the index set $\mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$ and hyperplanes $(\boldsymbol{\alpha}_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $i \in \mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$[11], where

$$\begin{aligned}
|\mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}| &= \underbrace{1}_{\text{(HARD-3a)}} + \underbrace{|\mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3b)}} + \underbrace{|\mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3c)}} + \underbrace{|\bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3d-UB)}} + \underbrace{|\underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3d-LB)}} + \underbrace{|\mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3e)}} + \underbrace{|\mathcal{T}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(EASY-3a)}} + \underbrace{|\mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(EASY-3bc-i')}} \\
&\le \underbrace{1}_{\text{(HARD-3a)}} + \underbrace{T}_{\text{(HARD-3b)}} + \underbrace{T}_{\text{(HARD-3c)}} + \underbrace{T+1}_{\text{(HARD-3d-UB)}} + \underbrace{T+1}_{\text{(HARD-3d-LB)}} + \underbrace{1}_{\text{(HARD-3e)}} + \underbrace{T+1}_{\text{(EASY-3a)}} + \underbrace{T}_{\text{(EASY-3bc-i')}} \\
&\le 5 + 6T.
\end{aligned}$$

It follows from the above notation, from the fact that we have used substitution to eliminate the variable $w_{t+1,s}$ for each $t \in [T]$ and $s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}$ with $s \le t+1$, and from the fact that (EASY-3b-ii), (EASY-3c-ii), (EASY-3d-ii), and (EASY-3e-ii) can be eliminated without loss of generality that $\left(\bar{\mathbf{y}}, \left(\bar{w}_{t+1,s} : t \in [T],\ s \notin \left(\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}\right) \cap \{1,\ldots,t+1\}\right)\right)$ is the unique solution to the following system of equations:

$$\begin{aligned}
&\boldsymbol{\alpha}_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \sum_{t=1}^{T} \sum_{s \notin \left(\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \{1,\ldots,t+1\}\right)} \gamma_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}} w_{t+1,s} = \beta_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}} && \forall i \in \mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \\
&y_{t,s} = 0 && \forall t \in [T+1],\ s \in \bar{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \underline{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t \qquad \text{(EASY-3d-i)} \\
&w_{1,1} = 0 && \forall t \in [1], s \in \mathcal{T}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t \qquad \text{(EASY-3e-i)} \\
&\sum_{\ell=t_{s,k}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}+1}^{t_{s,k+1}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}} y_{\ell,s} = 0 && \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}},\ k \in \{1,\ldots,|\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}|-1\}.
\end{aligned} \tag{EASY-3bc''}$$

Notice that any $y_{\ell,s}$ only appear once in (EASY-3bc''). It thus follows from Lemma 3 that the number of nonzero entries in $\bar{\mathbf{y}}$ satisfies

$$\|\bar{\mathbf{y}}\|_0 \le \left\|\left(\bar{\mathbf{y}}, \left(\bar{w}_{t+1,s} : t \in [T],\ s \notin \left(\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}\right) \cap \{1,\ldots,t+1\}\right)\right)\right\|_0 \le 2\,|\mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}| \le 10 + 12T,$$

which concludes our proof of Theorem 3. □

[11] The vector $\boldsymbol{\gamma}_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$ in each hyperplane $i \in \mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$ contains an element corresponding to the variable $w_{t+1,s}$ for each $t \in [T]$ and $s \notin \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}$.

Appendix D: Proof of Theorem 4

Proof of Theorem 4. The basic idea of the proof is reducing the problem with non-separable uncertainty set to a problem with box uncertainty set, by dualizing the linking constraints, and then utilizing the same arguments as in the proof of Theorem 2.

With the same argument as the Step 1 in the proof of Theorem 2, we know that the epigraph form of the problem is equivalent to

$$\begin{aligned}
&\underset{\substack{c_0\in\mathbb{R}\\ \mathbf{y}_{t,1},\ldots,\mathbf{y}_{t,t}\in\mathbb{R}^E:\,\forall t\in[T+1]}}{\text{minimize}} && c_0 && (4)\\
&\text{subject to} && \sum_{s=1}^{T+1}\left(\sum_{t=s}^{T+1}\sum_{e=1}^{E} c_{te}y_{t,s,e}\right)\zeta_s \le c_0 && (4a)\\
& && \sum_{s=1}^{T+1}\left(\sum_{t=s}^{T+1} y_{t,s,e}\right)\zeta_s \le Q_e \qquad \forall e\in[E] && (4b)\\
& && 0\le \sum_{s=1}^{t} y_{t,s,e}\zeta_s + \sum_{s=t+1}^{T+1} 0\zeta_s \le p_{te} \quad \forall e\in[E],\ t\in[T+1] && (4c)\\
& && V_{\min} \le v_1 + \left(\sum_{\ell=1}^{t}\sum_{e\in[E]} y_{\ell,1,e}\right)\zeta_1 && \\
& && \qquad + \sum_{s=2}^{t}\left(-1+\sum_{\ell=s}^{t}\sum_{e\in[E]} y_{\ell,s,e}\right)\zeta_s - \zeta_{t+1} + \sum_{s=t+2}^{T+1} 0\zeta_s \le V_{\max} \quad \forall t\in[T] && (4d)\\
& && \forall \boldsymbol{\zeta}\equiv(\zeta_1,\ldots,\zeta_{T+1})\in\mathcal{U}.
\end{aligned}$$

Next, we denote $f_s=\frac{1}{\bar{D}_s-\underline{D}_s}$ for $s=2,\ldots,T+1$, $f_s=0$ for $s=1$, and $g=T-k+\sum_{s=2}^{T+1}\frac{\underline{D}_s}{\bar{D}_s-\underline{D}_s}$. Then, we can rewrite the budget uncertainty set $\mathcal{U}$ as $\{\zeta_1\in\mathcal{U}_1,\ldots,\zeta_T\in\mathcal{U}_T:\mathbf{f}^\intercal\boldsymbol{\zeta}\le g\}$. Now we consider the generic form of the problem with the notation of (EC.1), where the feasible region becomes

$$\mathcal{Y}\triangleq\left\{(\mathbf{y},c_0):\max_{\zeta_1\in\mathcal{U}_1,\ldots,\zeta_T\in\mathcal{U}_T,\mathbf{f}^\intercal\boldsymbol{\zeta}\le g}\left\{\sum_{t=1}^{T}\mathbf{a}_{i,t}^\intercal\left(\sum_{s=1}^{t}\mathbf{y}_{t,s}\zeta_s\right)-\sum_{t=1}^{T}b_{i,t}\zeta_t\right\}\le c_i\ \ \forall i\in\{0,\ldots,m\}\right\}.$$

Furthermore, we denote $\mu_i\ge 0$ as the dual variable for $\mathbf{f}^\intercal\boldsymbol{\zeta}\le g$ in the i-th constraint, and denote

$$\mathcal{Y}_\mu\triangleq\left\{(\mathbf{y},c_0):\max_{\zeta_1\in\mathcal{U}_1,\ldots,\zeta_T\in\mathcal{U}_T}\left\{\sum_{t=1}^{T}\mathbf{a}_{i,t}^\intercal\left(\sum_{s=1}^{t}\mathbf{y}_{t,s}\zeta_s\right)-\sum_{t=1}^{T}b_{i,t}\zeta_t+\mu_i(g-\mathbf{f}^\intercal\boldsymbol{\zeta})\right\}\le c_i\ \ \forall i\in\{0,\ldots,m\}\right\}.$$

The next lemma presents the connection between $\mathcal{Y}$ and $\mathcal{Y}_{\boldsymbol{\mu}}$:

LEMMA EC.1. $\mathcal{Y}=\cup_{\boldsymbol{\mu}\ge\mathbf{0}}\mathcal{Y}_{\boldsymbol{\mu}}$.

Proof. Consider any $\boldsymbol{\mu}\ge\mathbf{0}$, and let $(\mathbf{y},c_0)\in\mathcal{Y}_{\boldsymbol{\mu}}$. Then it holds for each $i\in\{0,\ldots,m\}$ that

$$\begin{aligned}
&\max_{\zeta_1\in\mathcal{U}_1,\ldots,\zeta_T\in\mathcal{U}_T,\mathbf{f}^\intercal\boldsymbol{\zeta}\le g}\left\{\sum_{t=1}^{T}\mathbf{a}_{i,t}^\intercal\left(\sum_{s=1}^{t}\mathbf{y}_{t,s}\zeta_s\right)-\sum_{t=1}^{T}b_{i,t}\zeta_t\right\}\\
&\le\max_{\zeta_1\in\mathcal{U}_1,\ldots,\zeta_T\in\mathcal{U}_T}\left\{\sum_{t=1}^{T}\mathbf{a}_{i,t}^\intercal\left(\sum_{s=1}^{t}\mathbf{y}_{t,s}\zeta_s\right)-\sum_{t=1}^{T}b_{i,t}\zeta_t+\mu_i(g-\mathbf{f}^\intercal\boldsymbol{\zeta})\right\}\\
&\le c_i,
\end{aligned}$$

where the first inequality uses weak duality and the second inequality is from the definition of $\mathcal{Y}_{\boldsymbol{\mu}}$. Thus, it holds for all $\mu_i \ge 0$ that $\mathcal{Y}_{\boldsymbol{\mu}} \subseteq \mathcal{Y}$, whereby

$$\cup_{\boldsymbol{\mu}\ge\mathbf{0}}\mathcal{Y}_{\boldsymbol{\mu}} \subseteq \mathcal{Y}. \tag{EC.4}$$

On the other hand, for any $(\mathbf{y}, c_0) \in \mathcal{Y}$, it holds by strong duality of linear optimization that there exists $\boldsymbol{\mu} \ge \mathbf{0}$ such that

$$\begin{aligned}&\max_{\zeta_1\in\mathcal{U}_1,\ldots,\zeta_T\in\mathcal{U}_T}\left\{\sum_{t=1}^{T}\mathbf{a}_{i,t}^{\intercal}\left(\sum_{s=1}^{t}\mathbf{y}_{t,s}\zeta_s\right)-\sum_{t=1}^{T}b_{i,t}\zeta_t+\mu_i(g-\mathbf{f}^{\intercal}\xi)\right\}\\ =&\max_{\zeta_1\in\mathcal{U}_1,\ldots,\zeta_T\in\mathcal{U}_T,\mathbf{f}^{\intercal}\zeta\le g}\left\{\sum_{t=1}^{T}\mathbf{a}_{i,t}^{\intercal}\left(\sum_{s=1}^{t}\mathbf{y}_{t,s}\zeta_s\right)-\sum_{t=1}^{T}b_{i,t}\zeta_t\right\}\le c_i, \forall i\in\{0,\ldots,m\}\ ,\end{aligned}$$

thus it holds that $y \in \mathcal{Y}_{\boldsymbol{\mu}}$, whereby

$$\mathcal{Y} \subseteq \cup_{\boldsymbol{\mu}\ge\mathbf{0}}\mathcal{Y}_{\boldsymbol{\mu}}\ . \tag{EC.5}$$

We finish the proof by combining (EC.4) and (EC.5). $\square$

The next lemma shows that an extreme point of $\mathcal{Y}$ is also an extreme point of $\mathcal{Y}_{\boldsymbol{\mu}}$.

Lemma EC.2. *For any extreme point* $(\mathbf{y}, c_0)$ *of* $\mathcal{Y}$*, there exists* $\boldsymbol{\mu} \ge \mathbf{0}$*, such that* $(\mathbf{y}, c_0)$ *is an extreme point of* $\mathcal{Y}_{\boldsymbol{\mu}}$*.*

Proof. It follows from Lemma EC.1 that there exists $\boldsymbol{\mu} \ge \mathbf{0}$ such that $(\mathbf{y}, c_0) \in \mathcal{Y}_{\boldsymbol{\mu}}$. Suppose $(\mathbf{y}, c_0)$ is not an extreme point of $\mathcal{Y}_{\boldsymbol{\mu}}$, then there exist $(\mathbf{y}^1, c_0^1), (\mathbf{y}^2, c_0^2) \in \mathcal{Y}_{\boldsymbol{\mu}}$ and $0 < \lambda < 1$ such that $(\mathbf{y}, c_0) = \lambda(\mathbf{y}^1, c_0^1) + (1-\lambda)(\mathbf{y}^2, c_0^2)$. Notice that it follows from Lemma EC.1 that $(\mathbf{y}^1, c_0^1), (\mathbf{y}^2, c_0^2) \in \mathcal{Y}$. This shows that $(\mathbf{y}, c_0)$ is not an extreme point of $\mathcal{Y}$, which leads to a contradiction. Therefore, it must be the case that $(\mathbf{y}, c_0)$ is an extreme point of $\mathcal{Y}_{\boldsymbol{\mu}}$. $\square$

Notice that the uncertainty set for $\mathcal{Y}_{\boldsymbol{\mu}}$ only involves box constraint. Now we can use the above argument to reduce (4) with non-box constraint to a similar problem with box constraint by dualizing the budget constraint.

Another key observation is that the constraint (4c) for $t \le T+1-k$ only involves $\zeta_1, \ldots, \zeta_t$ and is independent of $\zeta_{T+1-k+1}, \zeta_{T+1-k+2}, \ldots, \zeta_{T+1}$. Furthermore, notice that for any $\zeta_1 \in \mathcal{U}_1, \zeta_2 \in \mathcal{U}_2, \ldots, \zeta_{T+1-k} \in \mathcal{U}_{T+1-k}$, we can find $\boldsymbol{\zeta} \equiv (\zeta_1, \ldots, \zeta_{T+1}) \in \mathcal{U}$ with the same value of $\zeta_1, \ldots, \zeta_{T+1-k}$ due to the definition of $\mathcal{U}$. This means that the effective uncertainty set for (4c) with $t \le T+1-k$ is exactly a box constraint. The same argument works for (4d) for $t \le T+1-k$.

Therefore, for any extreme point $\bar{\mathbf{y}}$ of the constraint set in (4), it follows from Lemma EC.2 that there exists $\mu^a \ge 0$, $\mu_e^b \ge 0$ for $e \in [E]$, $\underline{\mu}_{et}^c, \bar{\mu}_{et}^c \ge 0$ for $e \in [E]$, $t \in \{T+2-k, \ldots, T+1\}$, $\underline{\mu}_t^d, \bar{\mu}_t^d \ge 0$ for $t \in \{T+2-k, \ldots, T+1\}$, such that $\bar{\mathbf{y}}$ is an extreme point to the following set:

$$\sum_{s=1}^{T+1}\left(-\mu^a f_s + \sum_{t=s}^{T+1}\sum_{e=1}^{E} c_{te} y_{t,s,e}\right)\zeta_s \le c_0 - \mu^a g \tag{5a}$$

$$\sum_{s=1}^{T+1}\left(-\mu_e^b f_s + \sum_{t=s}^{T+1} y_{t,s,e}\right)\zeta_s \le Q_e - \mu_e^b g \qquad \forall e \in [E] \tag{5b}$$

$$0 \le \sum_{s=1}^{t} y_{t,s,e}\zeta_s \le p_{te} \quad \forall e \in [E],\ t \in \{1,...,T+1-k\} \tag{5c-i}$$

$$-\underline{\mu}_{et}^{c} g \le \sum_{s=1}^{t}\left(y_{t,s,e} - \underline{\mu}_{et}^{c} f_s\right)\zeta_s \quad \forall e \in [E],\ t \in \{T+2-k,...,T+1\} \tag{5c-ii-LB}$$

$$\sum_{s=1}^{t}\left(y_{t,s,e} - \bar{\mu}_{et}^{c} f_s\right)\zeta_s \le p_{te} - \bar{\mu}_{et}^{c} g \quad \forall e \in [E],\ t \in \{T+2-k,...,T+1\} \tag{5c-ii-UB}$$

$$\begin{aligned} V_{\min} \le v_1 &+ \left(\sum_{\ell=1}^{t}\sum_{e\in[E]} y_{\ell,1,e}\right)\zeta_1 \\ &+ \sum_{s=2}^{t}\left(-1 + \sum_{\ell=s}^{t}\sum_{e\in[E]} y_{\ell,s,e}\right)\zeta_s - \zeta_{t+1} \le V_{\max} \quad \forall t \in \{1,...,T+1-k\} \end{aligned} \tag{5d-i}$$

$$\begin{aligned} V_{\min} - \underline{\mu}_t^d g \le v_1 &+ \left(-\underline{\mu}_t^d f_1 + \sum_{\ell=1}^{t}\sum_{e\in[E]} y_{\ell,1,e}\right)\zeta_1 \\ &+ \sum_{s=2}^{t}\left(-1 - \underline{\mu}_t^d f_s + \sum_{\ell=s}^{t}\sum_{e\in[E]} y_{\ell,s,e}\right)\zeta_s - \zeta_{t+1} \quad \forall t \in \{T+2-k,...,T\} \end{aligned} \tag{5d-ii-LB}$$

$$\begin{aligned} v_1 &+ \left(-\bar{\mu}_t^d f_1 + \sum_{\ell=1}^{t}\sum_{e\in[E]} y_{\ell,1,e}\right)\zeta_1 \\ &+ \sum_{s=2}^{t}\left(-1 - \bar{\mu}_t^d f_s + \sum_{\ell=s}^{t}\sum_{e\in[E]} y_{\ell,s,e}\right)\zeta_s - \zeta_{t+1} \le V_{\max} - \bar{\mu}_t^d g \quad \forall t \in \{T+2-k,...,T\} \end{aligned} \tag{5d-ii-UB}$$

$$\forall \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_{T+1} \in \mathcal{U}_{T+1}.$$

Now notice that the uncertainty set is a box constraint, and we can utilize the same arguments as in the proof of Theorem 2 for the proof. We can create a similar set of equality systems as in (HARD-2a)-(EASY-2d-iv). Consider an extreme point $\bar{y}$ to the constraint set (5a)-(5d-ii-UB). It follows from Lemma 2 that there exists

- index sets $\mathcal{I}^{5b,\bar{\mathbf{y}}} \subseteq [E]$, $\underline{\mathcal{I}}^{5c-i,\bar{\mathbf{y}}}, \bar{\mathcal{I}}^{5c-i,\bar{\mathbf{y}}} \subseteq [T+1-k]\times[E]$, $\mathcal{I}^{5c-ii-LB,\bar{\mathbf{y}}}, \mathcal{I}^{5c-ii-UB,\bar{\mathbf{y}}} \subseteq \{T+2-k,...,T+1\}\times[E]$, $\underline{\mathcal{I}}^{5d-i,\bar{\mathbf{y}}}, \bar{\mathcal{I}}^{5d-i,\bar{\mathbf{y}}} \subseteq [T+1-k]$ and $\mathcal{I}^{5d-ii-LB,\bar{\mathbf{y}}}, \mathcal{I}^{5d-ii-UB,\bar{\mathbf{y}}} \subseteq \{T+2-k,...,T+1\}$;
- index sets $\mathcal{T}^{5a,\bar{\mathbf{y}}} \subseteq [T+1]$, $\mathcal{T}_e^{5b,\bar{\mathbf{y}}} \subseteq [T+1]$ for each $e \in \mathcal{I}^{5b,\bar{\mathbf{y}}}$, $\underline{\mathcal{T}}_{t,e}^{5c-i,\bar{\mathbf{y}}} \subseteq [T+1-k]$ for each $(t,e) \in \underline{\mathcal{I}}^{5c-i,\bar{\mathbf{y}}}$, $\bar{\mathcal{T}}_{t,e}^{5c-i,\bar{\mathbf{y}}} \subseteq [T+1-k]$ for each $(t,e) \in \bar{\mathcal{I}}^{5c-i,\bar{\mathbf{y}}}$, $\mathcal{T}_{t,e}^{5c-ii-LB,\bar{\mathbf{y}}} \subseteq \{T+2-k,...,T+1\}$ for each $(t,e) \in \mathcal{I}^{5c-ii-LB,\bar{\mathbf{y}}}$, $\mathcal{T}_{t,e}^{5c-ii-UB,\bar{\mathbf{y}}} \subseteq \{T+2-k,...,T+1\}$ for each $(t,e) \in \mathcal{I}^{5c-ii-UB,\bar{\mathbf{y}}}$, $\underline{\mathcal{T}}_t^{5d-i,\bar{\mathbf{y}}} \subseteq [T+1-k]$ for each $t \in \underline{\mathcal{I}}^{5d-i,\bar{\mathbf{y}}}$, $\bar{\mathcal{T}}_t^{5d-i,\bar{\mathbf{y}}} \subseteq [T+1-k]$ for each $t \in \bar{\mathcal{I}}^{5d-i,\bar{\mathbf{y}}}$, $\mathcal{T}_t^{5d-ii-LB,\bar{\mathbf{y}}} \subseteq \{T+2-k,...,T+1\}$ for each $t \in \mathcal{I}^{5d-ii-LB,\bar{\mathbf{y}}}$, $\mathcal{T}_t^{5d-ii-UB,\bar{\mathbf{y}}} \subseteq \{T+2-k,...,T+1\}$ for each $t \in \mathcal{I}^{5d-ii-UB,\bar{\mathbf{y}}}$;
- hyperplanes $(\boldsymbol{\alpha}^{5a,\bar{\mathbf{y}}}, \beta^{5a,\bar{\mathbf{y}}})$, $(\boldsymbol{\alpha}_e^{5b,\bar{\mathbf{y}}}, \beta_e^{5b,\bar{\mathbf{y}}})$ for each $e \in \mathcal{I}^{5b,\bar{\mathbf{y}}}$, $(\underline{\boldsymbol{\alpha}}_{t,e}^{5c-i,\bar{\mathbf{y}}}, \underline{\beta}_{t,e}^{5c-i,\bar{\mathbf{y}}})$ for each $(t,e) \in \underline{\mathcal{I}}^{5c-i,\bar{\mathbf{y}}}$, $(\bar{\boldsymbol{\alpha}}_{t,e}^{5c-i,\bar{\mathbf{y}}}, \bar{\beta}_{t,e}^{5c-i,\bar{\mathbf{y}}})$ for each $(t,e) \in \bar{\mathcal{I}}^{5c-i,\bar{\mathbf{y}}}$, $(\boldsymbol{\alpha}_{t,e}^{5c-ii-LB,\bar{\mathbf{y}}}, \beta_{t,e}^{5c-ii-LB,\bar{\mathbf{y}}})$ for each $(t,e) \in \mathcal{I}^{5c-ii-LB,\bar{\mathbf{y}}}$, $(\boldsymbol{\alpha}_{t,e}^{5c-ii-UB,\bar{\mathbf{y}}}, \beta_{t,e}^{5c-ii-UB,\bar{\mathbf{y}}})$ for each $(t,e) \in \mathcal{I}^{5c-ii-UB,\bar{\mathbf{y}}}$, $(\underline{\boldsymbol{\alpha}}_t^{5d-i,\bar{\mathbf{y}}}, \underline{\beta}_t^{5d-i,\bar{\mathbf{y}}})$ for each $t \in \underline{\mathcal{I}}^{5d-i,\bar{\mathbf{y}}}$, $(\bar{\boldsymbol{\alpha}}_t^{5d-i,\bar{\mathbf{y}}}, \bar{\beta}_t^{5d-i,\bar{\mathbf{y}}})$ for each $t \in \bar{\mathcal{I}}^{5d-i,\bar{\mathbf{y}}}$, $(\boldsymbol{\alpha}_t^{5d-ii-LB,\bar{\mathbf{y}}}, \beta_t^{5d-ii-LB,\bar{\mathbf{y}}})$ for each $t \in \mathcal{I}^{5d-ii-LB,\bar{\mathbf{y}}}$, $(\bar{\boldsymbol{\alpha}}_t^{5d-ii-UB,\mathbf{y}}, \beta_t^{5d-ii-UB,\bar{\mathbf{y}}})$ for each $t \in \mathcal{I}^{5d-ii-UB,\bar{\mathbf{y}}}$,

such that $\bar{\mathbf{y}}$ is the unique solution to the following system of equations:

$$\boldsymbol{\alpha}^{5a,\bar{\mathbf{y}}} \cdot \mathbf{y} = \beta^{5a,\bar{\mathbf{y}}} \tag{HARD-5a}$$

$$\boldsymbol{\alpha}_e^{5b,\bar{\mathbf{y}}} \cdot \mathbf{y} = \beta_e^{5b,\bar{\mathbf{y}}} \quad \forall e \in \mathcal{I}^{5b,\bar{\mathbf{y}}} \tag{HARD-5b}$$

$$\underline{\boldsymbol{\alpha}}_{t,e}^{5c-i,\bar{\mathbf{y}}} \cdot \mathbf{y} = \underline{\beta}_{t,e}^{5c-i,\bar{\mathbf{y}}} \quad \forall (t,e) \in \underline{\mathcal{I}}^{5c-i,\bar{\mathbf{y}}} \tag{HARD-5c-i-LB}$$

$$\bar{\boldsymbol{\alpha}}_{t,e}^{5c-i,\bar{\mathbf{y}}} \cdot \mathbf{y} = \bar{\beta}_{t,e}^{5c-i,\bar{\mathbf{y}}} \quad \forall (t,e) \in \bar{\mathcal{I}}^{5c-i,\bar{\mathbf{y}}} \tag{HARD-5c-i-UB}$$

$$\boldsymbol{\alpha}_{t,e}^{5c-ii-UB,\bar{\mathbf{y}}} \cdot \mathbf{y} = \beta_{t,e}^{5c-ii-UB,\bar{\mathbf{y}}} \quad \forall (t,e) \in \mathcal{I}^{5c-ii-UB,\bar{\mathbf{y}}} \tag{HARD-5c-ii-LB}$$

$$\boldsymbol{\alpha}_{t,e}^{5c-ii-LB,\bar{\mathbf{y}}} \cdot \mathbf{y} = \beta_{t,e}^{5c-ii-LB,\bar{\mathbf{y}}} \quad \forall (t,e) \in \mathcal{I}^{5c-ii-LB,\bar{\mathbf{y}}} \tag{HARD-5c-ii-UB}$$

$$\underline{\boldsymbol{\alpha}}_{t}^{5d-i,\bar{\mathbf{y}}} \cdot \mathbf{y} = \underline{\beta}_{t}^{5d-i,\bar{\mathbf{y}}} \quad \forall t \in \underline{\mathcal{I}}^{5d-i,\bar{\mathbf{y}}} \tag{HARD-5d-i-LB}$$

$$\bar{\boldsymbol{\alpha}}_{t}^{5d-i,\bar{\mathbf{y}}} \cdot \mathbf{y} = \bar{\beta}_{t}^{5d-i,\bar{\mathbf{y}}} \quad \forall t \in \bar{\mathcal{I}}^{5d-i,\bar{\mathbf{y}}} \tag{HARD-5d-i-UB}$$

$$\boldsymbol{\alpha}_{t}^{5d-ii-UB,\bar{\mathbf{y}}} \cdot \mathbf{y} = \beta_{t}^{5d-ii-UB,\bar{\mathbf{y}}} \quad \forall t \in \mathcal{I}^{5d-ii-UB,\bar{\mathbf{y}}} \tag{HARD-5d-ii-LB}$$

$$\boldsymbol{\alpha}_{t}^{5d-ii-LB,\bar{\mathbf{y}}} \cdot \mathbf{y} = \beta_{t}^{5d-ii-LB,\bar{\mathbf{y}}} \quad \forall t \in \mathcal{I}^{5d-ii-LB,\bar{\mathbf{y}}} \tag{HARD-5d-ii-UB}$$

$$\sum_{t=s}^{T+1} \sum_{e=1}^{E} c_{te} y_{t,s,e} = \mu^a f_s \quad \forall s \in \mathcal{T}^{5a,\bar{\mathbf{y}}} \tag{EASY-5a}$$

$$\sum_{t=s}^{T+1} y_{t,s,e} = \mu_e^b g \quad \forall e \in [E],\ s \in \mathcal{T}_e^{5b,\bar{\mathbf{y}}} \tag{EASY-5b}$$

$$y_{t,s,e} = 0 \quad \forall e \in [E],\ t \in [T+1-k],\ s \in \underline{\mathcal{T}}_{t,e}^{5c-i,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_{t,e}^{5c-i,\bar{\mathbf{y}}} \text{ if } s \le t \tag{EASY-5c-i}$$

$$y_{t,s,e} = \underline{\mu}_{et}^c f_s \quad \forall e \in [E],\ t \in \{T+2-k, T+1\},\ s \in \mathcal{T}_{t,e}^{5c-ii-LB,\bar{\mathbf{y}}} \text{ if } s \le t \tag{EASY-5c-ii-LB}$$

$$y_{t,s,e} = \bar{\mu}_{et}^c f_s \quad \forall e \in [E],\ t \in \{T+2-k, T+1\},\ s \in \mathcal{T}_{t,e}^{5c-ii-UB,\bar{\mathbf{y}}} \text{ if } s \le t \tag{EASY-5c-ii-UB}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]} y_{\ell,s,e} = 0 \quad \forall t \in [T+1-k],\ s \in \underline{\mathcal{T}}_t^{5d-i,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{5d-i,\bar{\mathbf{y}}} \text{ if } s = 1 \tag{EASY-5d-i-i}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]} y_{\ell,s,e} = 1 \quad \forall t \in [T+1-k],\ s \in \underline{\mathcal{T}}_t^{5d-i,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{5d-i,\bar{\mathbf{y}}} \text{ if } s \in \{2,\ldots,t\} \tag{EASY-5d-i-ii}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]} y_{\ell,s,e} = \underline{\mu}_t^d f_1 \quad \forall t \in \{T+2-k, T+1\},\ s \in \mathcal{T}_t^{5d-ii-LB,\bar{\mathbf{y}}} \text{ if } s = 1 \tag{EASY-5d-ii-LB-i}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]} y_{\ell,s,e} = 1 + \underline{\mu}_t^d f_s \quad \forall t \in \{T+2-k, T+1\},\ s \in \mathcal{T}_t^{5d-ii-LB,\bar{\mathbf{y}}} \text{ if } s \in \{2,\ldots,t\} \tag{EASY-5d-ii-LB-ii}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]} y_{\ell,s,e} = \bar{\mu}_t^d f_1 \quad \forall t \in \{T+2-k, T+1\},\ s \in \mathcal{T}_t^{5d-ii-UB,\bar{\mathbf{y}}} \text{ if } s = 1 \tag{EASY-5d-ii-UB-i}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]} y_{\ell,s,e} = 1 + \bar{\mu}_t^d f_s \quad \forall t \in \{T+2-k, T+1\},\ s \in \mathcal{T}_t^{5d-ii-UB,\bar{\mathbf{y}}} \text{ if } s \in \{2,\ldots,t\}. \tag{EASY-5d-ii-UB-ii}$$

With exactly the same argument as in the step 3 of the proof of Theorem 2, we have

$$\begin{aligned}|\mathcal{I}| \le\ & \underbrace{1}_{\text{(HARD-5a)}} + \underbrace{E}_{\text{(HARD-5b)}} + \underbrace{(T+1-k)E}_{\text{(HARD-5c-i-LB)}} + \underbrace{(T+1-k)E}_{\text{(HARD-5c-i-UB)}} + \underbrace{kE}_{\text{(HARD-5c-ii-LB)}} + \underbrace{kE}_{\text{(HARD-5c-ii-UB)}} \\ & + \underbrace{T+1-k}_{\text{(HARD-5d-i-LB)}} + \underbrace{T+1-k}_{\text{(HARD-5d-i-UB)}} + \underbrace{k-1}_{\text{(HARD-5d-ii-LB)}} + \underbrace{k-1}_{\text{(HARD-5d-ii-UB)}} + \underbrace{T+1}_{\text{(EASY-5a)}} + \underbrace{(T+1)E}_{\text{(EASY-5b)}}\end{aligned}$$

$$
\begin{aligned}
&+ \underbrace{kT}_{\text{(EASY-5c-ii-LB)}} + \underbrace{kT}_{\text{(EASY-5c-ii-UB)}} + \underbrace{T-k}_{\text{(EASY-5d-i-i)}} + \underbrace{T-k}_{\text{(EASY-5d-i-ii)}} \\
&+ \underbrace{kT}_{\text{(EASY-5d-ii-LB-i)}} + \underbrace{kT}_{\text{(EASY-5d-ii-LB-ii)}} + \underbrace{kT}_{\text{(EASY-5d-ii-UB-i)}} + \underbrace{kT}_{\text{(EASY-5d-ii-UB-ii)}} \\
&= 1 + 4E + 5T + 3ET + 6kT \ ,
\end{aligned}
$$

where we treat (EASY-5a), (EASY-5b), (EASY-5c-ii-LB), (EASY-5c-ii-UB), (EASY-5d-ii-LB-i), (EASY-5d-ii-LB-ii) (EASY-5d-ii-UB-i), (EASY-5d-ii-LB-ii), (EASY-5d-ii-UB-ii) as hard constraints, and we utilize the same argument as in the proof of Theorem 2 for (EASY-5d-i-ii). Similar to the proof of Theorem 2, it follows from Lemma 3 that

$$\|\bar{\mathbf{y}}\|_0 \le 2|\mathcal{I}| \le 2\left(1 + 4E + 5T + 3ET + 6kT\right),$$

which finishes the proof. □

Appendix E: Proof of Proposition 1

Proof of Proposition 1. Consider any active set $\mathcal{A}$. We observe for each $s \in \{1, \ldots, T+1\}$ that the quantities

$$-b_{i,s} + \sum_{t,j:(t,s,j)\in\mathcal{A}} a_{i,t,j} y_{t,s,j}$$

in the constraints of the optimization problem (LDR-1) are given by

$$0 + \sum_{t,e:(t,s,e)\in\mathcal{A}} c_{te} y_{tse} \tag{CONS-s-1a}$$

$$0 + \sum_{t:(t,s,e)\in\mathcal{A}} y_{tse} \qquad \forall e \in [E] \tag{CONS-s-1b}$$

$$0 + y_{tse} \qquad \forall t \in \{s, \ldots, T+1\}, e \in [E] : (t,s,e) \in \mathcal{A} \tag{CONS-s-1c-UB}$$

$$0 - y_{tse} \qquad \forall t \in \{s, \ldots, T+1\}, e \in [E] : (t,s,e) \in \mathcal{A} \tag{CONS-s-1c-LB}$$

$$0 + \sum_{\ell=1}^{t} \sum_{e\in[E]:(t,s,e)\in\mathcal{A}} y_{\ell se} \qquad \forall t \in [T] \quad [\text{if } s = 1] \tag{CONS-s-1d-UB-i}$$

$$0 - \sum_{\ell=1}^{t} \sum_{e\in[E]:(t,s,e)\in\mathcal{A}} y_{\ell se} \qquad \forall t \in [T] \quad [\text{if } s = 1] \tag{CONS-s-1d-LB-i}$$

$$-1 + \sum_{\ell=s}^{t} \sum_{e\in[E]:(\ell,s,e)\in\mathcal{A}} y_{\ell se} \qquad \forall t \in \{s, \ldots, T\} \quad [\text{if } s \in \{2, \ldots, T\}] \tag{CONS-s-1d-UB-ii}$$

$$1 - \sum_{\ell=s}^{t} \sum_{e\in[E]:(\ell,s,e)\in\mathcal{A}} y_{\ell se} \qquad \forall t \in \{s, \ldots, T\} \quad [\text{if } s \in \{2, \ldots, T\}] \tag{CONS-s-1d-LB-ii}$$

$$1 \qquad [\text{if } s \in \{2, \ldots, T+1\}] \tag{CONS-s-1d-UB-iii}$$

$$-1 \qquad [\text{if } s \in \{2, \ldots, T+1\}] \tag{CONS-s-1d-LB-iii}$$

We now count the number of unique quantities across the periods $s \in \{1,\dots,T+1\}$ of the optimization problem. First, we observe that the number of unique quantities among (CONS-s-1a), (CONS-s-1b), (CONS-s-1c-UB), and (CONS-s-1c-LB) across periods $s \in \{1,\dots,T+1\}$ is upper bounded by

$$\sum_{s=1}^{T+1}\left(\underbrace{1}_{\text{(CONS-}s\text{-1a)}} + \underbrace{\sum_{e\in[E]} 1}_{\text{(CONS-}s\text{-1b)}} + \underbrace{\sum_{t=s}^{T+1}\sum_{e\in[E]:(t,s,e)\in\mathcal{A}} 2}_{\text{(CONS-}s\text{-1c-UB)+(CONS-}s\text{-1c-LB)}} \right) = (E+1)(T+1)+2|\mathcal{A}|. \tag{EC.6}$$

Second, we observe for each $s \in \{2,\dots,T\}$ and $t \in \{s,\dots,T\}$ that

$$[(t,s,e)\notin\mathcal{A} \text{ for all } e\in[E]] \implies \left[\sum_{\ell=s}^{t-1}\sum_{e\in[E]:(\ell,s,e)\in\mathcal{A}} y_{\ell se} = \sum_{\ell=s}^{t}\sum_{e\in[E]:(\ell,s,e)\in\mathcal{A}} y_{\ell se}\right].$$

Therefore, we observe that the number of unique quantities among (CONS-s-1d-UB-i), (CONS-s-1d-LB-i), (CONS-s-1d-UB-ii), and (CONS-s-1d-LB-ii) across periods $s\in\{1,\dots,T\}$ is upper bounded by

$$\begin{aligned}\underbrace{\sum_{s=1}^{T}\left(2+\sum_{t\in\{s,\dots,T\}:\text{there exists } e\in[E] \text{ such that } (t,s,e)\in\mathcal{A}} 2\right)}_{\substack{\text{(CONS-}s\text{-1d-UB-i)+(CONS-}s\text{-1d-LB-i)}\\ \text{+(CONS-}s\text{-1d-UB-ii)+(CONS-}s\text{-1d-LB-ii)}}} &\le \sum_{s=1}^{T}\left(2+\sum_{t=s}^{T}\sum_{e\in[E]:(t,s,e)\in\mathcal{A}} 2\right)\\ &\le 2T+2|\mathcal{A}|.\end{aligned} \tag{EC.7}$$

Finally, we observe that the number of unique quantities among (CONS-s-1d-UB-iii) and (CONS-s-1d-LB-iii) is given by

$$\sum_{s=2}^{T+1} \underbrace{2}_{\text{(CONS-}s\text{-1d-UB-iii)+(CONS-}s\text{-1d-LB-iii)}} = 2T. \tag{EC.8}$$

Combining lines (EC.6), (EC.7), and (EC.8), we conclude for the production-inventory problem that

$$\begin{aligned}K^{\mathcal{A},1}+\dots+K^{\mathcal{A},T+1} &\le \underbrace{(E+1)(T+1)+2|\mathcal{A}|}_{\text{(EC.6)}} + \underbrace{2T+2|\mathcal{A}|}_{\text{(EC.7)}} + \underbrace{2T}_{\text{(EC.8)}}\\ &= 4|\mathcal{A}|+ET+5T+E+1.\end{aligned}$$

□

Appendix F: Proof of Propositions 2 and 3 and Lemma 6

Our proofs of Propositions 2 and 3 and Lemma 6 make use of several intermediary observations and results. First, we observe using the robust counterpart technique that the dual linear optimization reformulation of (LDR) is stated as follows:

$$\begin{array}{lll} \underset{\substack{\lambda_0,\dots,\lambda_m\in\mathbb{R},\\ \zeta_{0,s},\dots,\zeta_{m,s}\in\mathbb{R}\ \forall s\in[T]}}{\text{maximize}} & -\sum_{i=1}^{m} c_i\lambda_i - \sum_{i=0}^{m}\sum_{s=1}^{T} b_{i,s}\zeta_{i,s} & \\ \text{subject to} & \sum_{i=0}^{m} a_{i,t,j}\zeta_{i,s} = 0 & \forall 1\le s\le t\le T, j\in[n] \\ & \underline{D}_s\lambda_i \le \zeta_{i,s} \le \bar{D}_s\lambda_i & \forall i\in\{0,\dots,m\},\ s\in[T] \\ & \lambda_0 = 1 & \\ & \lambda_i \ge 0 & \forall i \in\{1,\dots,m\}. \end{array} \tag{D}$$

The optimal objective value of (D) is equal to the optimal objective value of (LDR), which implies that the optimal objective value of (D) is always less than or equal to the optimal objective value of (D-$\mathcal{A}$). Moreover, we observe that the dual of the linear optimization problem (P-$\mathcal{A}$') is given by

$$\begin{aligned}
\underset{\substack{\lambda_0,\ldots,\lambda_m \in \mathbb{R},\\ \zeta_{0,s},\ldots,\zeta_{m,s}\in\mathbb{R}\;\forall s\in[T]}}{\text{maximize}} \quad & -\sum_{i=1}^m c_i\lambda_i - \sum_{i=0}^m\sum_{s=1}^T b_{i,s}\zeta_{i,s} \\
\text{subject to} \quad & \sum_{i=0}^m a_{i,t,j}\zeta_{i,s} = 0 && \forall (t,s,j)\in\mathcal{A} \\
& \underline{D}_s\lambda_i \le \zeta_{i,s} \le \bar{D}_s\lambda_i && \forall i\in\{0,\ldots,m\},\; s\in[T] \\
& \lambda_0 = 1 \\
& \lambda_i \ge 0 && \forall i \in\{1,\ldots,m\}.
\end{aligned} \tag{D-$\mathcal{A}$'}$$

Finally, for the sake of convenience, we repeat the optimization problem (D-$\mathcal{A}$) below:

$$\begin{aligned}
\underset{\substack{\lambda_0,\ldots,\lambda_m \in \mathbb{R},\\ \zeta^{\mathcal{A},s}_{k,s}\in\mathbb{R}\;\forall s\in[T],k\in[K^{\mathcal{A},s}]}}{\text{maximize}} \quad & -\sum_{i=1}^m c_i\lambda_i - \sum_{s=1}^T\sum_{k=1}^{K^{\mathcal{A},s}} b^{\mathcal{A},s}_{k,s}\zeta^{\mathcal{A},s}_{k,s} \\
\text{subject to} \quad & \sum_{k=1}^{K^{\mathcal{A},s}} a^{\mathcal{A},s}_{k,t,j}\zeta^{\mathcal{A},s}_{k,s} = 0 && \forall (t,s,j)\in\mathcal{A} \\
& \underline{D}_s\left(\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i\right) \le \zeta^{\mathcal{A},s}_{k,s} \le \bar{D}_s\left(\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i\right) && \forall s\in[T], k\in[K^{\mathcal{A},s}] \\
& \lambda_0 = 1 \\
& \lambda_i \ge 0 && \forall i \in\{1,\ldots,m\}.
\end{aligned} \tag{D-$\mathcal{A}$}$$

In the following lemma, we provide a transformation of a feasible solution of the optimization problem (D-$\mathcal{A}$) into a (possibly-infeasible) solution for the optimization problem (D) with the same objective value.

Lemma EC.3. *Consider any feasible solution of* (D-$\mathcal{A}$)*. For each* $i\in\{0,\ldots,m\}$ *and* $s\in[T]$*, let*

$$\zeta_{i,s} \triangleq \left(\frac{\lambda_i}{\sum_{i'\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i')=\pi^{\mathcal{A},s}(i)}\lambda_{i'}}\right)\zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}$$

where we define $0/0$ *to be equal to* 0*. Then the following conditions are satisfied:*

$$-\sum_{i=1}^m c_i\lambda_i - \sum_{i=0}^m\sum_{s=1}^T b_{i,s}\zeta_{i,s} = -\sum_{i=1}^m c_i\lambda_i - \sum_{s=1}^T\sum_{k=1}^{K^{\mathcal{A},s}} b^{\mathcal{A},s}_{k,s}\zeta^{\mathcal{A},s}_{k,s} \tag{EC.9a}$$

$$\sum_{i=0}^m a_{i,t,j}\zeta_{i,s} = 0 \qquad \forall (t,s,j)\in\mathcal{A} \tag{EC.9b}$$

$$\underline{D}_s\lambda_i \le \zeta_{i,s} \le \bar{D}_s\lambda_i \qquad \forall i\in\{0,\ldots,m\},\; s\in[T] \tag{EC.9c}$$

$$\lambda_0 = 1 \tag{EC.9d}$$

$$\lambda_i \ge 0 \qquad \forall i\in\{1,\ldots,m\}. \tag{EC.9e}$$

Before presenting the proof of the above Lemma EC.3, let us offer an interpretation of its conditions. Condition (EC.9a) shows that the transformation from the above lemma yields a solution for the optimization problem (D) with the same objective value as the feasible solution for (D-$\mathcal{A}$). Conditions (EC.9b)-(EC.9e) show that the transformation yields a solution that satisfies all of the constraints of the optimization problem (D) with the possible exception of equality constraints of the form $\sum_{i=0}^m a_{i,t,j}\zeta_{i,s} = 0$ for some subset of tuples $(t,s,j) \notin \mathcal{A}$. Having interpreted those conditions, we now present the proof of Lemma EC.3:

Proof of Lemma EC.3. Consider any feasible solution of (D-$\mathcal{A}$). For each $i \in \{0,\dots,m\}$ and $s \in [T]$, let

$$\zeta_{i,s} \triangleq \left(\frac{\lambda_i}{\sum_{i' \in \{0,\dots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} \tag{EC.10}$$

where we define 0/0 to be equal to 0.

We first show that condition (EC.9a) holds. Indeed,

$$\begin{aligned}
-\sum_{i=1}^m c_i\lambda_i - \sum_{i=0}^m\sum_{s=1}^T b_{i,s}\zeta_{i,s} &= \sum_{i=1}^m c_i\lambda_i - \sum_{s=1}^T \sum_{k=1}^{K^{\mathcal{A},s}} \sum_{i\in\{0,\dots,m\}:\pi^{\mathcal{A},s}(i)=k} b_{i,s}\zeta_{i,s} \\
&= \sum_{i=1}^m c_i\lambda_i - \sum_{s=1}^T \sum_{k=1}^{K^{\mathcal{A},s}} b^{\mathcal{A},s}_{k,s} \sum_{i\in\{0,\dots,m\}:\pi^{\mathcal{A},s}(i)=k} \zeta_{i,s} \\
&= \sum_{i=1}^m c_i\lambda_i - \sum_{s=1}^T \sum_{k=1}^{K^{\mathcal{A},s}} b^{\mathcal{A},s}_{k,s} \sum_{i\in\{0,\dots,m\}:\pi^{\mathcal{A},s}(i)=k} \left(\frac{\lambda_i}{\sum_{i' \in \{0,\dots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} \\
&= \sum_{i=1}^m c_i\lambda_i - \sum_{s=1}^T \sum_{k=1}^{K^{\mathcal{A},s}} b^{\mathcal{A},s}_{k,s} \sum_{i\in\{0,\dots,m\}:\pi^{\mathcal{A},s}(i)=k} \left(\frac{\lambda_i}{\sum_{i' \in \{0,\dots,m\}: \pi^{\mathcal{A},s}(i') = k} \lambda_{i'}} \right) \zeta^{\mathcal{A},s}_{k,s} \\
&= \sum_{i=1}^m c_i\lambda_i - \sum_{s=1}^T \sum_{k=1}^{K^{\mathcal{A},s}} b^{\mathcal{A},s}_{k,s}\zeta^{\mathcal{A},s}_{k,s}.
\end{aligned}$$

The first and second equalities follow from algebra. The third equality follows from line (EC.10). The fourth and fifth equalities follow from algebra.

We first show that condition (EC.9b) holds. Indeed, we observe for each $(t,s,j) \in \mathcal{A}$ that

$$\begin{aligned}
\sum_{i=0}^m a_{i,t,j}\zeta_{i,s} &= \sum_{k=1}^{K^{\mathcal{A},s}} \sum_{i\in\{0,\dots,m\}:\pi^{\mathcal{A},s}(i)=k} a_{i,t,j}\zeta_{i,s} \\
&= \sum_{k=1}^{K^{\mathcal{A},s}} a^{\mathcal{A},s}_{k,t,j} \sum_{i\in\{0,\dots,m\}:\pi^{\mathcal{A},s}(i)=k} \zeta_{i,s} \\
&= \sum_{k=1}^{K^{\mathcal{A},s}} a^{\mathcal{A},s}_{k,t,j} \sum_{i\in\{0,\dots,m\}:\pi^{\mathcal{A},s}(i)=k} \left(\frac{\lambda_i}{\sum_{i' \in \{0,\dots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} \\
&= \sum_{k=1}^{K^{\mathcal{A},s}} a^{\mathcal{A},s}_{k,t,j} \sum_{i\in\{0,\dots,m\}:\pi^{\mathcal{A},s}(i)=k} \left(\frac{\lambda_i}{\sum_{i' \in \{0,\dots,m\}: \pi^{\mathcal{A},s}(i') = k} \lambda_{i'}} \right) \zeta^{\mathcal{A},s}_{k,s} \\
&= \sum_{k=1}^{K^{\mathcal{A},s}} a^{\mathcal{A},s}_{k,t,j}\zeta^{\mathcal{A},s}_{k,s} \\
&= 0.
\end{aligned}$$

The first and second equalities follow from algebra. The third equality follows from line (EC.10). The fourth and fifth equalities follow from algebra. The sixth equality follows from the fact that we are considering a feasible solution of the linear optimization problem (D-$\mathcal{A}$).

We next show that condition (EC.9c) holds. Indeed, we observe for each $i \in \{0,\ldots,m\}$ and $s \in [T]$ that

$$\begin{aligned}
\zeta_{i,s} &= \left(\frac{\lambda_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} \\
&\in \left[\left(\frac{\lambda_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \underline{D}_s \left(\sum_{i \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i) = k} \lambda_i \right), \right. \\
&\qquad \left. \left(\frac{\lambda_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \bar{D}_s \left(\sum_{i \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i) = k} \lambda_i \right) \right] \\
&= \left[\underline{D}_s \lambda_i, \bar{D}_s, \lambda_i \right].
\end{aligned}$$

The first equality follows from line (EC.10). The inclusion follows from the fact that we are considering a feasible solution of the linear optimization problem (D-$\mathcal{A}$). The second equality follows from algebra.

Because conditions (EC.9d) and (EC.9e) follow from the fact that we are considering a feasible solution of the linear optimization problem (D-$\mathcal{A}$), our proof of Lemma EC.3 is complete. □

Equipped with Lemma EC.3, we are now ready for the proofs of Propositions 2 and 3 and Lemma 6.

Proof of Proposition 2. Consider any optimal solution for (D-$\mathcal{A}$), and suppose that the solution satisfies

$$\sum_{i=0}^{m} a_{i,t,j} \left(\frac{\lambda_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} = 0 \quad \forall (t,s,j) \notin \mathcal{A}. \tag{1}$$

Now, for each $i \in \{0,\ldots,m\}$ and $s \in [T]$, let

$$\zeta_{i,s} \triangleq \left(\frac{\lambda_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \lambda_{i'}} \right) \zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}.$$

In that case, it follows from (1) that

$$\sum_{i=0}^{m} a_{i,t,j} \zeta_{i,s} = 0 \quad \forall (t,s,j) \notin \mathcal{A}. \tag{EC.11}$$

Therefore, it follows from line (EC.11) and Lemma EC.3 that there exists a feasible solution for the optimization problem (D) with the same objective value as the optimal objective value of (D-$\mathcal{A}$). Hence, we have shown that the optimal objective value of (D) is equal to the optimal objective value of (D-$\mathcal{A}$). Because the optimal objective value of (D-$\mathcal{A}$) is equal to the optimal objective value of (LDR), and because the optimal objective value of (D) is equal to the optimal objective value of (LDR), our proof of Proposition 2 is complete. □

Proof of Proposition 3. Let $\hat{\lambda}_0,\ldots,\hat{\lambda}_m \in \mathbb{R}$ and $\hat{\zeta}^{\mathcal{A},s}_{k,s} \in \mathbb{R}\, \forall s \in [T], k \in [K^{\mathcal{A},s}]$ denote an optimal solution for the optimization problem (D-$\mathcal{A}$). Therefore, it follows from Lemma EC.3 and from the fact that the optimal objective value of (D-$\mathcal{A}$) is equal to the optimal objective value of (D-$\mathcal{A}$') that an optimal solution for the optimization problem (D-$\mathcal{A}$') is given by $\bar{\lambda}_0,\ldots,\bar{\lambda}_m \in \mathbb{R}$ and $\bar{\zeta}_{i,s} \in \mathbb{R}\, \forall s \in [T], k \in [K^{\mathcal{A},s}]$, where

$$\begin{aligned}
\bar{\lambda}_i &\triangleq \hat{\lambda}_i && \forall i \in \{0,\ldots,m\}; \\
\bar{\zeta}_{i,s} &\triangleq \left(\frac{\hat{\lambda}_i}{\sum_{i' \in \{0,\ldots,m\}: \pi^{\mathcal{A},s}(i') = \pi^{\mathcal{A},s}(i)} \hat{\lambda}_{i'}} \right) \hat{\zeta}^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} && \forall i \in \{0,\ldots,m\}, s \in \{1,\ldots,T\}.
\end{aligned}$$

With the above notation, let us define

$$\begin{aligned}\mathcal{A}^{=} &\triangleq \left\{(t,s,j)\notin\mathcal{A}: \quad \sum_{i=0}^{m} a_{i,t,j}\left(\frac{\hat{\lambda}_i}{\sum_{i'\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i')=\pi^{\mathcal{A},s}(i)}\hat{\lambda}_{i'}}\right)\hat{\zeta}^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}=0\right\}\\ &=\left\{(t,s,j)\notin\mathcal{A}: \quad \sum_{i=0}^{m} a_{i,t,j}\bar{\zeta}_{i,s}=0\right\}.\end{aligned}$$

If $\widetilde{\mathcal{A}}\subseteq\mathcal{A}^{=}$, then we observe that $\bar{\lambda}_0,\ldots,\bar{\lambda}_m\in\mathbb{R}$ and $\bar{\zeta}_{i,s}\in\mathbb{R}\;\forall s\in[T], k\in[K^{\mathcal{A},s}]$ must also be an optimal solution for the linear optimization problem (D-$\mathcal{A}\cup\widetilde{\mathcal{A}}$'). Since the optimal objective value of (D-$\mathcal{A}\cup\widetilde{\mathcal{A}}$') is equal to the optimal objective value of (LDR-$\mathcal{A}\cup\widetilde{\mathcal{A}}$), our proof is complete. □

Proof of Lemma 6. We recall from Assumption 1 that the optimal objective value of (LDR) is finite. Moreover, we observe that the optimal objective value of (LDR) is equal to the optimal objective value of (D), which implies that the optimal objective value of (LDR) is finite. Finally, we observe that the set of feasible solutions of the linear optimization problem (D-$\mathcal{A}$) for any active set $\mathcal{A}$ is a superset of the set of feasible solutions for the linear optimization problem (D). Hence, we conclude that (D-$\mathcal{A}$) always has at least one feasible solution. □

Appendix G: Proofs from §4.4

Proof of Lemma 4. To efficiently determine whether line (1) is satisfied, let the matrix $\mathbf{A}_i\triangleq(\mathbf{a}_{i,1}^\intercal,\ldots,\mathbf{a}_{i,T}^\intercal)\in\mathbb{R}^{T\times n}$ be defined for each $i\in\{0,\ldots,m\}$. We observe from algebra that the following equalities hold for each $(t,s,j)\notin\mathcal{A}$:

$$\begin{aligned}\sum_{i=0}^{m} a_{i,t,j}\left(\frac{\lambda_i}{\sum_{i'\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i')=\pi^{\mathcal{A},s}(i)}\lambda_{i'}}\right)\zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s} &= \sum_{k=1}^{K^{\mathcal{A},s}}\underbrace{\left(\frac{\zeta^{\mathcal{A},s}_{k,s}}{\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i}\right)}_{\alpha^{\mathcal{A},s}_{k,s}}\left(\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k} a_{i,t,j}\lambda_i\right)\\ &=\sum_{k=1}^{K^{\mathcal{A},s}}\alpha^{\mathcal{A},s}_{k,s}\left(\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k} a_{i,t,j}\lambda_i\right)\\ &=\sum_{k=1}^{K^{\mathcal{A},s}}\alpha^{\mathcal{A},s}_{k,s}\left[\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i\mathbf{A}_i\right]_{t,j} .\alpha^{\mathcal{A},s}_{k,s}\beta^{\mathcal{A},s}_{k,t,j}.\end{aligned}$$

Therefore, determining whether line (1) is satisfied can be split into the following three parts.

(a) We compute the quantity $\alpha^{\mathcal{A},s}_{k,s}$ for each $s\in[T]$ and $k\in[K^{\mathcal{A},s}]$.

(b) For each period $s\in[T]$:

  (i) We compute the sparse matrix $\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i\mathbf{A}_i$ for each $k\in[K^{\mathcal{A},s}]$.

  (ii) We compute $\sum_{k=1}^{K^{\mathcal{A},s}}\alpha^{\mathcal{A},s}_{k,s}\left[\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i\mathbf{A}_i\right]_{t,j}$ for each t,j such that $(t,s,j)\notin\mathcal{A}$.

We observe that the computation time and space for performing part (a) is $\mathcal{O}(K^{\mathcal{A}})$. We observe that each iteration of part (b-i) requires a total of $\mathcal{O}(Z)$ time and space to compute and store the sparse matrices $\sum_{i\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i)=k}\lambda_i\mathbf{A}_i$ for each $k\in[K^{\mathcal{A},s}]$. Since the memory for storing the sparse matrices can be reused across iterations of part (b-i), and since part (b-i) will be performed in T iterations, we conclude that a total of $\mathcal{O}(TZ)$ time and $\mathcal{O}(Z)$ space is needed to perform part (b-i) across all iterations. Finally, we observe part (b-ii) for each $(t,s,j)\notin\mathcal{A}$ requires $\mathcal{O}(K^{\mathcal{A},s})$ time and $\mathcal{O}(1)$ space. Since part (b-ii) will be performed once for each $(t,s,j) \notin \mathcal{A}$, we observe that the total computation time for performing part (b-ii) across all iterations is

$$\mathcal{O}\left(\sum_{(t,s,j)\notin\mathcal{A}} K^{\mathcal{A},s}\right) = \mathcal{O}\left(\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{t=s}^{T} K^{\mathcal{A},s}\right) = \mathcal{O}\left(\sum_{j=1}^{n}\sum_{s=1}^{T} TK^{\mathcal{A},s}\right) = \mathcal{O}\left(nTK^{\mathcal{A}}\right),$$

and the total space for performing part (b-ii) across all iterations is

$$\mathcal{O}\left(\sum_{(t,s,j)\notin\mathcal{A}} K^{\mathcal{A},s}\right) = \mathcal{O}(nT^2).$$

Combining the above analysis, we conclude that the total computation time for computing the quantities $\sum_{i=0}^{m} a_{i,t,j}\left(\frac{\lambda_i}{\sum_{i'\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i')=\pi^{\mathcal{A},s}(i)}\lambda_{i'}}\right)\zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}$ for each $(t,s,j)\notin\mathcal{A}$ is

$$\mathcal{O}\left(\underbrace{K^{\mathcal{A}}}_{\text{(a)}} + \underbrace{TZ}_{\text{(b-i)}} + \underbrace{nTK^{\mathcal{A}}}_{\text{(b-ii)}}\right) = \mathcal{O}\left(T\left(Z + nK^{\mathcal{A}}\right)\right),$$

and the total space for computing the quantities $\sum_{i=0}^{m} a_{i,t,j}\left(\frac{\lambda_i}{\sum_{i'\in\{0,\ldots,m\}:\pi^{\mathcal{A},s}(i')=\pi^{\mathcal{A},s}(i)}\lambda_{i'}}\right)\zeta^{\mathcal{A},s}_{\pi^{\mathcal{A},s}(i),s}$ for each $(t,s,j)\notin\mathcal{A}$ is

$$\mathcal{O}\left(\underbrace{K^{\mathcal{A}}}_{\text{(a)}} + \underbrace{Z}_{\text{(b-i)}} + \underbrace{nT^2}_{\text{(b-ii)}}\right) = \mathcal{O}\left(Z + nT^2\right).$$

Our proof of Lemma 4 is complete. $\square$

Proof of Lemma 5. Suppose for the sake of developing a contradiction that the optimal objective value of the linear optimization problem (D-$\mathcal{A}$) is changed after removing the constraints from line (2). Then it follows from strong duality for linear optimization that there must exist an optimal solution for (P-$\mathcal{A}$) and a corresponding tuple $(k,s)\in\mathcal{C}$ such that the optimal solution either satisfies $\bar{\omega}^{\mathcal{A},s}_{k,s} > 0$ or satisfies $\underline{\omega}^{\mathcal{A},s}_{k,s} > 0$, but not both. However, every feasible solution of the linear optimization problem (P-$\mathcal{A}$) must satisfy the following constraints:

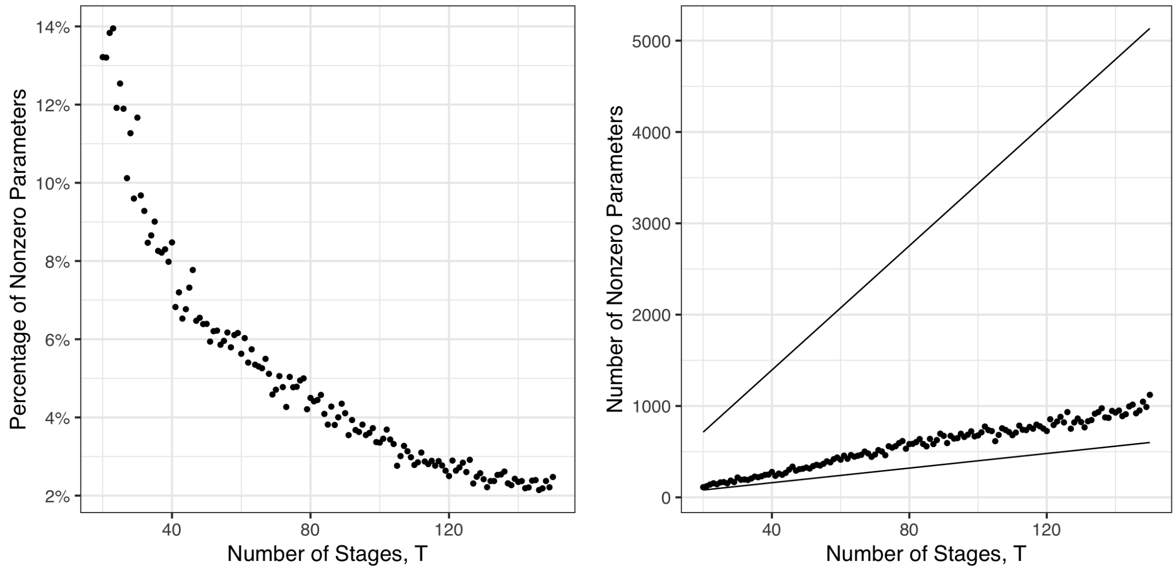
$$\begin{aligned} &\bar{\omega}^{\mathcal{A},s}_{k,s} - \underline{\omega}^{\mathcal{A},s}_{k,s} = -b^{\mathcal{A},s}_{k} + \sum_{t,j:(t,s,j)\in\mathcal{A}} a^{\mathcal{A},s}_{k,t,j} y_{t,s,j} && \forall (k,s)\in\mathcal{C} \\ &\bar{\omega}^{\mathcal{A},s}_{k,s}, \underline{\omega}^{\mathcal{A},s}_{k,s} \ge 0 && \forall (k,s)\in\mathcal{C} \end{aligned}$$

Since the equality $(b^{\mathcal{A},s}_{k,s}, (a^{\mathcal{A},s}_{k,t,j})_{t,j:(t,s,j)\in\mathcal{A}}) = (0,(0)_{t,j:(t,s,j)\in\mathcal{A}})$ holds for all $(k,s)\in\mathcal{C}$, it follows from the above reasoning that every feasible solution of the linear optimization problem (P-$\mathcal{A}$) must satisfy the equality $\bar{\omega}^{\mathcal{A},s}_{k,s} = \underline{\omega}^{\mathcal{A},s}_{k,s}$ for all $(k,s)\in\mathcal{C}$. Therefore, we conclude that there must exist an optimal solution of the linear optimization problem (P-$\mathcal{A}$) and a corresponding tuple $(k,s)\in\mathcal{C}$ such that the optimal solution satisfies $\bar{\omega}^{\mathcal{A},s}_{k,s}, \underline{\omega}^{\mathcal{A},s}_{k,s} > 0$. We have thus obtained a contradiction, which completes the proof of Lemma 5. $\square$

Appendix H: Additional Numerical Results

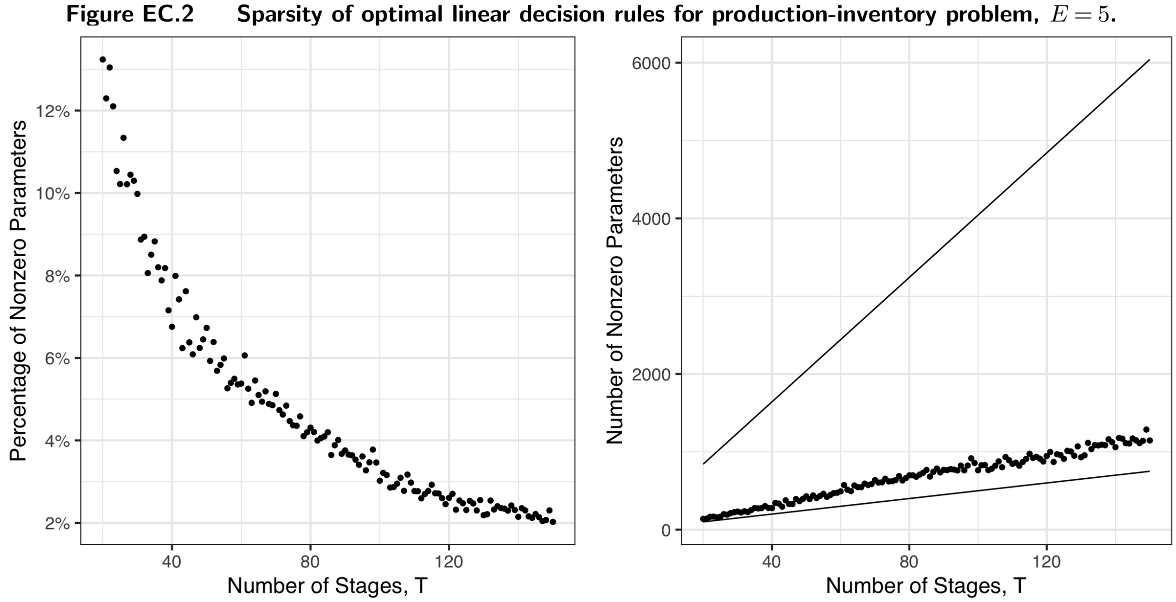
In Figures EC.1 and EC.2, we present numerical results which are similar to those from Figure 1 for cases where the number of factories is $E=4$ and $E=5$. The results show that the findings from Figure 1 are not exclusive to the case with $E=3$ factories.

Figure EC.1 Sparsity of optimal linear decision rules for production-inventory problem, $E = 4$.



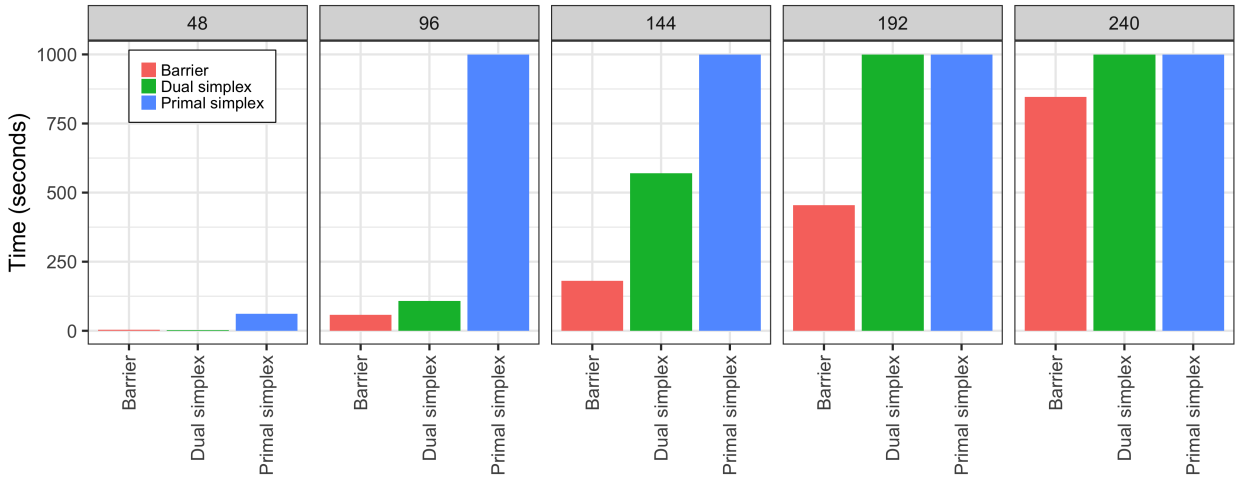
Note. Each point represents the optimal linear decision rules computed for the corresponding number of stages T and for $E = 4$ factories. Left figure shows the percentage of parameters of optimal linear decision rules which are nonzero. Right figure shows the number of nonzero parameters in optimal linear decision rules compared to the upper bound from Theorem 1 (top solid black line) and the number of parameters in static decision rules (bottom solid black line).

Figure EC.2 Sparsity of optimal linear decision rules for production-inventory problem, $E = 5$.



Note. Each point represents the optimal linear decision rules computed for the corresponding number of stages T and for $E = 5$ factories. Left figure shows the percentage of parameters of optimal linear decision rules which are nonzero. Right figure shows the number of nonzero parameters in optimal linear decision rules compared to the upper bound from Theorem 1 (top solid black line) and the number of parameters in static decision rules (bottom solid black line).

Figure EC.3 Computation times for robust counterpart with $E = 5$ factories.



Note. Results shown for experiments with $E = 5$ factories and $T \in \{48, 96, 144, 192, 240\}$ time periods. Bars show the computation time (in seconds) for solving the robust counterpart using primal simplex, dual simplex, and barrier method. All methods are either solved to optimality or terminated at 1000 seconds.

In Figure EC.3, we present numerical results which are similar to those from Figure 2 for solving the robust counterpart with Gurobi using primal simplex, dual simplex, and barrier methods. In these additional numerical experiments, the methods are run to optimality (but terminated if the computation time exceeds 1000 seconds). In all of the experiments in Figure EC.3 for which the computation time is equal to 1000 seconds, the solver returned neither a primal feasible solution nor a dual feasible solution. These additional experiments thus demonstrate that the slow computation time of the robust counterpart in Figure 2 is not a consequence of using barrier method to solve the linear optimization problem obtained from the robust counterpart technique, but rather a consequence of the shear size of the linear optimization problem that must be solved.