Distributionally Robust Inventory Management
with Advance Purchase Contracts

Yilin Xue\textsuperscript{1}, Yongzhen Li\textsuperscript{2}, Napat Rujeerapaiboon\textsuperscript{1}
\textsuperscript{1}Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore
\textsuperscript{2}Department of Management Science and Engineering, School of Economics and Management, Southeast University, China
yilin.xue@u.nus.edu, liyongzhen@seu.edu.cn, napat.rujeerapaiboon@nus.edu.sg

We propose a distributionally robust inventory model for finding an optimal ordering policy that attains the minimum worst-case expected total cost. In a classical stochastic setting, this problem is typically addressed by dynamic programming and is solved by the famous base-stock policy. This approach however crucially relies on two controversial assumptions: the demands are serially independent and the demand distribution is perfectly known. Aiming to address these issues, inspired by the seminal work of Scarf (1958), we adopt a mean-variance ambiguity set that imposes neither the shape of each marginal demand distribution nor their independence structure, and we focus on the case of advance purchase agreements which are prevalent in the robust inventory literature and have drawn renewed attention because of the Covid-19 vaccine procurement.

The proposed distributionally robust inventory model provably reduces to a finite conic optimization problem with however an exponential number of constraints. To gain tractability and to err on the safe side, we propose two conservative approximations. The first approximation is obtained by recognizing the problem as an artificial two-stage robust optimization problem and then by restricting each adaptive decision to a linear decision rule. The second approximation, on the other hand, is obtained by a constraint partitioning and by upper bounding each resultant maximum sum with a sum of maxima. We then present a progressive approximation based on a scenario reduction technique to gauge the quality of the proposed conservative approximations. We prove that this progressive approximation is exact when the inventory problem consists of two periods, and besides we use it to show that our conservative solutions are still close to being optimal when the planning horizon is longer. All of our exact and approximate inventory models are expressed as standard conic programs which allow for the incorporation of additional distributional information. The extensions are readily obtained by deriving a new cone that corresponds to the restricted ambiguity set and embedding it in the original problems. We analytically derive the worst-case demand distribution from the mean-variance ambiguity set and numerically use it to show that our robust inventory policy is more resilient to the misspecification of the demand distribution than the state-of-the-art non-robust policies.

\textit{Keywords}: Inventory management; distributionally robust optimization; conic programming.

\section{Introduction}

We consider an inventory management problem where the manager is monetarily liable for holding a non-empty stock and/or for not being able to meet the demand in addition to the purchasing cost.
We assume that the problem consists of $T \in \mathbb{N}$ periods, and we denote the amount of the product to be delivered in period $t \in \{1, \ldots, T\}$ by $x_t \geq 0$. In the simplest case where $T = 1$, the problem contains a single decision variable $x_1$ and is intrinsically similar to a newsvendor problem (Shapiro et al. 2014). Typically, the solution to the risk-neutral newsvendor problem could be characterized by the inverse of the cumulative distribution function of the unknown demand $\tilde{\xi}_1$. However, the reliance on the demand distribution, which itself is difficult to accurately empirically estimate, may limit the practical relevance of this formula. Scarf (1958) addressed this issue by seeking to find an optimal $x_1^*$ that performs best in view of the worst-case probability distribution amongst all those that share the same mean $\mu$ and variance $\sigma^2$. This seminal work has been extended in several directions. For instance, Gallego and Moon (1993) allowed the newsvendor to make a second purchase even after the demand is realized, whereas Ben-Tal and Hochman (1976) and Das et al. (2021) adopted a different ambiguity set of plausible demand distributions by replacing the assumption of a known variance with the assumption of a known mean absolute deviation and a known $n^{th}$, $n > 1$, moment of the demand distribution, respectively. Besides, Gallego and Moon (1993) emphasized the practical relevance of the mean-variance ambiguity set by highlighting that, if the demand distribution is normal, then the expected value of distributional information (EVDI), which is defined as the gap between the expected costs attained by the stochastically optimal solution and the robustly optimal solution due to Scarf (1958), is seemingly negligible. We remark that EVDI has also been used as a decision making criterion; see e.g. Yue et al. (2006). For the sake of completeness, while our focus is on the mean-variance ambiguity set, which has an advantage of being characterized by only a few statistics, there are other non-parametric newsvendor models which directly incorporate the historical observations of the demand and/or other influential features and simultaneously account for the demand distribution ambiguity, such as Chen and Xie (2021) and Lee et al. (2021) as well as Fu et al. (2021).

For a multi-period problem ($T \geq 2$), we again consider a risk-neutral inventory manager whose objective is still to minimize the total expected cost. When the demands are serially independent and follow a known distribution, the purchasing costs consist of a fixed and a linearly variable part, and the holding and the backlogging costs are linear, Scarf (1960) formulated this problem as a dynamic program and showed that there exists an optimal base-stock policy, which is characterized by a collection of reordering points $s \in \mathbb{R}^T$ and order-up-to levels $S \in \mathbb{R}^T$ ($S \geq s$). In the absence of the fixed ordering cost, it can be imposed without loss of optimality that $s = S$, and thus the optimal policy simplifies; see e.g. Bertsekas (1995). Despite the breakthrough in this characterization of the optimal policy, the computation is provably difficult when $T < \infty$ and when the demand distribution is discrete (Halman et al. 2009), and only a few specific variants of the dynamic inventory problem admit a reasonably efficient exact solution approach (Veinott and Wagner 1965, Federgruen and
Zipkin 1984, Zheng and Federgruen 1991; all of which assume $T = \infty$), and otherwise one may need to resort to a simulation-based method (Fu 1994) or to interpolating the value functions (Halman et al. 2009).

In reality, the demand distribution is hardly available, but it is required for the computation of the optimal policy. Incorporating inaccurate information (such as, using an empirical distribution in lieu of the true demand distribution) in a dynamic program systematically induces error propagation, which is increasingly difficult to keep track of with larger $T$; see e.g. Zhang et al. (2021). Because of this, several authors have turned towards robust (e.g. Ben-Tal et al. 2005) and distributionally robust (e.g. See and Sim 2010, Bertsimas et al. 2019) optimization. To gain tractability, these authors solved their respective variant of the inventory problem using a decision rule approximation. Though, Lu and Sturt (2022) showed that the majority of linear decision rule coefficients which arise in different variants of the robust inventory problem optimally vanish. Solyalı et al. (2016) relied on a different formulation which is inspired by the facility layout and knapsack problem to improve the computational efficiency in the presence of the fixed ordering cost. As an alternative solution approach, Xin and Goldberg (2022) assumed that the product demands are a bounded martingale and derived the optimal policy which can be interpreted as a base-stock policy where each order-up-to level is dependent on the most recent demand.

“Unless commitment is made, there are only promises and hopes; but no plans.”

— Drucker (1974)

Moreover, despite the undoubted usefulness of a dynamic inventory policy, such as the base-stock policy, it may be difficult to bring this forth as a supply chain contract. Indeed, Mamani et al. (2017) argued that, when products have short life cycles (e.g. electronic gadgets and apparels), the amount of purchasing opportunities may not be aplenty. Likewise, Basciftci et al. (2021) further argued that, in order to make a better arrangement of their operations and to minimize disruptions, suppliers typically prefer decisive contractual partners who will make their requirements (e.g., how much and when the products should be delivered) known early. All of these arguments show the significance and the relevance of an advance purchase contract, where the inventory manager is to submit a single purchase order to the supplier with however an option to specify different delivery dates. To give a contemporary example, the global population is currently rampaged by Covid-19, and virtually every nation competes to acquire sufficient doses of protective vaccines. Advance purchase agreements have repeatedly been signed by government representatives and biotechnology companies; see e.g. Jalelah (2021) and Andres (2022).

In this paper, we will focus on the optimal management of inventory levels which evolve in accordance to the advance purchase agreement, which is primarily suitable in the aforementioned context.
and when the planning horizon $T$ is only moderately large. When no other distributional information besides the support of the uncertain demand is available, Bertsimas and Thiele (2006) identified a robustly optimal policy and interpreted it as a base-stock policy, and subsequently, Mamani et al. (2017) determined a closed-form solution of the same problem when the uncertainty set satisfies a certain symmetrical criterion. It should however be noted that the robust model adopted in these two papers is superfluously conservative as it imposes that both the delivery quantities $\{x_t\}_{t=1}^T$ and the per-period worst-case holding and the backlogging cost incurred $\{p_t\}_{t=1}^T$ are static variables in the sense that

$$p_t = \max_{\xi} \left\{ \max \left\{ h \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \right\} \right\},$$

where $y_0$ denotes the initial inventory level and $(h, b)$ contains the unit cost of holding and backlogging, respectively. Here, the maximization optimizes over all possible realizations of the demands $\xi$ within the uncertainty set, and the inventory manager will seek to minimize the sum of the total purchasing cost and $\sum_{t=1}^T p_t$. In mathematical actuality, the per-period $p_t$ could and should be adaptive to the previously revealed demands $\{\xi_\tau\}_{\tau=1}^t$ in the sense that

$$p_t(\xi_1, \ldots, \xi_t) = \max \left\{ h \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \right\},$$

and the inventory manager should instead minimize the summation of the total purchasing cost and $\max_{\xi} \sum_{t=1}^T p_t(\xi_1, \ldots, \xi_t)$ as this latter approach will give a single worst-case scenario. We refer our readers to Gorissen and den Hertog (2013) for a more detailed discussion on this subtle intricacy, which was overcome by tighter approximations from Gorissen and den Hertog (2013) and Ardestani-Jaafari and Delage (2016) or by iterative heuristics from Bienstock and Özbay (2008) and Rodrigues et al. (2021). To our knowledge, these techniques are not directly applicable to the distributionally robust variant of the problem considered in this paper, which takes a viewpoint that nature chooses the worst-case demand distribution as opposed to the worst-case demand scenario.

Despite these exciting developments in developing and solving various robust inventory models, they are ignorant of any possible distributional information, such as moments, and are thus arguably overly conservative. To alleviate this conservativeness, we will choose to develop a distributionally robust inventory model in this paper, where in line with the above literature we will assume that the purchase orders are static decisions. In particular, we aim at minimizing the worst-case expected total cost of managing an inventory, which is obtained by maximizing the expected total cost over the planning horizon across all possible demand distributions that are consistent with a marginal mean $\mu$ and variance $\sigma^2$. We remark that the mean-variance ambiguity set has been extensively studied in the past, see e.g. Delage and Ye (2010), Zymler et al. (2013), and similar results are
also available for a more general moment-based ambiguity set, see e.g. Bertsimas and Popescu (2005). Besides, we note that in this paper the holding and the backlogging costs for all periods are modelled faithfully in the sense that they can be adaptive to the unfolding demand observations.

Relying on the duality theory for generalized moment problems, we formulate our distributionally robust inventory problem as a finite convex optimization problem, however, with an exponential number of constraints. In order to gain tractability, we propose two conservative approximations of the problem. The first conservative approximation is due to us interpreting the problem as a concise two-stage robust optimization problem and then approximating each second-stage decision with a linear decision rule (Ben-Tal et al. 2004). This approximation is inspired by a clever trick of Roos et al. (2020) who observed that wait-and-see decisions can be artificially introduced to a robust optimization problem that does not admit an exact tractable robust counterpart in order to exploit approximation techniques tailored for solving adaptive robust problems. The second conservative approximation is due to a meticulous partitioning of constraints and a series of interchanges between summation and maximization. In the remainder of the paper, we will henceforth refer to these two approximations as ‘L-conservative’ and ‘Q-conservative’, respectively. In addition, we also come up with a scenario reduction technique (Hadjjiyiannis et al. 2011) that helps us develop a progressive approximation of the proposed model, from which we can calculate an optimality gap of the conservative solutions that are due to the L- and Q-approximations. To our knowledge, computing the optimality gap is an issue of significant importance, but it has been often overlooked in the robust inventory management literature.

All of our exact and approximate models are representable as standard conic optimization problems. Adopting the plain vanilla mean-variance ambiguity set, the cone therein is an intersection of several second-order cones. It is worth mentioning that these models, thanks to their conic representation, can be readily extended in several dimensions. We give two explicit examples in the paper: (i) when the demands are almost surely non-negative (i.e., when the inventory manager implements a no-return policy) and (ii) when the demands are uncorrelated. In both cases, the exact model still appears intractable and its tractable approximations, both progressive and conservative, can be effortlessly derived by changing the cone. We find that extension (ii) considerably lowers the worst-case total expected cost. This observation is line with our anticipation as the worst-case demand distribution we derive from the plain vanilla model shows a strong dependency pattern.

Finally, as the dynamics of the inventory level is a cornerstone of many queuing models, our techniques could have a far-reaching impact on other operational problems of similar nature, such as an appointment scheduling problem; see Mak et al. (2015) and Padmanabhan et al. (2021). These papers aim at providing second-order cone and semidefinite programs for minimizing the worst-case
expected waiting and overtime cost when the service times are ambiguously random. This appointment scheduling problem involves a convex maximization subproblem, which had been thought to be intractable, until Mak et al. (2015) ingeniously formulated this subproblem as a mixed-integer linear program with a totally unimodular constraint matrix and is consequently equivalent to its linear program relaxation. Notwithstanding that we are unable to extend this exact solution method, our proposed conic approximations are numerically accurate and have the flexibility to be used with other types of the ambiguity set with relative ease as discussed above.

For succinctness, we summarize the main contributions of the paper below.

- To the best of our knowledge, we are the first to consider the distributionally robust inventory problem, where our decisions consist of the delivery amount $x_t$ for each period $t$ in the planning horizon $\{1, \ldots, T\}$, under the ambiguity set that is specified by the marginal means and variances of the demands. We then show that this problem admits an equivalent reformulation as a robust second-order cone program with a discrete uncertainty set which is characterized by the unit holding and backlogging costs.

- We then quantify the optimality gaps of our $L$- and $Q$- conservative solutions with the proposed progressive approximation. Numerically, these gaps appear insignificant, and thus the studied distributionally robust inventory problem can be solved almost exactly. Additionally, we give a rationale behind the progressive approximation by proving that it is exact when $T = 2$.

- We demonstrate that both the exact and the approximate reformulations of the distributionally robust inventory problem can be presented compactly using a cone that is representable as an intersection of several standard second-order cones. By means of a cone replacement, these conic representations are highly flexible in that they can incorporate additional distributional information of the demands, such as non-negativity and pairwise uncorrelatedness.

- Besides, we identify the worst-case demand distribution from the ambiguity set that attains the maximum expected total cost. We utilize it in a stress test to show that our distributionally robust inventory policy is more resilient than the stochastic policy that assumes an incorrect demand distribution. In addition, an adaptive version of our policy which can be obtained by re-solving the problem repeatedly in a shrinking horizon fashion, is almost on par with the optimal base-stock policy even when the probability distribution of the demand is known, is correct, and is serially independent.

The remainder of the paper is structured as follows. Section 2 details the distributionally robust inventory problem we are interested in as well as several other representations of the problem. One of them is then used to derive the worst-case distribution that plays an important role in a stress test analysis. Sections 3 and 4 discuss the proposed conservative and progressive approximations of the exact model, respectively. The conic extensions of our model are presented in Section 5,
and finally Section 6 reports the results of our numerical studies which measure the quality of the proposed approximations and compare our approach with the state-of-the-art benchmarks.

**Notation:** We use boldface lowercase letters (e.g. \( \mathbf{v} \)) and uppercase letters (e.g. \( M \)) to represent vectors and matrices, respectively. Specifically, \( \mathbf{1} \) (0) denotes a vector or a matrix of all ones (zeros), and \( \mathbf{1}_i \) denotes the \( i \)th canonical basis vector. The set of all real numbers is denoted by \( \mathbb{R} \) and its subset of non-negative and strictly positive numbers are denoted by \( \mathbb{R}_+ \) and \( \mathbb{R}_{++} \), respectively. \( \mathbb{S}^n \) represents a set of symmetric matrices in \( \mathbb{R}^{n \times n} \). For any \( A, B \in \mathbb{S}^n \), \( A \preceq B \) indicates that \( A - B \) is positive semidefinite, and \( \langle A, B \rangle \) denotes their trace scalar product. Besides these blackboard letters, we use capital caligraphic letters to denote other sets (including cones). If \( K \) is a cone, then \( K^* \) denotes its dual. For any vector \( \mathbf{v} \in \mathbb{R}^n \), \( \text{diag}(\mathbf{v}) \) is a diagonal matrix in \( \mathbb{S}^n \) that has elements of \( \mathbf{v} \) sitting on its main diagonal, and \( \#(\mathbf{v}, m) \) counts the number of occurrences of \( m \) in \( \mathbf{v} \), and \( \mathbf{v}^m \), \( 1 \leq m \leq n \), is a subvector of \( \mathbf{v} \) containing its first \( m \) elements, i.e., \( \mathbf{v}^m = (v_1, \ldots, v_m)^\top \). Throughout, a division by zero is allowed and is defined as

\[
\frac{a}{0} = \begin{cases} +\infty & \text{if } a > 0, \\ 0 & \text{if } a = 0, \\ -\infty & \text{if } a < 0. 
\end{cases}
\]

### 2. Mean-variance inventory model and its finite representations

We consider an uncapacitated inventory system that stores a single product for sale, and we denote by \( c, h \) and \( b \) the unit ordering, holding and backlogging costs, respectively. We assume that these cost parameters are positive and that they are time-invariant, although these are assumptions that could readily be lifted. Supposing that the inventory manager is risk-neutral, that the demand distribution is perfectly known, and that any unfulfilled order could be indefinitely backlogged, facing the uncertain demands \( \xi = (\xi_1, \ldots, \xi_T)^\top \), the inventory manager would be interested in minimizing the total expected cost, i.e., solving

\[
\begin{align*}
\text{minimize} \quad & \mathbb{E}_P \left[ \sum_{t=1}^{T} (cx_t + \max \{h_{y_t}(\xi^t), -b_{y_t}(\xi^t)\}) \right] \\
\text{subject to} \quad & \mathbf{x} \in \mathcal{X}, \quad y_t : \mathbb{R}_+^t \mapsto \mathbb{R} \quad \forall t \in \{1, \ldots, T\} \\
& y_t(\xi^t) = y_{t-1}(\xi^{t-1}) + x_t - \xi_t \quad \mathbb{P}\text{-a.s.} \quad \forall t \in \{1, \ldots, T\},
\end{align*}
\]

(1)

where \( \mathbf{x} = (x_1, \ldots, x_T)^\top \) and \( \mathbf{y} = (y_1, \ldots, y_T)^\top \) collect all decision variables representing the order quantities and the end-of-period inventory levels, respectively, and \( y_0 \in \mathbb{R} \) denotes the initial inventory level, which is given. For the ease of exposition, we assume that all orders are instantaneous, i.e., the lead times are zero. Besides, we follow Bertsimas and Thiele (2006) and Mamani et al. (2017) in imposing that \( \mathbf{x} \) is chosen here-and-now and the feasible set \( \mathcal{X} \subseteq \mathbb{R}_+^T \) is uncertainty-free. Unless stated otherwise, \( \mathcal{X} = \mathbb{R}_+^T \).
In the actual reality, not only $\xi$ is uncertain but its probability distribution $\mathbb{P}$, which needs to be empirically estimated, is ambiguous. Therefore, to hedge against such an estimation risk, the inventory manager may choose to robustify Problem (1) and solve

$$\begin{align*}
\text{minimize} \quad & \max_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[ \sum_{t=1}^{T} \left( cx_t + \max \{ hy_t(\xi_t), -by_t(\xi_t) \} \right) \right] \\
\text{subject to} \quad & x \in \mathcal{X}, \quad y_t : \mathbb{R}^I \to \mathbb{R} \quad \forall t \in \{1, \ldots, T\} \\
\quad & y_t(\xi_t) = y_{t-1}(\xi_t^{t-1}) + x_t - \xi_t \quad \forall \xi \in \mathbb{R}^T, \quad \forall t \in \{1, \ldots, T\}.
\end{align*}$$

(2)

Note that in both Problems (1) and (2), the decision variables $y_t(\xi_t)$ could be easily eliminated by replacing each $y_t(\xi_t)$ with $y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau)$. Therefore, although $x$ is decided here-and-now, the per-period holding and backlogging costs depend on the realization of past demands. In contrast, Mamani et al. (2017) ignored such randomness in exchange for a closed-form solution.

We henceforth express Problem (2) compactly as

$$\begin{align*}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & x \in \mathcal{X},
\end{align*}$$

where the objective function $f$ characterizes the worst-case expected cost incurred from $x$, i.e.,

$$f(x) = \max_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[ \sum_{t=1}^{T} \left( cx_t + \max \left\{ h \left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right) \right) \right) \right].$$

(3)

We note that $f$ is convex but not necessarily easy to evaluate because it involves solving an infinite optimization problem over the probability distribution $\mathbb{P} \in \mathcal{P}$. Following the seminal work of Scarf (1958), we choose to work with the following mean-variance ambiguity set:

$$\mathcal{P} = \{ \mathbb{P} \in \mathcal{M}_+^{\mathbb{R}^T} : \mathbb{P}(\xi \in \mathbb{R}^T) = 1, \mathbb{E}_{\mathbb{P}}[\xi_t] = \mu, \mathbb{E}_{\mathbb{P}}[\xi_t^2] = \mu^2 + \sigma^2 \quad \forall t = 1, \ldots, T \},$$

(4)

where $\mathcal{M}_+$ denotes the cone of all non-negative measures supported on the input set, because these two summary statistics can be inferred relatively easily from empirical data. This choice of $\mathcal{P}$ is equivalent to us assuming that the demands $\xi$ have a stationary mean $\mu > 0$ and a stationary variance $\sigma^2 > 0$ but nothing else is known. We remark that this basic ambiguity set allows demands to be negative to capture possible product returns. When returns are disallowed (perhaps because of a hygienic reason or the product’s short shelf life), we will consider a restricted ambiguity set in Section 5.

Under this $\mathcal{P}$, we will first show that the worst-case expected total cost $f(x)$, with $x$ fixed, can be determined by solving a finite-optimization problem or its dual. Before presenting these results, we introduce a cone $\mathcal{K}_t$, $t \in \{1, \ldots, T\}$, that is instrumental to our subsequent analyses.

$$\mathcal{K}_t = \left\{ (\alpha, \beta, \gamma) \in \mathbb{R}^+_+ \times \mathbb{R}^+ \times \mathbb{R}^+_+ : 4\alpha \geq \sum_{\tau=1}^{t} \frac{\beta^2}{\gamma^2} \right\}.$$
Note that $\mathcal{K}_t$ is a proper cone, that is, it is closed, solid, pointed and convex. Specifically, we can express $\mathcal{K}_t$ as an intersection of multiple second-order cones, i.e.,

$$\mathcal{K}_t = \left\{ (\alpha, \beta, \gamma) \in \mathbb{R}_+ \times \mathbb{R}^t \times \mathbb{R}^t : \exists \theta \in \mathbb{R}_+, \alpha \geq \sum_{\tau=1}^t \theta_\tau, \| (\beta_\tau, \theta_\tau - \gamma_\tau) \|^2 \leq \theta_\tau + \gamma_\tau \ \forall \tau = 1, \ldots, t \right\}.$$ 

Moreover, we also introduce a discrete set $\mathcal{E} = \{ h, -b \}^T$ and note that its cardinality is $2^T$. We will use it to capture the $2^T$ underlying linear lower bounds of the total holding and backlogging cost, which appears as a part of the objective function of Problems (1) and (2), i.e.,

$$\sum_{t=1}^T \max \{ hy_0(\xi^t), -by_0(\xi^t) \} = \max_{e \in \mathcal{E}} \sum_{t=1}^T e_t \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right).$$

This exponential complexity arises as a result of that, at the end of each period $t \in \{1, \ldots, T\}$, the inventory manager will have to pay for either the holding cost if the end-of-period inventory level is positive or the backlogging cost otherwise.

**Theorem 1.** We have that

$$f(x) = \minimize \quad c^T x + \alpha^T \beta + (\mu^2 + \sigma^2)^T \gamma$$

subject to $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^T$, $\gamma \in \mathbb{R}^T$

$$(\alpha(x, e), \beta(e), \gamma) \succeq_{\mathcal{K}_T} 0 \ \forall e \in \mathcal{E},$$

where $\alpha(x, e) = \alpha - y_0 1^T e - \sum_{t=1}^T x_t \sum_{\tau=t}^T e_\tau$, $\beta_t(e) = \beta_t + \sum_{\tau=t}^T e_\tau$, for all $x \in \mathbb{R}^T$.

We can interpret Problem (5) as a robust conic program with an auxiliary uncertain vector $e$ that belongs to a discrete uncertainty set $\mathcal{E}$.

**Proof of Theorem 1.** Starting from its definition in (3), we may characterize the function $f$ as the optimal objective value of a generalized moment problem, i.e.,

$$f(x) = \max \int_{\mathbb{R}^T} \frac{1}{\mathbb{P}(d\xi)} \sum_{t=1}^T e_t \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \mathbb{P}(d\xi)$$

subject to $\mathbb{P} \in \mathcal{M}_+(\mathbb{R}^T)$

$$\int_{\mathbb{R}^T} \mathbb{P}(d\xi) = 1$$

$$\mathbb{E}[\xi_t] \mathbb{P}(d\xi) = \mu \ \forall t = 1, \ldots, T$$

$$\mathbb{E}[\xi_t^2] \mathbb{P}(d\xi) = \mu^2 + \sigma^2 \ \forall t = 1, \ldots, T.$$ 

By assigning the dual variables $\alpha$ to the normalization constraint, each $\beta_t$ to each of the first-order moment constraints, and each $\gamma_t$ to each of the second-order moment constraints and invoking
strong duality due to Shapiro (2001, Proposition 3.4), we find a minimization problem that attains the same optimal objective value as the above maximization problem:

$$\min \alpha + \mu_1 \gamma + (\mu^2 + \sigma^2) \gamma$$

subject to:

$$\alpha \in \mathbb{R}, \beta \in \mathbb{R}^T, \gamma \in \mathbb{R}^T$$

$$\alpha + \sum_{t=1}^T \beta_t \xi_t + \sum_{t=1}^T \gamma_t \xi_t^2 \geq \sum_{t=1}^T e_t \left( y_0 + \sum_{\tau=1}^t (x_{\tau} - \xi_{\tau}) \right) \quad \forall e \in \mathcal{E} \quad \forall \xi \in \mathbb{R}^T.$$

Note that the inequality constraint in the above minimization problem could be rearranged as

$$\left[ \alpha - y_0 \mathbf{1}^T \mathbf{e} - \sum_{t=1}^T e_t \sum_{\tau=1}^t x_{\tau} \right] + \left[ \sum_{t=1}^T \beta_t \xi_t + \sum_{t=1}^T e_t \sum_{\tau=1}^t \xi_{\tau} \right] + \sum_{t=1}^T \gamma_t \xi_t^2 \geq 0 \quad \forall e \in \mathcal{E} \quad \forall \xi \in \mathbb{R}^T$$

$$\iff \alpha(x, e) + \sum_{t=1}^T \beta_t(e) \xi_t + \sum_{t=1}^T \gamma_t \xi_t^2 \geq 0 \quad \forall e \in \mathcal{E} \quad \forall \xi \in \mathbb{R}^T.$$

For any fixed $e \in \mathcal{E}$, as the above quadratic inequality holds for every $\xi \in \mathbb{R}^T$, it follows that $\gamma_t$ must be non-negative. Furthermore, if $\gamma_t$ vanishes, then so does $\beta_t(e)$. Denoting by $\mathcal{T}$ the index set $\{t \in \{1, \ldots, T\} : \gamma_t > 0\}$, we can re-express the considered quadratic inequality as

$$\alpha(x, e) + \sum_{t \in \mathcal{T}} \beta_t(e) \xi_t + \sum_{t \in \mathcal{T}} \gamma_t \xi_t^2 \geq 0 \quad \forall \xi \in \mathbb{R}^T$$

$$\iff \alpha(x, e) + \sum_{t \in \mathcal{T}} \gamma_t \left[ \left( \xi_t + \frac{\beta_t(e)}{2 \gamma_t} \right)^2 - \frac{\beta_t(e)^2}{4 \gamma_t^2} \right] \geq 0 \quad \forall \xi \in \mathbb{R}^T$$

$$\iff 4 \alpha(x, e) \geq \sum_{t \in \mathcal{T}} \beta_t(e)^2 \gamma_t^2 \quad \forall \xi \in \mathbb{R}^T,$$

and the proof is completed by recalling the definition of the cone $K_T$.

**Remark 1.** As a by-product of the proof of Theorem 1, $\gamma \geq 0$ is imposed as an explicit constraint in Problem (5). As $\gamma_t$, $t \in \{1, \ldots, T\}$, is a dual variable of the constraint $\mathbb{E}_\mathcal{E}[\xi_t^2] = \mu^2 + \sigma^2$, the optimal objective value of Problem (5) coincides with the worst-case expected total cost across all probability distributions of the demand from the relaxed ambiguity set

$$\mathcal{P}' = \{ \mathbb{P} \in \mathcal{M}_+(\mathbb{R}^T) : \mathbb{P}(\xi \in \mathbb{R}^T) = 1, \mathbb{E}_\mathcal{P}[\xi_t] = \mu, \mathbb{E}_\mathcal{P}[\xi_t^2] \leq \mu^2 + \sigma^2 \quad \forall t = 1, \ldots, T \} \supset \mathcal{P}.$$

In other words, the worst-case distribution achieves maximum variances. This observation is consistent with various moment-based distributionally robust optimization problems; see e.g. Delage and Ye (2010) and Rujeerapaiboon et al. (2016).

**Remark 2.** The proof of Theorem 1 reveals that, for any fixed $y_0 \in \mathbb{R}$ and $x \in \mathcal{X}$ at optimality,

$$\alpha = \max_{e \in \mathcal{E}} \left\{ \max_{\xi} \left\{ y_0 \mathbf{1}^T \mathbf{e} + \sum_{t=1}^T e_t \sum_{\tau=1}^t \mathbf{x}_{\tau} - \sum_{t=1}^T \beta_t \xi_t - \sum_{t=1}^T e_t \sum_{\tau=1}^t \xi_{\tau} - \sum_{t=1}^T \gamma_t \xi_t^2 \right\} \right\}, \tag{7}$$
where the optimization problem over \( e \in \mathcal{E} \) on the right-hand side constitutes a convex maximization problem. Mak et al. (2015) and Padmanabhan et al. (2021) recently encountered a similar optimization problem when analyzing a robust appointment scheduling problem, where they have

\[
\alpha = \max_{e \in \mathcal{E}} \left\{ \sum_{t=1}^{T} \max_{\xi_t} g_t(e_t, \xi_t, \beta_t, \gamma_t) \right\} = \max_{e \in \text{conv}(\mathcal{E})} \left\{ \sum_{t=1}^{T} \max_{\xi_t} g_t(e_t, \xi_t, \beta_t, \gamma_t) \right\}
\]

for suitably defined functions \( g_1, \ldots, g_T \) that are convex in \( e \). Subsequently, they express (8) as a mixed-integer binary program with a totally unimodular constraint matrix so that the binarity requirement can be lifted. We cannot however directly express (7) as an instance of (8) because of the term \( \sum_{t=1}^{T} e_t \sum_{\tau=1}^{T} \xi_{\tau} = \sum_{t=1}^{T} \xi_{t} \sum_{\tau=1}^{T} e_{\tau} \). Although we can use a variable transformation \( \eta_t = \sum_{t=1}^{T} e_t, t \in \{1, \ldots, T\} \), and \( \eta_{T+1} = 0 \) and replace each \( e_t \) with \( \eta_t - \eta_{t+1} \) so that

\[
\alpha = \max_{\eta} \left\{ \sum_{t=1}^{T} \max_{\xi_t} \tilde{g}_t(\eta_t, \xi_t, \beta_t, \gamma_t) : -b \leq \eta_t - \eta_{t+1} \leq +h, \ \eta_{T+1} = 0 \right\}
\]

for suitable functions \( \tilde{g}_1, \ldots, \tilde{g}_T \) that are convex in \( e \), such a transformation appears to prevent us from following the steps in Mak et al. (2015) to obtain a polynomial-sized exact reformulation of Problem (5).

In the next step, we will describe another way of computing \( f(x) \) based on the dual reformulation of Problem (5). As one would expect, there is certain semblance between Problem (6) and the optimization problem in Theorem 2 below. Subsequently, an analysis of this latter formulation, which is still exact, will be provided. This will eventually lead us to the characterization of the worst-case probability distribution within the ambiguity set \( \mathcal{P} \).

**Theorem 2.** We have that

\[
f(x) = \text{maximize } c^{\top} x + \sum_{e \in \mathcal{E} : \pi(e) \neq 0} \alpha(e) \sum_{t=1}^{T} e_t \left( y_0 + \sum_{\tau=1}^{t} \left( x_\tau - \frac{\beta_\tau(e)}{\alpha(e)} \right) \right)
\]

subject to \( \pi : \mathcal{E} \mapsto \mathbb{R}, \ \beta : \mathcal{E} \mapsto \mathbb{R}^T, \ \gamma : \mathcal{E} \mapsto \mathbb{R}^T \)

\[
\langle \pi(e), \beta(e), \gamma(e) \rangle \geq K_\pi, 0 \quad \forall e \in \mathcal{E}
\]

\[
\sum_{e \in \mathcal{E}} \pi(e) = 1 \quad \sum_{e \in \mathcal{E}} \beta(e) = \mu \mathbf{1} \quad \sum_{e \in \mathcal{E}} \gamma(e) = (\mu^2 + \sigma^2) \mathbf{1},
\]

for all \( x \in \mathbb{R}_+^T \).

We note that Problems (5) and (9) are similarly intractable because they contains an exponential number of constraints and decision variables, respectively, as \( \mathcal{E} \) is exponentially-sized.
Proof of Theorem 2. Leveraging the exact characterization of the worst-case expected cost $f(x)$ from Theorem 1, we assign a set of dual variables $(\bar{\pi}(e), \bar{\beta}(e), \gamma(e)) \in K_T^*$ to the conic constraint in Problem (5) that is associated with each $e \in \mathcal{E}$:

$$((\alpha(x,e), \beta(e), \gamma) \geq_{K_T} 0$$

$$\iff \left( \alpha - y_0 1^T e - \sum_{t=1}^T x_t \sum_{\tau=t}^T e_{\tau}, \beta_1 + \sum_{t=1}^T \sum_{\tau=t}^T e_{\tau}, \beta_2 + \sum_{t=2}^T \sum_{\tau=t}^T e_{\tau}, \ldots, \beta_T + e_{T, \gamma_1, \gamma_2, \ldots, \gamma_T} \right) \geq_{K_T} 0,$$

and obtain the following dual formulation of Problem (5):

$$f(x) = \max_{x \in \mathcal{X}} c1^T x + \sum_{e \in \mathcal{E}} \alpha(e) \left( y_0 1^T e + \sum_{t=1}^T x_t \sum_{\tau=t}^T e_{\tau} \right) - \sum_{e \in \mathcal{E}} \sum_{t=1}^T \beta_1(e) \sum_{\tau=t}^T e_{\tau}$$

subject to $\bar{\alpha} : \mathcal{E} \mapsto \mathbb{R}$, $\bar{\beta} : \mathcal{E} \mapsto \mathbb{R}^T$, $\gamma : \mathcal{E} \mapsto \mathbb{R}^T$

$$(\bar{\alpha}(e), \bar{\beta}(e), \gamma(e)) \geq_{K_T} 0 \quad \forall e \in \mathcal{E}$$

$$\sum_{e \in \mathcal{E}} \alpha(e) = 1$$

$$\sum_{e \in \mathcal{E}} \beta(e) = \mu 1$$

$$\sum_{e \in \mathcal{E}} \gamma(e) = (\mu^2 + \sigma^2) 1.$$ 

By noting that the objective function of the above optimization problem, when evaluated at any feasible $(\bar{\alpha}, \bar{\beta}, \gamma)$, could be re-expressed as

$$c1^T x + \sum_{e \in \mathcal{E}} \alpha(e) \left( y_0 1^T e + \sum_{t=1}^T \left( x_t - \frac{\beta_1(e)}{\alpha(e)} \right) \sum_{\tau=t}^T e_{\tau} \right) - \sum_{e \in \mathcal{E}} \sum_{t=1}^T \beta_1(e) \sum_{\tau=t}^T e_{\tau}$$

$$= c1^T x + \sum_{e \in \mathcal{E}} \alpha(e) \left( y_0 1^T e + \sum_{t=1}^T e_t \sum_{\tau=t}^T \left( x_t - \frac{\beta_1(e)}{\alpha(e)} \right) \right)$$

$$= c1^T x + \sum_{e \in \mathcal{E}} \alpha(e) \left( \sum_{t=1}^T e_t \left( y_0 + \sum_{\tau=t}^T \left( x_t - \frac{\beta_1(e)}{\alpha(e)} \right) \right) \right),$$

where the first equality holds because $\bar{\beta}(e) = 0$ whenever $\bar{\alpha}(e) = 0$ (see Proposition 1 just below), the proof is completed. \hfill \Box

The characterization of Problem (9) involves the cone $K_T^*$ that is dual to $K_T$. In order to solve this problem, we thus need the exact description of the dual cone, which is given next.

Proposition 1. We have that

$$K_T^* = \left\{ \left( \bar{\alpha}, \bar{\beta}, \gamma \right) \in \mathbb{R}_+ \times \mathbb{R}^t \times \mathbb{R}_+^t : \gamma \geq \max_{\tau \in \{1, \ldots, t\}} \frac{\beta_\tau}{\gamma_\tau} \right\}.$$ (10)
Proof. We denote the set given on the right-hand side of (10) by \( K_t^i \). So as to show that \( K_t^* = K_t^i \), it suffices to prove that \( K_t^i \subseteq K_t^* \) (step 1) and \( K_t^* \subseteq K_t^i \) (step 2).

For the first part, we consider an arbitrary \((\alpha, \beta, \gamma) \in K_t\) and \((\bar{\alpha}, \bar{\beta}, \bar{\gamma}) \in K_t^i\). It follows that
\[
\alpha \bar{\alpha} + \gamma \bar{\gamma} \geq \frac{1}{4} \sum_{t=1}^{t} \frac{\beta_t^2}{\gamma_t} + \frac{t}{t} \gamma_t \bar{\gamma_t} \geq \sum_{t=1}^{t} \left( \frac{\beta_t^2}{4 \gamma_t \bar{\gamma_t}} + \gamma_t \bar{\gamma_t} \right) \geq - \frac{t}{t} \beta_t \bar{\beta_t} = - \beta^\top \bar{\beta},
\]
where the last inequality holds because \( \frac{\beta_t^2}{4 \gamma_t \bar{\gamma_t}} + \gamma_t \bar{\gamma_t} \geq |\beta_t \bar{\beta_t}|, \forall \tau \). Indeed, due to the relationship between arithmetic and geometric means this latter inequality holds whenever \( \gamma_t \) and \( \bar{\gamma_t} \) are both strictly positive. If \( \gamma_t = 0 \) then \( \beta_t = 0 \), and if \( \bar{\gamma_t} = 0 \) then \( \bar{\beta_t} = 0 \). In both cases, the same inequality is still valid. The above derivation implies that \( K_t^i \) contains \( K_t^* \) inside. Conversely, consider an arbitrary \((\bar{\alpha}, \bar{\beta}, \bar{\gamma}) \in K_t^*\). It holds that
\[
\alpha \bar{\alpha} + \beta^\top \bar{\beta} + \gamma \bar{\gamma} \geq 0 \quad \forall (\alpha, \beta, \gamma) \in K_t.
\]
As \((1,0,0) \in K_t\), it necessarily follows from the above inequality that \( \bar{\alpha} \geq 0 \). Similarly as \((0,0,1) \in K_t, \forall \tau \in \{1, \ldots, t\} \), it necessarily follows that \( \bar{\gamma} \geq 0 \). Moreover, for any fixed \( \tau \in \{1, \ldots, t\} \), if \( \bar{\beta} \neq 0 \), then \( \bar{\gamma} > 0 \). Suppose otherwise for the sake of a contradiction that there exists \((\bar{\alpha}, \bar{\beta}, \bar{\gamma}) \in K_t^*\) such that \( \bar{\beta} \neq 0 \) and \( \bar{\gamma} = 0 \), then as
\[
\left( \alpha, \bar{\beta}, 1, \frac{\beta^2}{4 \alpha} \right) \in K_t \quad \forall \alpha > 0,
\]
it must follow that \( \alpha \bar{\alpha} - \beta^2 \geq 0 \) for any \( \alpha > 0 \). This resultant inequality however cannot hold true because \( \alpha \) could be arbitrarily close to zero. Assuming now that \( \bar{\beta} \neq 0 \) and \( \bar{\gamma} > 0 \), we hence find
\[
\left( \bar{\alpha}, \bar{\alpha}, \bar{\beta}, \frac{\bar{\beta}^2}{4 \bar{\alpha}} \right) \in K_t,
\]
and consequently,
\[
\bar{\alpha} \bar{\beta}^2 - \bar{\beta} \frac{2 \bar{\beta}^2}{\bar{\alpha}} + \bar{\gamma} \geq 0 \quad \implies \quad \bar{\alpha} \geq \frac{\bar{\beta}^2}{\bar{\gamma}}.
\]
As \( \bar{\alpha} \geq 0 \), the same conclusion can also be reached even if \( \bar{\beta} = 0 \). Since the above inequality holds for any \( \tau \in \{1, \ldots, t\} \), we may conclude that \( K_t^* \) is contained in \( K_t^i \). Combining both halves of the argument yields the desired result.

We close this section by visualizing \( K_2 \) and \( K_2^* \) in Figure 1. The three axes of Figure 1 (left) represent the value of \( \beta_1, \beta_2 \) and (the minimum) \( \gamma_1 + \gamma_2 \), respectively, when \( \alpha \) is set to one. Algebraically, this surface is obtained from the following observation that
\[
\gamma_1 + \gamma_2 \geq (|\beta_1| + |\beta_2|)^2 \left( \frac{\beta_1^2}{\gamma_1} + \frac{\beta_2^2}{\gamma_2} \right)^{-1} \geq \frac{1}{4} (|\beta_1| + |\beta_2|)^2
\]
for any \((\beta, \gamma) \) such that \( (1, \beta, \gamma) \in K_2 \). Analogously, the three axes of Figure 1 (right) represent the value of \( \bar{\beta}_1, \bar{\beta}_2 \) and (the minimum) \( \bar{\gamma}_1 + \bar{\gamma}_2 \), respectively, when \( \bar{\alpha} \) is set to one. The surface shown can be obtained by noting that \( \bar{\gamma}_1 + \bar{\gamma}_2 \geq \bar{\beta}_1^2 + \bar{\beta}_2^2 \) for any \((\bar{\beta}, \bar{\gamma})\) such that \((1, \bar{\beta}, \bar{\gamma}) \in K_2^* \).
2.1. Worst-case demand distribution

This section is devoted to the determination of a probability distribution from the ambiguity set \( P \) that solves the maximization problem on the right-hand side of (3) for any fixed inventory decision \( x \in \mathcal{X} \). The construction of the worst-case distribution is often used in a stress test (Bertsimas et al. 2010), and the main ingredients are the dual formulation derived in Theorem 2 and the following technical lemmas which are pertain to the optimal values of the decision variables of Problem (9).

**Lemma 1.** There exists an optimal solution \((\tilde{\alpha}^\star, \tilde{\beta}^\star, \tilde{\gamma}^\star)\) of Problem (9) such that \( \tilde{\beta}^\star(e) = \tilde{\gamma}^\star(e) = 0, \forall e \in \mathcal{E} : \tilde{\alpha}^\star(e) = 0. \)

**Proof.** Denote by \((\bar{\alpha}^\dagger, \bar{\beta}^\dagger, \bar{\gamma}^\dagger)\) an arbitrary optimal solution of Problem (9) and by \( \mathcal{E}^0 \) a subset of \( \mathcal{E} \) which contains all scenarios \( e \) such that \( \bar{\alpha}^\dagger(e) = 0 \). Note that \( \mathcal{E}^0 \) must be strictly contained in \( \mathcal{E} \) and there must exist a scenario \( e^\dagger \in \mathcal{E} \setminus \mathcal{E}^0 \) such that \( \bar{\alpha}^\dagger(e^\dagger) \) is strictly positive. Then, we construct a new solution \((\tilde{\alpha}^\star, \tilde{\beta}^\star, \tilde{\gamma}^\star)\) with \( \tilde{\alpha}^\star = \bar{\alpha}^\dagger, \tilde{\beta}^\star = \bar{\beta}^\dagger \), and

\[
\tilde{\gamma}^\star(e) = \begin{cases} 
\bar{\gamma}^\dagger(e) + \sum_{e \in \mathcal{E}^0} \bar{\gamma}^\dagger(e') & \text{if } e = e^\dagger, \\
0 & \text{if } e \in \mathcal{E}^0, \\
\bar{\gamma}^\dagger(e) & \text{if } e \in \mathcal{E} \setminus (\mathcal{E}^0 \cup \{e^\dagger\}).
\end{cases}
\]

It is readily seen that \( \sum_{e \in \mathcal{E}} \tilde{\gamma}^\star(e) = \sum_{e \in \mathcal{E}} \bar{\gamma}^\dagger(e) = (\mu^2 + \sigma^2)1, (\tilde{\alpha}^\star(e^\dagger), \tilde{\beta}^\star(e^\dagger), \tilde{\gamma}^\star(e^\dagger)) \in \mathcal{K}^2 \) and that all other constraints remain satisfied by this newly constructed solution. Since \( \tilde{\gamma} \) does not appear in the objective function of Problem (9), we may thus conclude that \((\tilde{\alpha}^\star, \tilde{\beta}^\star, \tilde{\gamma}^\star)\) is feasible and optimal. The proof is then completed by noting that \( \tilde{\beta}^\star(e) \) automatically vanishes whenever \( \tilde{\alpha}^\star(e) \) does. \( \square \)

**Lemma 2.** Any optimal solution of \((\tilde{\alpha}^\star, \tilde{\beta}^\star, \tilde{\gamma}^\star)\) of Problem (9) satisfies

\[
\eta_t \left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\tilde{\beta}^\star_{\alpha}(e)}{\tilde{\alpha}^\star_{\alpha}(e)} \right) \right) = \max \left\{ h \left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\tilde{\beta}^\star_{\alpha}(e)}{\tilde{\alpha}^\star_{\alpha}(e)} \right) \right), -b \left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\tilde{\beta}^\star_{\alpha}(e)}{\tilde{\alpha}^\star_{\alpha}(e)} \right) \right) \right\}
\]

for any \( t \in \{1, \ldots, T\} \) and any \( e \in \mathcal{E} \) such that \( \tilde{\alpha}^\star(e) > 0 \).
For any fixed $\mathbf{e} \in \mathcal{E}$, where $\overline{\alpha}^*(\mathbf{e}) > 0$, if $\xi_\tau = \frac{\overline{\beta}^*_\tau(\mathbf{e})}{\overline{\alpha}(\mathbf{e})}$, $\forall \tau$, then the inventory level at the end of period $t$ is given by $y_0 + \sum_{\tau=1}^{t} \left( x_{\tau} - \frac{\overline{\beta}_\tau^*(\mathbf{e})}{\overline{\alpha}(\mathbf{e})} \right)$, and the right-hand side of the above equation characterizes either the holding or the backlogging cost that would incur. Lemma 2 therefore provides a simpler expression of such a cost. This lemma is particularly useful as the worst-case distribution, which we will later construct, is a mixture distribution comprising of components whose mean demands at period $\tau$, $\forall \tau$, are equal to $\frac{\overline{\beta}_\tau^*(\mathbf{e})}{\overline{\alpha}(\mathbf{e})}$.

Proof of Lemma 2. Consider any scenario $\mathbf{e} \in \mathcal{E}$ such that $\overline{\alpha}^*(\mathbf{e}) > 0$. To simplify the exposition, we abbreviate (and slightly abuse the notation) the dual optimal solution $\overline{\alpha}^*(\mathbf{e}), \overline{\beta}^*(\mathbf{e}), \overline{\gamma}^*(\mathbf{e})$ by $\overline{\alpha}^*, \overline{\beta}^*, \overline{\gamma}^*$, respectively. Besides, we denote by $(\alpha^*, \beta^*, \gamma^*)$ an optimal solution of the primal problem (5). The KKT optimality condition of this primal-dual pair includes a complementary slackness condition which necessarily holds and reads

$$
\overline{\alpha}^* \left( \alpha^* - y_0 \mathbf{1}^T - \sum_{t=1}^{T} x_t \sum_{\tau=t}^{T} e_{\tau} \right) + \sum_{t=1}^{T} \overline{\beta}_t^* \left( \beta_t^* + \sum_{\tau=t}^{T} e_{\tau} \right) + \sum_{t=1}^{T} \overline{\gamma}_t^* = 0
$$

$$
\iff \overline{\alpha}^* \left( y_0 \mathbf{1}^T + \sum_{t=1}^{T} x_t \sum_{\tau=t}^{T} e_{\tau} \right) - \sum_{t=1}^{T} \overline{\beta}_t^* \sum_{\tau=t}^{T} e_{\tau} = \alpha^* \overline{\alpha}^* + (\beta^*)^T \overline{\beta}^* + (\gamma^*)^T \overline{\gamma}^*
$$

$$
\iff \overline{\alpha}^* \sum_{t=1}^{T} x_t \left( y_0 + \sum_{\tau=t}^{T} \left( x_{\tau} - \frac{\overline{\beta}_\tau^*}{\alpha^*} \right) \right) = \alpha^* \overline{\alpha}^* + (\beta^*)^T \overline{\beta}^* + (\gamma^*)^T \overline{\gamma}^*.
$$

For any time period $t' \in \{1, \ldots, T\}$, it thus follows that

$$
\overline{\alpha}^* e_{t'} \left( y_0 + \sum_{\tau=1}^{t'} \left( x_{\tau} - \frac{\overline{\beta}_\tau^*}{\alpha^*} \right) \right)
$$

$$
= \alpha^* \overline{\alpha}^* + (\beta^*)^T \overline{\beta}^* + (\gamma^*)^T \overline{\gamma}^* - \alpha^* \sum_{t \neq t'} e_{t'} \left( y_0 + \sum_{\tau=t'}^{T} \left( x_{\tau} - \frac{\overline{\beta}_\tau^*}{\alpha^*} \right) \right)
$$

$$
= \alpha^* \left( \alpha^* - y_0 \sum_{t \neq t'} e_{t'} \sum_{\tau=t}^{T} x_{\tau} - \sum_{\tau=t' \neq \tau}^{T} e_{\tau} \right) + \sum_{t=1}^{T} \beta_t^* \left( \beta_t^* + \sum_{\tau=t}^{T} e_{\tau} \right) + \sum_{t=1}^{T} \gamma_t^*.
$$

Next, we collect the coefficients of the dual solution on the right-hand side of the above equation

$$
\zeta = \left( \alpha^* - y_0 \sum_{t \neq t'} e_{t'} - \sum_{t=1}^{T} x_t \sum_{\tau=t}^{T} e_{\tau} \beta_\tau^* + \sum_{\tau=t}^{T} e_{\tau} \beta_\tau^* + \sum_{t=1}^{T} e_{t} \gamma_t^* \right)^T
$$

and observe that $\zeta$ can be expressed as a convex combination of the following two vectors

$$
\overline{\zeta} = \left( \alpha^* - y_0 \mathbf{1}^T e - \sum_{t=1}^{T} x_t \sum_{\tau=t}^{T} e_{\tau} \beta_\tau^* + \sum_{\tau=t}^{T} e_{\tau} \beta_\tau^* + \sum_{t=2}^{T} e_{t} \gamma_t^* \right)^T
$$

$$
\zeta = \left( \alpha^* - y_0 \mathbf{1}^T e - \sum_{t=1}^{T} x_t \sum_{\tau=t}^{T} e_{\tau} \beta_\tau^* + \sum_{\tau=t}^{T} e_{\tau} \beta_\tau^* + \sum_{t=2}^{T} e_{t} \gamma_t^* \right)^T.
$$

$^1$ That is we drop their dependency on $\mathbf{e}$.
where the scenarios \( e \) and \( \xi \) are characterized by

\[
\overline{e}_t = e_t \quad \forall t \in \{1, \ldots, T\} \setminus \{t'\} \quad \text{and} \quad \overline{e}_t' = +h \quad \text{as well as} \quad \overline{e}_\nu = -b.
\]

In particular, we can express \( \xi \) as \( \frac{b}{h+\varepsilon}\overline{\xi} + \frac{h}{h+\varepsilon}\xi \). Note that the feasibility of \((\alpha^*, \beta^*, \gamma^*)\) in view of Problem (5) implies that \( \xi, \overline{\xi} \in \mathcal{K}_T \). As \( \mathcal{K}_T \) constitutes a convex cone, we have that \( \xi \in \mathcal{K}_T \) and

\[
\varepsilon_t' \left( y_0 + \sum_{t'=1}^{t''} \left( x_{t'} - \frac{\beta_{t'}}{\alpha_{t'}} \right) \right) \geq 0
\]

because \((\overline{\alpha}^*, \overline{\beta}^*, \overline{\gamma}^*) \in \mathcal{K}_T^* \) and because of (11). The above inequality and the fact that \( \varepsilon_t' \in \{+h,-b\} \), together, finally completes the proof.

Next, we propose a candidate for the worst-case probability distribution that attains the maximum expected cost for a given inventory decision \( x \in \mathcal{X} \). In particular, we solve Problem (9) and determine an optimal solution \((\overline{\alpha}^*, \overline{\beta}^*, \overline{\gamma}^*)\) that satisfies the requirement of Lemma 1. Note that if a solver gives a solution that does not meet the requirement of Lemma 1, we can use the construction provided in the proof of Lemma 1 to determine one that does. Then, for any \( e \in E \) such that \( \overline{\alpha}^*(e) \neq 0 \), we construct a probability distribution \( \mathbb{P}^{e,\varepsilon} \in \mathcal{M}_+(\mathbb{R}^T) \), \( \varepsilon \in (0,1) \), via

\[
\mathbb{P}^{e,\varepsilon} \left( \xi = \xi^{e,\varepsilon} \right) = 1 - \varepsilon \quad \text{and} \quad \mathbb{P}^{e,\varepsilon} \left( \xi = \xi^{e,\varepsilon} \right) = \varepsilon \quad \text{for all} \quad e \in E,\quad \varepsilon \in (0,1),
\]

where the atoms \( \xi^{e,\varepsilon} \in \mathbb{R}^T \) and \( \overline{\xi}^{e,\varepsilon} \in \mathbb{R}^T \) are chosen as

\[
\xi^{e,\varepsilon}_t = \frac{\beta_t^*(e)}{\alpha_t^*(e)} - \sqrt{\frac{\varepsilon}{1 - \varepsilon}} \sqrt{\frac{\alpha_t^*(e) \gamma_t^*(e) - (\beta_t^*(e))^2}{\alpha_t^*(e)}} \quad \forall t = 1, \ldots, T,
\]

\[
\overline{\xi}^{e,\varepsilon}_t = \frac{\beta_t^*(e)}{\alpha_t^*(e)} + \sqrt{\frac{\varepsilon}{1 - \varepsilon}} \sqrt{\frac{\alpha_t^*(e) \gamma_t^*(e) - (\beta_t^*(e))^2}{\alpha_t^*(e)}} \quad \forall t = 1, \ldots, T.
\]

Note that \( \sqrt{\alpha_t^*(e) \gamma_t^*(e) - (\beta_t^*(e))^2} \) is well-defined because term inside is non-negative, which is due to the characterization of \( \mathcal{K}_T^* \) from Proposition 1. Next, we construct a mixture distribution

\[
\mathbb{P}^e = \sum_{e \in E} \overline{\alpha}^*(e) \mathbb{P}^{e,\varepsilon}
\]

and assert that this demand distribution attains the worst-case inventory cost in (3) as \( \varepsilon \downarrow 0 \).

**Theorem 3.** The demand distribution \( \mathbb{P}^e \), \( e \in (0,1) \), has the following properties:

- \( \mathbb{P}^e \in \mathcal{P} \) for all \( e \in (0,1) \),
- \( \lim_{\varepsilon \downarrow 0} \mathbb{E}_{\mathbb{P}^e} \left[ \sum_{t=1}^{T} \max \left\{ h \left( y_0 + \sum_{\tau=1}^{t'} (x_{\tau} - \xi_{\tau}) \right), -b \left( y_0 + \sum_{\tau=1}^{T} (x_{\tau} - \xi_{\tau}) \right) \right\} \right] = f(x) - \epsilon 1^T x \).
Proof. To establish that the mixture distribution $\mathbb{P}^\varepsilon$ belongs to the ambiguity set $\mathcal{P}$, we first compute from (12) the first two moments of each probability distribution component and find
\[
\mathbb{E}_{\mathbb{P}^\varepsilon} \left[ \xi_t \right] = \frac{\beta^*_t (\varepsilon)}{\alpha^*(\varepsilon)} \quad \text{and} \quad \mathbb{E}_{\mathbb{P}^\varepsilon} \left[ \xi^2_t \right] = \frac{\pi^*_t (\varepsilon)}{\alpha^*(\varepsilon)} \quad \forall t \in \{1, \ldots, T\}.
\]
Subsequently, from (13) we obtain
\[
\mathbb{E}_{\mathbb{P}^\varepsilon} \left[ \xi_t \right] = \sum_{\varepsilon \in \mathcal{E}} \beta^*_t (\varepsilon) = \sum_{\varepsilon \in \mathcal{E}} \sum_{e \in \mathcal{C}} \beta^*_t (\varepsilon) = \mu \quad \text{and} \quad \mathbb{E}_{\mathbb{P}^\varepsilon} \left[ \xi^2_t \right] = \sum_{\varepsilon \in \mathcal{E}} \pi^*_t (\varepsilon) = \sum_{\varepsilon \in \mathcal{E}} \pi^*_t (\varepsilon) = \mu^2 + \sigma^2
\]
for any $t \in \{1, \ldots, T\}$ because of Lemma 1, and we can conclude that $\mathbb{P}^\varepsilon \in \mathcal{P}$.

By construction, it similarly follows that
\[
\mathbb{E}_{\mathbb{P}^\varepsilon} \left[ \sum_{t=1}^T \max \left\{ h \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \right\} \right]
\]
\[
= (1 - \varepsilon) \sum_{t=1}^T \max \left\{ h \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi^{\varepsilon,e}_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi^{\varepsilon,e}_\tau) \right) \right\} + \varepsilon \sum_{t=1}^T \max \left\{ h \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi^{\varepsilon,e}_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi^{\varepsilon,e}_\tau) \right) \right\}.
\]
Since $\lim_{\varepsilon \to 0} \xi^{\varepsilon,e}_\tau = 0$ for all $\tau \in \{1, \ldots, T\}$, we further find
\[
\lim_{\varepsilon \to 0} \mathbb{E}_{\mathbb{P}^\varepsilon} \left[ \sum_{t=1}^T \max \left\{ h \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \right\} \right]
\]
\[
= \lim_{\varepsilon \to 0} \sum_{t=1}^T \max \left\{ h \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi^{\varepsilon,e}_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi^{\varepsilon,e}_\tau) \right) \right\}
\]
\[
= \sum_{t=1}^T \max \left\{ h \left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\beta^*_t (\varepsilon)}{\alpha^*(\varepsilon)} \right) \right), -b \left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\beta^*_t (\varepsilon)}{\alpha^*(\varepsilon)} \right) \right) \right\}
\]
\[
= \sum_{t=1}^T \epsilon_t \left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\beta^*_t (\varepsilon)}{\alpha^*(\varepsilon)} \right) \right),
\]
where the last equality is due to Lemma 2. Utilizing the fact that $\mathbb{P}^\varepsilon$ is a mixture of $\mathbb{P}^{e,e}$ and invoking Theorem 2 complete the proof. 

Under the constructed worst-case distribution, the demands $\{\xi_t\}_{t=1}^T$ appear to be highly dependent across time periods, and we ascertain this claim in Section 6. An ambiguity-averse inventory manager is therefore advised to take extra caution when it comes to stochastic independence, which is typically assumed (e.g. Scarf 1960).

3. Efficiently solvable conservative approximation

Theorems 1 and 2 ensure that we can evaluate $f(x)$ by solving a finite convex optimization problem. Nevertheless, the difficulty arises when $T$ gets large as the uncertainty set $\mathcal{E}$ is exponential-sized. In
this section, we provide two distinct ways of approximating \( f(x) \) from above. The first approximation is a result of interpreting Problem (5) as an artificial two-stage robust optimization problem and restricting the second-stage decisions using affine adaptation, whereas the second is due to constraint categorization and a series of interchanges between maximization and summation.

To this end, we first leverage Theorem 1 to provide a new characterization of \( f(x) \) as the optimal objective value of a two-stage robust optimization problem.

**Theorem 4.** We have that

\[
\begin{align*}
    f(x) = \min_{\alpha, \beta, \gamma} & \quad \alpha \mathbf{1}^T x + \alpha \mathbf{1}^T \beta + (\mu^2 + \sigma^2) \mathbf{1}^T \gamma \\
    \text{subject to} & \quad \alpha \in \mathbb{R}, \quad \beta \in \mathbb{R}^T, \quad \gamma \in \mathbb{R}^T, \quad \phi : \mathbb{R}^{2T} \to \mathbb{R}_+, \quad \psi : \mathbb{R}^{2T} \to \mathbb{R}_+ \\
    & \quad \alpha + \beta^T u + \gamma^T v \geq \mathbf{1}^T \phi(u, v) + h \mathbf{1}^T \psi(u, v) \\
    & \quad \forall (u, v) : (1, u, v) \in K_T^* \\
    & \quad y_0 + \phi_t(u, v) - \psi_t(u, v) = \sum_{\tau=1}^T (u_{\tau} - x_{\tau}) \\
    & \quad \forall (u, v) : (1, u, v) \in K_T^*, \quad \forall t = 1, \ldots, T,
\end{align*}
\]

for all \( x \in \mathbb{R}_+^T \).

In a nutshell, as soon revealed Theorem 4 is similar to Theorem 2 in that they both characterize the worst-case expected total cost \( f(x) \) by applying a duality technique to Problem (5). Observe that the dual cone \( K_T^* \) shows up in both of the theorems.

**Proof of Theorem 4.** Considering Problem (5), which characterizes \( f(x) \), we could re-express the robust constraints therein as:

\[
\min_{e \in \mathcal{E}} \max_{\omega \in \mathbb{R}} \{ \omega : (\alpha(x, e), \beta(e), \gamma) \geq K_T(\omega, 0) \} \geq 0.
\]

Then, by dualizing the maximization problem on the left-hand side of this inequality, we attain:

\[
\begin{align*}
    \min_{e \in \mathcal{E}} & \quad \min_{u, v \in \mathbb{R}^T} \left\{ \alpha(x, e) + u^T \beta(e) + v^T v : (1, u, v) \geq K_T^* 0 \right\} \geq 0 \\
    \iff & \quad \min_{u, v \in \mathbb{R}^T} \left\{ \gamma^T v + \min_{e \in \mathcal{E}} \left\{ \alpha(x, e) + u^T \beta(e) \right\} : (1, u, v) \geq K_T^* 0 \right\} \geq 0 \quad (15) \\
    \iff & \quad \min_{u, v \in \mathbb{R}^T} \left\{ \gamma^T v + \min_{e \in \text{conv}(\mathcal{E})} \left\{ \alpha(x, e) + u^T \beta(e) \right\} : (1, u, v) \geq K_T^* 0 \right\} \geq 0,
\end{align*}
\]

where the second equivalence is because \( \alpha(x, e) \) and \( \beta(e) \) are linear in \( e \) and because any solvable linear program has a vertex solution. Next, we focus on the inner minimization problem, which we could expand as

\[
\begin{align*}
    \text{minimize} & \quad \left[ \alpha - y_0 \mathbf{1}^T e - \sum_{t=1}^T x_{\tau} \sum_{\tau=1}^T e_{\tau} \right] + \left[ \beta^T u + \sum_{t=1}^T u_{\tau} \sum_{\tau=1}^T e_{\tau} \right] \\
    \text{subject to} & \quad e \geq -b \mathbf{1}, \quad e \leq h \mathbf{1}
\end{align*}
\]
thanks to definition of \( \alpha(x, e), \beta(e) \) and \( E \). Assigning the dual variables \( \phi \in \mathbb{R}_+^T \) and \( \psi \in \mathbb{R}_+^T \) to the two constraints of this linear program, respectively, and rewriting the objective function as

\[
\alpha + \beta^T u + \sum_{t=1}^T c_t \left( -y_0 + \sum_{\tau=1}^t (u_\tau - x_\tau) \right)
\]

allows us to derive the dual:

\[
\begin{align*}
\text{maximize} & \quad \alpha + \beta^T u - b1^T \phi - h1^T \psi \\
\text{subject to} & \quad \phi_t - \psi_t = -y_0 + \sum_{\tau=1}^t (u_\tau - x_\tau) \quad \forall t = 1, \ldots, T.
\end{align*}
\]

Replacing the inner minimization problem in (15) with its dual finally completes the proof.

We consider Problem (14) as a two-stage robust optimization that contains first-stage decision variables \( (\alpha, \beta, \gamma) \) as well as second-stage decision variables \( (\phi, \psi) \) whose values could depend on yet another auxiliary uncertain vector \( (u, v) \) which belongs to a projection of the dual cone \( \mathcal{K}_T \).

Although the number of constraints in Problem (14) grows linearly with \( T \), the main difficulty of solving it lies in the fact that it is a two-stage problem with wait-and-see decision variables \( \phi \) and \( \psi \). Our next result describes how to use a linear decision rule approximation (Ben-Tal et al. 2004) in conjunction with Theorem 4 to derive the following upper bound on \( f(x) \):

\[
\tilde{T}_L(x) = \min \left\{ c1^T x + \alpha + \mu 1^T \beta + (\mu^2 + \sigma^2)1^T \gamma \right\}
\]

subject to

\[
\begin{align*}
\alpha & \in \mathbb{R}, \quad \beta \in \mathbb{R}^T, \quad \gamma \in \mathbb{R}^T, \quad \kappa \in \mathbb{R}^{2T+1}, \quad (\pi_t, \pi_t^u, \pi_t^\gamma) \in \mathbb{R} \times \mathbb{R}^T \times \mathbb{R}^T \quad \forall t = 1, \ldots, T \\
(\pi_t, \pi_t^u, \pi_t^\gamma) & \geq_{\mathcal{K}_T} 0 \quad \forall t = 1, \ldots, T \\
\left( \pi_t + y_0 + \sum_{\tau=1}^t x_\tau - \kappa_{T+1}, \pi_t^u - (1, \ldots, 1, 0, \ldots, 0)^T, \pi_t^\gamma \right) & \geq_{\mathcal{K}_T} 0 \quad \forall t = 1, \ldots, T \\
\left( \alpha - hTy_0 - h \sum_{t=1}^T x_t(T-t+1) - (b+h) \sum_{t=1}^T \pi_t - \kappa_{2T+1}, \beta - (b+h) \sum_{t=1}^T \pi_t^u + h(T, T-1, \ldots, 1)^T, \gamma - (b+h) \sum_{t=1}^T \pi_t^\gamma \right) & \geq_{\mathcal{K}_T} 0,
\end{align*}
\]

where the vector \( (1, \ldots, 1, 0, \ldots, 0)^T \) in the second line of constraints has \( t \) components equal to one and \( T - t \) components equal to zero. Perhaps interestingly, unlike Problem (14) this minimization problem involves only the primal cone \( \mathcal{K}_T \). From a computational perspective, we only use the dual cone \( \mathcal{K}_T^\ast \) to construct the worst-case demand distribution; see Section 2.1.

We emphasize that the linear decision rule approximation adopted here is different than those used in the earlier papers, such as Ben-Tal et al. (2005), See and Sim (2010), and Bertsimas et al. (2019), because unlike theirs our main operational decisions \( x \in X \) are to be chosen here and now. We only apply the affine restriction to the artificial wait-and-see decisions that appear in Problem (14), which we introduce as a means to capture the worst-case expected total cost \( f(x) \).
Distributionally Robust Inventory Management with Advance Purchase Contracts

**Proposition 2 (L-conservative approximation).** We have that $f(x) \leq \mathcal{J}_L(x)$ for all $x \in \mathbb{R}_+^T$.

**Proof.** Our first step entails simplifying the exact characterization of $f(x)$ from Theorem 4 by replacing the adaptive decisions $\psi_t(u, v)$ by $y_0 + \phi_t(u, v) + \sum_{\tau=1}^t (x_\tau - u_\tau)$, $1 \leq t \leq T$. Hence,

$$f(x) = \text{minimize } c^T x + \alpha + \mu^T \beta + (\mu^2 + \sigma^2) \gamma$$

subject to $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^T$, $\gamma \in \mathbb{R}^T$, $\phi : \mathbb{R}^{2T} \rightarrow \mathbb{R}^T$

$$\phi_t(u, v) \geq 0 \quad \forall (u, v) : (1, u, v) \in \mathcal{K}_T^*, \ t = 1, \ldots, T$$

$$y_0 + \phi_t(u, v) + \sum_{\tau=1}^t (x_\tau - u_\tau) \geq 0 \quad \forall (u, v) : (1, u, v) \in \mathcal{K}_T^*, \ t = 1, \ldots, T$$

$$\alpha + \beta^T u + \gamma^T v \geq h Ty_0 + (b + h) \sum_{t=1}^T \phi_t(u, v) + h \sum_{t=1}^T \sum_{\tau=1}^t (x_\tau - u_\tau)$$

$$\forall (u, v) : (1, u, v) \in \mathcal{K}_T^*. \quad \text{max } \{ (u, v) : (1, u, v) \in \mathcal{K}_T^* \} \geq 0$$

We subsequently obtain a conservative approximation by restricting each $\phi_t(u, v)$ to an affine form: $\phi_t(u, v) = \pi_t + u^T \pi_t^u + v^T \pi_t^v$ for some linear decision rule coefficients $\pi_t \in \mathbb{R}$, $\pi_t^u \in \mathbb{R}^T$, $\pi_t^v \in \mathbb{R}^T$. As a result, we have the following upper bound of $f(x)$ that involves only here-and-now decisions.

$$\text{minimize } c^T x + \alpha + \mu^T \beta + (\mu^2 + \sigma^2) \gamma$$

subject to $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^T$, $\gamma \in \mathbb{R}^T$, $(\pi_t, \pi_t^u, \pi_t^v) \in \mathbb{R} \times \mathbb{R}^T \times \mathbb{R}^T \forall t = 1, \ldots, T$

$$\pi_t + u^T \pi_t^u + v^T \pi_t^v \geq 0 \quad \forall (u, v) : (1, u, v) \in \mathcal{K}_T^*, \ t = 1, \ldots, T$$

$$\pi_t + y_0 + \sum_{\tau=1}^t x_\tau + u^T \pi_t^u - \sum_{\tau=1}^t u_\tau + v^T \pi_t^v \geq 0 \quad \forall (u, v) : (1, u, v) \in \mathcal{K}_T^*, \ t = 1, \ldots, T$$

$$\alpha - h Ty_0 - h \sum_{t=1}^T x_t(T - t + 1) - (b + h) \sum_{t=1}^T \pi_t + u^T (\beta - (b + h) \sum_{t=1}^T \pi_t^u)$$

$$+ h \sum_{t=1}^T u_t(T - t + 1) + v^T (\gamma - (b + h) \sum_{t=1}^T \pi_t^v) \geq 0 \quad \forall (u, v) : (1, u, v) \in \mathcal{K}_T^*.$$

Leveraging the primal-dual pair involving cone $\mathcal{K}_T$ familiar from the proof of Theorem 4, we can derive the robust counterpart of the first constraint as follows.

$$\pi_t + u^T \pi_t^u + v^T \pi_t^v \geq 0 \quad \forall (u, v) : (1, u, v) \in \mathcal{K}_T^*.$$

$$\iff \min_{u, v} \{ \pi_t + u^T \pi_t^u + v^T \pi_t^v : (1, u, v) \in \mathcal{K}_T^* \} \geq 0$$

$$\iff \max \{ \kappa_t : (\pi_t - \kappa_t, \pi_t^u, \pi_t^v) \in \mathcal{K}_T \} \geq 0.$$

Applying the same routine for the remaining constraints, which are similarly linear in the conically constrained uncertain parameters $u$ and $v$, completes the proof. \qed
Next, we derive an alternative conservative approximation of \( f(x) \) which is obtained by constraint categorization and a series of interchanges between maximization and summation.

\[
\tilde{f}_Q(x) = \min \quad c^T x + \alpha + \mu_1 \beta + (\mu_2 + \sigma^2) \gamma
\]

subject to \( \alpha \in \mathbb{R}, \beta \in \mathbb{R}^T, \gamma \in \mathbb{R}^T, \lambda^0, \ldots, \lambda^T \in \mathbb{R}^T \)

\[
\alpha + by_0(T - k) - hy_0k \geq 1^T \lambda^k \quad \forall k = 0, \ldots, T
\]

\[
\left( \lambda^k - x_t \sum_{\tau = t}^T \tilde{e}^k, \beta_i(\tilde{e}^k), \gamma_t \right) \geq x_{k+1} \quad \forall k = 0, \ldots, T \quad \forall t = 1, \ldots, T
\]

\[
\left( \lambda^k - x_t \sum_{\tau = t}^T \tilde{e}^k, \beta_i(\tilde{e}^k), \gamma_t \right) \geq 0 \quad \forall k = 0, \ldots, T \quad \forall t = 1, \ldots, T
\]

where the auxiliary vectors \( \tilde{e}^k \in \mathbb{R}^T \) and \( \hat{e}^k \in \mathbb{R}^T \) are defined through

\[
\hat{e}^k = (-b, \ldots, -b + h, \ldots, +h)^T \quad \text{and} \quad \tilde{e}^k = (+h, \ldots, +h, -b, \ldots, -b)^T.
\]

The correctness of this upper bound is validated in the next theorem.

**Theorem 5 (Q-conservative approximation).** We have that \( f(x) \leq \tilde{f}_Q(x) \) for all \( x \in \mathbb{R}^T_+ \).

**Proof.** It is sufficient to show that, for any \((\alpha, \beta, \gamma, \lambda^0, \ldots, \lambda^T)\) that is feasible in Problem (16), \((\alpha, \beta, \gamma)\) is feasible in Problem (5). To achieve this, we first observe that, for each \( k \in \{0, \ldots, T\} \) and \( t \in \{1, \ldots, T\} \), the two \( \mathcal{K}_1 \)-inequalities of Problem (16) imply that

\[
4\lambda^k_t \geq \max \left\{ \frac{\beta_i(\tilde{e}^k)^2}{\gamma_t} + 4x_t \sum_{\tau = t}^T \tilde{e}^k, \frac{\beta_i(\tilde{e}^k)^2}{\gamma_t} + 4x_t \sum_{\tau = t}^T \tilde{e}^k \right\}
\]

Next, fixing the index \( k \in \{0, \ldots, T\} \), summing up the above inequality over \( t \in \{1, \ldots, T\} \) and finally using the remaining (linear) constraint of Problem (16) results in

\[
4\alpha + 4by_0(T - k) - 4hy_0k \geq \sum_{t=1}^T \max \left\{ \frac{\beta_i(\tilde{e}^k)^2}{\gamma_t} + 4x_t \sum_{\tau = t}^T \tilde{e}^k, \frac{\beta_i(\tilde{e}^k)^2}{\gamma_t} + 4x_t \sum_{\tau = t}^T \tilde{e}^k \right\}
\]

\[
= \sum_{t=1}^T \max_{e \in \mathcal{E}} \left\{ \frac{\beta_i(e)^2}{\gamma_t} + 4x_t \sum_{\tau = t}^T e, \#(e, h) = k \right\}
\]

\[
\geq \max_{e \in \mathcal{E}} \left\{ \sum_{t=1}^T \left( \frac{\beta_i(e)^2}{\gamma_t} + 4x_t \sum_{\tau = t}^T e \right), \#(e, h) = k \right\},
\]

where the equality holds because the quadratic expression \( \frac{\beta_i(e)^2}{\gamma_t} + 4x_t \sum_{\tau = t}^T e \) is convex in \( \sum_{\tau = t}^T e \) and hence its maximum value over the discrete feasible set \( \{ e \in \mathcal{E} : \#(e, h) = k \} \) is attained when \( e = \hat{e}^k \) or \( e = \tilde{e}^k \), and the last inequality follows as a result of the interchange between summation and maximization. A slight rearrangement of terms further yields

\[
\min_{e \in \mathcal{E}} \left\{ 4\alpha - 4y_01^T e - 4 \sum_{t=1}^T x_t \sum_{\tau = t}^T e - \sum_{t=1}^T \frac{\beta_i(e)^2}{\gamma_t} : \#(e, h) = k \right\} \geq 0 \quad \forall k = 0, \ldots, T
\]

\[
\iff (\alpha(x, e), \beta(e), \gamma) \geq \mathcal{K}_T \quad \forall e \in \mathcal{E} : \#(e, h) = k \quad \forall k = 0, \ldots, T
\]

\[
\iff (\alpha(x, e), \beta(e), \gamma) \geq \mathcal{K}_T \quad \forall e \in \mathcal{E}
\]

and correspondingly the feasibility of \((\alpha, \beta, \gamma)\) in Problem in (5). The proof is now completed. \( \square \)
4. Efficiently solvable progressive approximation

Previously, we have proposed two conservative approximations for \( f(x), x \in \mathbb{R}^T \), with however no way of quantifying how accurate they are. In this section, we will complement our earlier results with a progressive approximation of \( f(x) \). This approximation is based on a scenario reduction technique for a robust optimization problem, whereby we choose to remove purportedly less important scenarios from the uncertainty set. Tractability of the approximation is ensured if the number of retained scenarios grows linearly with the problem size.

Specifically, the proposed lower bound is

\[
\underline{f}(x) = \text{minimize } c^T x + \alpha + \mu 1^T \beta + (\mu^2 + \sigma^2) 1^T y \\
\text{subject to } \alpha \in \mathbb{R}, \beta \in \mathbb{R}^T, \gamma \in \mathbb{R}^T \\
(\alpha(x, \bar{e^k}), \beta(\bar{e^k}), \gamma) \succeq_k 0 \quad \forall k = 0, \ldots, T,
\]

where, for any \( x \in \mathbb{R}^T_+ \) and \( e \in \mathcal{E} \), \( \alpha(x, e) \) and \( \beta(e) \) are defined as in Theorem 1.

**Theorem 6 (Progressive approximation).** We have that \( f(x) \geq \underline{f}(x) \) for all \( x \in \mathbb{R}^T_+ \).

**Proof.** This is an immediate consequence of Theorem 1 as we can interpret \( f(x) \) as the optimal objective value of Problem (5) with the uncertain vector \( e \in \mathcal{E} \) and \( \underline{f}(x) \) as the optimal objective value of the same robust program with, however, a smaller uncertainty set \( \{\bar{e^0}, \ldots, \bar{e^T}\} \subseteq \mathcal{E} \).

To motivate the rationale behind this progressive approximation, we consider a stylised example of two-period distributionally robust uncapacitated inventory problem where the inventory is initially empty. Thanks to Theorem 1, the minimization problem \( \min_{x \geq 0} f(x) \) could be expanded to

\[
\text{minimize } c(x_1 + x_2) + \alpha + \mu (\beta_1 + \beta_2) + (\mu^2 + \sigma^2)(\gamma_1 + \gamma_2) \quad (17a) \\
\text{subject to } x_1 \in \mathbb{R}_+, \ x_2 \in \mathbb{R}_+, \ \alpha \in \mathbb{R}, \ \beta_1 \in \mathbb{R}, \ \beta_2 \in \mathbb{R}, \ \gamma_1 \in \mathbb{R}_+, \ \gamma_2 \in \mathbb{R}_+ \quad (17b) \\
\alpha - 2h x_1 - h x_2 - (\beta_1 + 2h)^2/4 \gamma_1 - (\beta_2 + h)^2/4 \gamma_2 \geq 0 \quad (17c) \\
\alpha + 2b x_1 + b x_2 - (\beta_1 - 2b)^2/4 \gamma_1 - (\beta_2 - b)^2/4 \gamma_2 \geq 0 \quad (17d) \\
\alpha + (b - h) x_1 - h x_2 - (\beta_1 + h - b)^2/4 \gamma_1 - (\beta_2 + h)^2/4 \gamma_2 \geq 0 \quad (17e) \\
\alpha + (b - h) x_1 + b x_2 - (\beta_1 + h - b)^2/4 \gamma_1 - (\beta_2 - b)^2/4 \gamma_2 \geq 0 \quad (17f)
\]

where the four constraints correspond to \( e = (h, h)^T = \bar{e}^2 \), \( e = (-b, -b)^T = \bar{e}^0 \), \( e = (-b, h)^T \) and \( e = (h, -b)^T = \bar{e}^1 \), respectively. Theorem 6 suggests that the scenario \( e = (-b, h)^T \) is least important. Indeed, we show below that the shadow price of the constraint \((17e)\) is equal to zero.

**Proposition 3.** The dual of Problem (17) is given by

\[
\text{maximize } (h + b) \sigma \left( \sqrt{4q_1q_2 + q_1q_3 + q_1q_4 + q_2q_3 + q_2q_4 + \sqrt{q_1q_2} + q_1q_4 + q_2q_3 + q_3q_4} \right) -
\]
\[
\mu (3hq_1 - 3bq_2 + (2h - b)q_3 + (h - 2b)q_4)
\]  \hspace{2cm} (18a)

subject to \ \begin{align*}
q_1, q_2, q_3, q_4 & \geq 0 \quad \text{(18b)} \\
q_1 + q_2 + q_3 + q_4 & = 1 \quad \text{(18c)} \\
c + 2hq_1 - 2bq_2 + (h - b)q_3 + (h - b)q_4 & \geq 0 \quad \text{(18d)} \\
c + hq_1 - bq_2 + hq_3 - bq_4 & \geq 0. \quad \text{(18e)}
\end{align*}

Proof. This dual formulation is obtained by assigning Lagrangian multipliers \(q_1, q_2, q_3, q_4 \geq 0\) to the constraints (17c)--(17f), respectively. The exact derivation is lengthy and is hence relegated to the appendix.

Theorem 7. Any optimal solution \(q^*\) of Problem (18) satisfies \(q_3^* = 0\).

As Problem (17) constitutes a convex minimization problem and as \(q_3\) represents the shadow price of the constraint (17e). The theorem implies that the constraint (17e) is redundant and can thus be removed without any impact on the optimal objective value. It could nonetheless be possible that, after the constraint elimination, the relaxed problem admits multiple optimal solutions and some of those may not be feasible in Problem (17).

Proof of Theorem 7. We divide our analysis into two cases depending on the value of \(q_4^*\). Our goal is to show that \(q_3^* = 0\).

Case 1 (\(q_4^* > 0\)): Suppose for the sake of a contradiction that \(q_3^* > 0\) and consider an alternative solution \(q^\dagger = (q_1^\dagger + \delta, q_2^\dagger + \delta, q_3^* - \delta, q_4^* - \delta)\) for some sufficiently small \(\delta > 0\) such that \(q^\dagger\) remains in the nonnegative orthant. It can be directly verified that \(q^\dagger\) is feasible in Problem (18). Besides, it holds that

\[
q_1^\dagger q_2^\dagger + q_1^\dagger q_3^\dagger + q_2^\dagger q_3^\dagger + q_3^\dagger q_4^\dagger = (q_1^* + \delta)(q_2^* + \delta) + (q_1^* + \delta)(q_4^* - \delta) + (q_2^* + \delta)(q_3^* - \delta) + (q_3^* - \delta)(q_4^* - \delta)
\]

\[
= q_1^* q_2^* + q_1^* q_4^* + q_2^* q_3^* + q_3^* q_4^*
\]

and that

\[
4q_1^\dagger q_2^\dagger + q_1^\dagger q_3^\dagger + q_1^\dagger q_4^\dagger + q_2^\dagger q_3^\dagger + q_2^\dagger q_4^\dagger = 4(q_1^* + \delta)(q_2^* + \delta) + (q_1^* + \delta)(q_4^* - \delta) + (q_2^* + \delta)(q_3^* - \delta) + (q_3^* + \delta)(q_4^* - \delta) + (q_2^* + \delta)(q_3^* - \delta) + (q_3^* + \delta)(q_4^* - \delta)
\]

\[
= 4q_1^* q_2^* + q_1^* q_4^* + q_2^* q_3^* + q_2^* q_4^* + q_3^* q_4^* + 2\delta
\]

as well as that

\[
3hq_1^\dagger - 3bq_2^\dagger + (2h - b)q_3^\dagger + (h - 2b)q_4^\dagger = 3h(q_1^* + \delta) - 3b(q_2^* + \delta) + (2h - b)(q_3^* - \delta) + (h - 2b)(q_4^* - \delta)
\]

\[
= 3hq_1^* - 3bq_2^* + (2h - b)q_3^* + (h - 2b)q_4^*.
\]

Together, the above three observations imply that \(q^\dagger\) attains an objective function value that is strictly larger than \(q^*\), which in turns contradicts with the supposed optimality of \(q^*\) and therefore renders \(q_3^* > 0\) impossible to materialize.
Case 2 \((q^*_4 = 0)\): Next, we consider the other case with vanishing \(q^*_4\). Under this condition, Problem (18) simplifies to a trivariate optimization problem:

\[
\begin{align*}
\text{maximize} & \quad (h + b)\sigma \left( \sqrt{4q_1q_2 + q_1q_3 + q_2q_3 + \sqrt{q_1q_2 + q_2q_3}} \right) - \mu (3hq_1 - 3bq_2 + (2h - b)q_3) \\
\text{subject to} & \quad q_1, \ q_2, \ q_3 \geq 0 \\
& \quad q_1 + q_2 + q_3 = 1 \\
& \quad c + 2hq_1 - 2bq_2 + (h - b)q_3 \geq 0 \\
& \quad c + bq_1 - bq_2 + bq_3 \geq 0
\end{align*}
\]

and subsequently to a bivariate optimization problem (barring the shifting and positive scaling of the objective function):

\[
\begin{align*}
\text{maximize} & \quad \sigma \left( \sqrt{q_1 + q_2 - (q_1 - q_2)^2 + \sqrt{q_2(1 - q_2)}} \right) + \mu (2q_2 - q_1) \\
\text{subject to} & \quad q_1, \ q_2 \geq 0 \\
& \quad q_1 + q_2 \leq 1 \\
& \quad q_2 - q_1 \leq \frac{c + h - b}{h + b} \\
& \quad q_2 \leq \frac{c + h}{h + b}
\end{align*}
\]

(19a) (19b) (19c) (19d) (19e)

We remark that Problem (19) constitutes a concave maximization problem because square root function is concave and increasing and a quadratic function with a non-positive leading coefficient is also concave. In view of Problem (19), our original goal to establish that \(q^*_4 = 0\) is tantamount to showing that the constraint (19c) is binding at optimality. We begin by arguing that the constraint (19e) cannot be binding because otherwise it would hold that

\[
q^*_1 + q^*_2 \geq 2q^*_2 - \frac{c + h - b}{h + b} = \frac{2(c + h)}{h + b} - \frac{c + h - b}{h + b} > 1,
\]

where the first inequality is due to (19d) and hence a contradiction. Therefore, the constraint (19e) can be omitted without any loss of optimality. Suppose next that the constraint (19d) is binding, Problem (19) simplifies (barring again the shifting and positive scaling of the objective function) to the following univariate concave maximization problem:

\[
\begin{align*}
\text{maximize} & \quad \sqrt{2q_2 - \frac{c + h - b}{h + b} - \left( \frac{c + h - b}{h + b} \right)^2} + \sqrt{q_2(1 - q_2)} + rq_2 \\
\text{subject to} & \quad \left( \frac{c + h - b}{h + b} \right)^+ \leq q_2 \leq \frac{2h + c}{2h + 2b},
\end{align*}
\]

(20)

where \(r\) is a positive constant equal to \(\frac{\mu}{\sigma}\). In Problem (20), the lower bound of \(q_2\) ensures that both \(q_1\) and \(q_2\) are nonnegative, whereas the upper bound ensures that \(q_1 + q_2 \leq 1\), which is an explicit
constraint in Problem (19). To show that (19c) is a binding constraint, it suffices to show that the upper bound \(\frac{2b+c}{2h+2b}\) is attained by \(q^*_3\), and to achieve this we will show that the objective function of (20) is increasing in \(q_2\) over its admissible range. Observe that Problem (20) can only be feasible when \(2b \geq c\), which we shall assume, and the derivative of the objective function denoted by \(g_{(20)}\) is:

\[
\frac{\partial g_{(20)}}{\partial q_2} = \frac{1}{\sqrt{2q_2 - \frac{c+h-b}{h+b} - \left(\frac{c+h-b}{h+b}\right)^2}} + \frac{1-2q_2}{2\sqrt{q_2(1-q_2)}} + r
\]

\[= \frac{h+b}{\sqrt{2(q_2 + b + c) - (c+h-b)(2h+c)}} + \frac{b-h-c}{(2h+b-c)(h+b)} + r
\]

When \(q_2 = \frac{2b+c}{2h+2b}\),

\[
\left.\frac{\partial g_{(20)}}{\partial q_2}\right|_{q_2 = \frac{2b+c}{2h+2b}} = \frac{h+b}{\sqrt{(h+b)(2h+c) - (c+h-b)(2h+c)}} + \frac{(b-h-c)/(h+b)}{\sqrt{(2h+c)(2b-c)/(h+b)}} + r
\]

\[= \frac{h+b}{\sqrt{(2h+c)(2b-c)}} + \frac{b-h-c}{\sqrt{(2h+c)(2b-c)}} + r
\]

\[= \sqrt{\frac{2b-c}{2h+c}} + r \geq 0.
\]

By its concavity, \(g_{(20)}\) is increasing in \(q_2 \in \left[\frac{c+h-b}{h+b}, \frac{2b+c}{2h+2b}\right]\) and as a result \(q^*_3\) indeed vanishes.

Henceforth, we can assume without any loss of optimality that constraints (19d) and (19e) are not binding. If at optimality the constraint (19c) is also not binding (i.e., if \(q^*_3 > 0\)), it must be possible to identify \(q^*\) that is optimal in Problem (19) by solving

\[
\begin{align*}
\text{maximize} & \quad \sqrt{q_1 + q_2 - (q_1 - q_2)^2} + \sqrt{q_2(1-q_2)} + r(2q_2 - q_1) \\
\text{subject to} & \quad 0 \leq q_1 \leq 1, \quad 0 \leq q_2 \leq 1.
\end{align*}
\]

Note that the square roots are well-defined over the feasible region as \(q_1 + q_2 \geq |q_1 - q_2| \geq (q_1 - q_2)^2\).

By a familiar argument that we use to analyze Problem (19) earlier, Problem (21) remains a concave maximization problem. Then, by expressing its objective function \(g_{(21)}\) as

\[
g_{(21)}(q_1, q_2) = \sqrt{q_1(1-q_1) + q_2(1-q_2) + 2q_1q_2} + \sqrt{q_2(1-q_2)} + r(2q_2 - q_1)
\]

and noting that the mapping \(q_2 \mapsto q_2(1-q_2)\) is strictly increasing over the interval \([0, \frac{1}{2}]\), it necessarily holds that \(q^*_2 \geq \frac{1}{2}\). Next, we derive the partial derivative of \(g_{(21)}\) with respect to \(q_1\):

\[
\frac{\partial g_{(21)}}{\partial q_1} = \frac{1-2q_1+2q_2}{2\sqrt{q_1 + q_2 - (q_1 - q_2)^2}} - r.
\]
When \( q_2 \) is chosen optimally as \( q_2^* \), we can determine the values of \( q_1 \) such that the above partial derivative vanishes as follows:

\[
(1 - 2q_1 + 2q_2^*)^2 = 4r^2 (q_1 + q_2^* - (q_1 - q_2^*)^2)
\]

\[
\iff 4(q_2^* - q_1)^2 + 4(q_2^* - q_1) + 1 = 8r^2 q_2^* - 4r^2(q_2^* - q_1) - 4r^2(q_2^* - q_1)^2
\]

\[
\iff 4(1 + r^2)(q_1^* - q_1)^2 + 4(1 + r^2)(q_2^* - q_1) + 1 - 8r^2 q_2^* = 0
\]

\[
\iff q_2^* - q_1 = -\frac{1}{2} \pm \frac{1}{2} \sqrt{\frac{r^2(1 + 8q_2^*)}{1 + r^2}}
\]

\[
\iff q_1 = q_2^* + \frac{1}{2} \pm \frac{1}{2} \sqrt{\frac{r^2(1 + 8q_2^*)}{1 + r^2}}.
\]

The above two values of \( q_1 \) together with the boundary points (i.e., 0 and 1) are the candidate values for \( q_1^* \). However, as \( q_2^* \geq \frac{1}{2} \) and as \( q_1^* + q_2^* \leq 1 \), we can rule out the values of \( q_1^* \) that are greater than \( \frac{1}{2} \). Hence, we find that \( q_1^* \in \{0, q_1^*\} \), where \( q_1^* = q_2^* + \frac{1}{2} - \frac{1}{2} \sqrt{\frac{r^2(1 + 8q_2^*)}{1 + r^2}} \). We can now divide our remaining analysis into two subcases.

**Case 2a (\( q_1^* = 0 \)):** In this case, as \( g_{(21)} \) is concave in \( q_1 \), we have \( q_1^* = 0 \). We will assume here that \( q_1^* < 0 \) and treat the remaining possibility of \( q_1^* = q_1^* = 0 \) in **Case 2b**. Here, Problem (21) reduces to a univariate optimization problem in \( q_2 \). By a slight abuse of notation, we denote the resulting objective \( g_{(21)}(0, q_2) \) by \( g_{(21)}(q_2) \), and we find

\[
\frac{\partial g_{(21)}}{\partial q_2} = \frac{1 - 2q_2}{\sqrt{q_2(1 - q_2)}} + 2r.
\]

This gradient simply evaluates to \( 2r > 0 \) when \( q_2 = \frac{1}{2} \), and it evaluates to a negative number when \( q_2 \) approaches one from below. By the intermediate value theorem, there exists a value of \( q_2 \in (\frac{1}{2}, 1) \) such that the gradient vanishes. Since \( g_{(21)} \) is concave in \( q_2 \), it then follows that

\[
\frac{1 - 2q_2^*}{\sqrt{q_2^*(1 - q_2^*)}} + 2r = 0 \implies r = \frac{2q_2^* - 1}{2\sqrt{q_2^*(1 - q_2^*)}} \implies \sqrt{\frac{r^2}{1 + r^2}} = 2q_2^* - 1.
\]

As a result, we obtain that

\[
q_2^* + \frac{1}{2} - \frac{1}{2} (2q_2^* - 1) \sqrt{1 + 8q_2^*} = q_1^* < 0 \implies (2q_2^* + 1)^2 < (2q_2^* - 1)^2 (1 + 8q_2^*).
\]

The latter inequality implies that \( 32(q_2^*)^2(q_2^* - 1) > 0 \), which is an impossible consequence in view of the feasibility of Problem (21).

**Case 2b (\( q_1^* = q_1^* \)):** Regardless of the value of \( q_1 \), we find

\[
\frac{\partial g_{(21)}}{\partial q_2} = \frac{1 + 2q_1 - 2q_2}{2\sqrt{q_1^* + q_2^* - (q_1^* - q_2^*)^2}} + \frac{1 - 2q_2}{2\sqrt{q_2^*(1 - q_2^*)}} + 2r.
\]

Similarly to the above case, for any fixed \( q_1 \leq \frac{1}{2} \), there exists a value of \( q_2 \in (\frac{1}{2}, 1) \) such that this gradient vanishes, and thus at optimality, we have

\[
\frac{1 + 2q_1^* - 2q_2^*}{2\sqrt{q_1^* + q_2^* - (q_1^* - q_2^*)^2}} + \frac{1 - 2q_2^*}{2\sqrt{q_2^*(1 - q_2^*)}} + 2r = 0. \tag{22}
\]
Note that, as \( q_1^* = q_1 \), it follows that

\[
2 \sqrt{q_1^2 + q_2^2 - (q_1^2 - q_2^2)}^2 = 2 \sqrt{2q_1^4 + \frac{1}{2} - \frac{1}{2} \sqrt{\frac{2}{1 + r^2}} - \frac{1}{4} \left( 1 - \sqrt{\frac{2}{1 + r^2}} \right)^2} = \sqrt{\frac{1 + 8q_2^2}{1 + r^2}}
\]

and that

\[
1 + 2q_1^* - 2q_2^* = 2 - r \sqrt{\frac{1 + 8q_2^2}{1 + r^2}}.
\]

Substituting these observations into (22), we obtain

\[
2 \sqrt{\frac{1 + r^2}{1 + 8q_2^2}} + \frac{1 - 2q_2^*}{2\sqrt{q_2^2(1 - q_2^2)}} + r = 0 \implies r < \frac{2q_2^* - 1}{2\sqrt{q_2^2(1 - q_2^2)}} \implies \sqrt{\frac{r^2}{1 + r^2}} < 2q_2^* - 1,
\]

where the rightmost implication holds because the mapping \( r \mapsto \sqrt{\frac{r^2}{1 + r^2}} \) is increasing in \( r \in [0, \infty) \).

As the inequality \( q_1 + q_2 \leq 1 \) is an explicit constraint of Problem (20), which is supposed to be non-binding, it must strictly hold when \( q_1 = q_1^* \) and \( q_2 = q_2^* \). Hence,

\[
2q_2^* + \frac{1}{2} - \frac{1}{2} \sqrt{\frac{2}{1 + r^2}} = q_1^* + q_2^* < 1 \implies (4q_2^* - 1)^2 < (2q_2^* - 1)^2(1 + 8q_2^2).
\]

The latter can hold only if \( 4q_2^*(q_2^* - 1)(8q_2^2 - 3) > 0 \) and hence a contradiction.

We thus conclude that, in all cases, it is impossible for \( q_1^* \) to be a strictly positive number and complete the proof. \( \square \)

As a progressive approximation, its optimal solution \( \mathbf{x}^* \) is not necessarily optimal in Problem (2), \textit{i.e.}, \( f(\mathbf{x}^*) \) could be strictly larger than \( f(\mathbf{x}^*) \), and this unfounded optimism could jeopardize the operations of the inventory manager. This statement holds even when \( T = 2 \) because \( \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \) involves optimizing over \( (\mathbf{x}, \alpha, \beta, \gamma) \) and it may admit multiple optimal solutions and some of those may not be feasible in \( \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \). We thus propose to use a conservative approximation which is due to Theorem 5 or Proposition 2 to obtain an approximate optimal solution \( \mathbf{x}^\dagger \) and approximate the associated worst-case cost \( f(\mathbf{x}^\dagger) \) by \( [f(\mathbf{x}^\dagger), f(\mathbf{x}^\dagger)] \) as well as upper bound the optimality gap \( f(\mathbf{x}^\dagger) - f(\mathbf{x}^*) \) by \( f(\mathbf{x}^\dagger) - f(\mathbf{x}^*) \). To avoid overloading the notation, we will henceforth use \( \mathbf{x}_1^\dagger \) and \( \mathbf{x}_0^\dagger \) to denote the \( L_\cdot \) and the \( Q\cdot \)-conservative solutions, respectively, in the remainder of the paper.

\textbf{Remark 3.} It is possible to extend our inventory management model and its approximations to cater for fixed ordering costs which are incurred whenever a strictly positive order is placed. Denoting by \( k \in \mathbb{R}_{++} \) a fixed ordering cost and introducing a decision vector \( \mathbf{z} \in \{0, 1\}^T \), the worst-case expected cost minimization problem could be written down as

\[
\begin{align*}
\text{minimize} \quad & k \mathbf{1}^\top \mathbf{z} + f(\mathbf{x}) \\
\text{subject to} \quad & \mathbf{x} \in \mathcal{X}, \quad \mathbf{z} \in \{0, 1\}^T \\
& \mathbf{x} \leq M \mathbf{z},
\end{align*}
\]
for a sufficiently large $M > 0$. As all of the previously proposed approximation methods are concerned with estimating $f(x)$ for any fixed $x \in X$, they could still be utilized here. However, due to the presence of $z$, the corresponding exact and approximation models will become mixed-integer second-order cone programs. Although non-convex, to some extent these problems can numerically be handled by off-the-shelf solvers, such as CPLEX and Gurobi.

5. Plug-and-play conic extensions

We now consider how to reduce the conservatism of the proposed distributionally robust inventory model and its approximations by injecting additional distributional information besides the mean and the variance. The results in this section primarily rely on the conic representations of our base and approximate models, which are the consequences of Theorem 1, Proposition 2, and Theorem 6. The main idea is that when the ambiguity set is shrunk, a new cone is derived, and it is then used in lieu of $K_T$ in our earlier results. We refer to this cone replacement as a plug-and-play feature of the studied distributionally robust inventory model.

5.1. Non-negative support

When returns are not allowed, naturally demands are non-negative with probability one, and to incorporate this information, we may consider the shrunk ambiguity set

$$P^* = \{ P \in \mathcal{M}_+(\mathbb{R}^T_+) : \mathbb{P}(\xi \in \mathbb{R}^T) = 1, \mathbb{E}_P[\xi_1] = \mu, \mathbb{E}_P[\xi_2] = \mu^2 + \sigma^2 \ \forall t = 1, \ldots, T \},$$

and the less conservative cost function

$$f^*(x) = \max_{P \in P^*} \mathbb{E}_P \left[ \sum_{t=1}^T \left( cx_t + \max \left\{ h \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \right) \right] .$$

It turns out that the majority of main results (Theorems 1, 4 and 6 as well as Proposition 2) readily extend by replacing the original cone $K_T$ therein with

$$K_T^* = \{ (\alpha, \beta, \gamma) \in \mathbb{R}_+ \times \mathbb{R}^T \times \mathbb{R}^T_+ : \exists (\delta, \theta) \in \mathbb{R}_+ \times \mathbb{R}^T, \alpha \geq 1^T \theta, (\theta_t, \beta_t - \delta_t, \gamma_t) \in K_1, \forall t = 1, \ldots, T \}.$$

One could verify that $K_T \subset K_T^*$, which could be perceived as a sign of reduced conservatism. It is a routine exercise to verify that the cone $K_T^*$ is proper.

Theorem 8. We have that

$$f^*(x) = \min \{ c^T x + \alpha + \mu^T \beta + (\mu^2 + \sigma^2) 1^T \gamma \}$$

subject to

$$\alpha \in \mathbb{R}, \ \beta \in \mathbb{R^T}, \ \gamma \in \mathbb{R^T}$$

$$(\alpha(x, e), \beta(e), \gamma) \geq K_T^*, \ 0 \ \forall e \in \mathcal{E},$$

where $\alpha(x, e)$ and $\beta(e)$ are defined as in Theorem 1, for all $x \in \mathbb{R}^T_+$. 
Proof. From an argument that is widely parallel to the proof of Theorem 1, we have that

\[
f^1(x) - c^T x = \text{minimize } \alpha + \mu 1^T \beta + (\mu^2 + \sigma^2) 1^T \gamma
\]

subject to \( \alpha \in \mathbb{R}, \beta \in \mathbb{R}^T, \gamma \in \mathbb{R}^T \)

\[
\alpha(x,e) + \sum_{t=1}^{T} \beta_t(e)\xi_t + \sum_{t=1}^{T} \gamma_t \xi_t^2 \geq 0 \quad \forall e \in \mathcal{E} \quad \forall \xi \in \mathbb{R}^T_+.
\]

For each fixed \( e \in \mathcal{E} \), the robust constraint holds for all \( \xi \in \mathbb{R}^T_+ \) iff there exists \( \theta \in \mathbb{R}^T \) such that

\[
\alpha(x,e) \geq 1^T \theta \quad \text{and} \quad \theta_t + \beta_t(e)\xi_t + \gamma_t \xi_t^2 \geq 0 \quad \forall \xi_t \geq 0 \quad \forall t = 1, \ldots, T.
\]

Note that it immediately holds that each \( \theta_t \) must be non-negative for otherwise the robust quadratic constraint fails to hold when \( \xi_t = 0 \). Furthermore, by the virtue of \( \mathcal{S} \)-lemma, each of these robust quadratic constraints holds iff there exists a \( \delta_t \geq 0 \), \( t = 1, \ldots, T \), such that

\[
\begin{bmatrix}
\gamma_t & \frac{1}{2} \beta_t(e) \\
\frac{1}{2} \beta_t(e) & \theta_t
\end{bmatrix} \succeq \mathbb{S}^+_+ \delta_t \begin{bmatrix}
0 & \frac{1}{2} \\
\frac{1}{2} & 0
\end{bmatrix} \iff \begin{cases}
\gamma_t \geq 0, & \theta_t \geq 0 \\
4\gamma_t \theta_t \geq (\beta_t(e) - \delta_t)^2
\end{cases}
\iff \begin{cases}
\gamma_t \geq 0, & \theta_t \geq 0 \\
(\theta_t, \beta_t(e) - \delta_t, \gamma_t) \succeq \mathcal{K}_T 0.
\end{cases}
\]

By leveraging the definition of \( \mathcal{K}_T^+ \), we finally complete the proof. \( \square \)

It appears in our experiments however that the difference between \( \min_{x \in \mathcal{X}} f(x) \) and \( \min_{x \in \mathcal{X}} f^1(x) \) is negligible. Hence, the non-negativity of the demands \( \xi \) may not be the most significant requirement to be embedded in the ambiguity set.

5.2. Uncorrelated demands

Next, we consider a scenario where the demands are uncorrelated, which itself could be an approximation of independent demands. Analogously to Section 5.1, we focus on the following restricted ambiguity set

\[
P^\perp = \left\{ P \in \mathcal{M}_+(\mathbb{R}^T) : \begin{array}{l}
P(\xi \in \mathbb{R}^T) = 1, \\
E_P[\xi_t] = \mu, \quad \forall t = 1, \ldots, T \\
E_P[\xi_t^2] = \mu^2 + \sigma^2, \quad \forall t = 1, \ldots, T \\
E_P[\xi_s \xi_t] = \mu^2, \quad \forall (s,t) : 1 \leq s < t \leq T
\end{array} \right\}
\]

and the corresponding less conservative cost function

\[
f^1(x) = \max_{P \in \mathcal{P}^\perp} E_P \left[ \sum_{t=1}^{T} e x_t + \max \left\{ h \left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right) \right\} \right].
\]

For the subsequent results, we use the following cone

\[
\mathcal{K}_T^+ = \left\{ (\alpha, \beta, \gamma, \Theta) \in \mathbb{R}^+ \times \mathbb{R}^T \times \mathbb{R}^T_+ \times \mathbb{S}^T : \begin{bmatrix} \text{diag}(\gamma) + \Theta \frac{1}{2} \beta^T \\ \frac{1}{2} \beta \end{bmatrix} \geq 0 \right\}.
\]

It follows from the Schur Complement Lemma that \( (\alpha, \beta, \gamma, 0) \in \mathcal{K}_T^+ \) iff \( (\alpha, \beta, \gamma) \in \mathcal{K}_T \). Besides, one could readily show that \( \mathcal{K}_T^+ \) is proper.
Theorem 9. We have that

\[
f^+(x) = \text{minimize} \quad c^T x + \alpha + \mu^T \beta + (\mu^2 + \sigma^2) \gamma + \mu^2 \Theta^T 1
\]

subject to \( \alpha \in \mathbb{R}, \beta \in \mathbb{R}^T, \gamma \in \mathbb{R}^T, \Theta \in \mathbb{S}^T \)

\[
(\alpha(x, e), \beta(e), \gamma, \Theta) \geq K_{\lambda}^+ 0 \quad \forall e \in \mathcal{E}
\]

\[
\text{diag}(\Theta) = 0,
\]

where \( \alpha(x, e) \) and \( \beta(e) \) are defined as in Theorem 1, for all \( x \in \mathbb{R}^T_+ \).

Proof. From an argument that is widely parallel to the proof of Theorem 1, we have that

\[
f^+(x) - c^T x = \text{minimize} \quad \alpha + \mu^T \beta + (\mu^2 + \sigma^2) \gamma + \mu^2 \sum_{1 \leq s < t \leq T} \theta_{st}
\]

subject to \( \alpha \in \mathbb{R}, \beta \in \mathbb{R}^T, \gamma \in \mathbb{R}^T, \Theta \in \mathbb{S}^T \times \mathbb{T} \)

\[
\alpha(x, e) + \sum_{t=1}^{T} \beta_t(e) \xi_t + \sum_{t=1}^{T} \gamma_t \xi_t^2 + \sum_{1 \leq s < t \leq T} \theta_{st} \xi_s \xi_t \geq 0 \quad \forall e \in \mathcal{E} \forall \xi \in \mathbb{R}^T,
\]

where each \( \theta_{st} \) is a dual variable assigned to the constraint that ensures \( \xi_s \) and \( \xi_t \) are uncorrelated. Note that \( \theta_{st} \) with \( s \geq t \) does not enter the above optimization problem.

Without any loss of optimality, we replace each \( \theta_{st}, s < t \), with \( 2 \theta_{st} \) and assign \( \theta_{ss} = 0 \) for all \( s \) and \( \theta_{st} = \theta_{st} \) for all pairs \( (s, t) \) such that \( s > t \) resulting in an equivalent problem of the form

\[
f^+(x) - c^T x = \text{minimize} \quad \alpha + \mu^T \beta + (\mu^2 + \sigma^2) \gamma + \mu^2 \Theta^T 1
\]

subject to \( \alpha \in \mathbb{R}, \beta \in \mathbb{R}^T, \gamma \in \mathbb{R}^T, \Theta \in \mathbb{S}^T \)

\[
\alpha(x, e) + \sum_{t=1}^{T} \beta_t(e) \xi_t + \sum_{t=1}^{T} \gamma_t \xi_t^2 + \xi^T \Theta \xi \geq 0 \quad \forall e \in \mathcal{E} \forall \xi \in \mathbb{R}^T
\]

\[
\text{diag}(\Theta) = 0.
\]

For any fixed \( e \in \mathcal{E} \), it can be recognized that the robust quadratic constraint (which has to hold for all \( \xi \in \mathbb{R}^T \)) is equivalent to

\[
\begin{bmatrix}
\xi \\
1
\end{bmatrix}
\begin{bmatrix}
\text{diag}(\gamma) + \Theta & \frac{1}{2} \beta(e) \\
\frac{1}{2} \beta(e)^T & \alpha(x, e)
\end{bmatrix}
\begin{bmatrix}
\xi \\
1
\end{bmatrix}
\geq 0 \quad \forall \xi \in \mathbb{R}^T \iff \begin{bmatrix}
\text{diag}(\gamma) + \Theta & \frac{1}{2} \beta(e) \\
\frac{1}{2} \beta(e)^T & \alpha(x, e)
\end{bmatrix}
\begin{bmatrix}
\xi \\
1
\end{bmatrix}
\geq 0.
\]

This latest observation together with the definition of \( K_{\lambda}^+ T \) completes the proof.

The above (exact) formulation is a correlation-aware counterpart of Theorem 1, which readily leads us to a progressive approximation akin to Theorem 6. For the conservative approximation, we can have the following results, which respectively are the counterparts of Theorem 4 and Proposition 2.
Theorem 10. We have that
\[
\begin{align*}
    f^+(x) = & \minimize \quad c x^T + \alpha + \mu 1^T \beta + (\mu^2 + \sigma^2) 1^T \gamma + \mu^2 1^T \Theta 1 \\
    \text{subject to} \quad & \alpha \in \mathbb{R}, \quad \beta \in \mathbb{R}^T, \quad \gamma \in \mathbb{R}^T, \quad \Theta \in \mathbb{S}^T, \quad \phi : \mathbb{R}^{2T+2} \rightarrow \mathbb{R}^T, \quad \psi : \mathbb{R}^{2T+2} \rightarrow \mathbb{R}^T \\
    & \alpha + \beta^T u + \gamma^T v + (\Theta, W) \geq \beta 1^T \phi(u, v, W) + h 1^T \psi(u, v, W) \\
    & \forall (u, v, W) : (1, u, v, W) \in (K_{P}^\perp)^* \\
    & y_0 + \phi_t(u, v, W) - \psi_t(u, v, W) = \sum_{\tau=1}^{T} (u_\tau - x_\tau) \quad \forall t = 1, \ldots, T \\
    & \forall (u, v, W) : (1, u, v, W) \in (K_{P}^\perp)^* \\
    \text{diag}(\Theta) = 0.
\end{align*}
\]

Proof. The proof widely parallels to that of Theorem 4 and is therefore omitted for brevity. \(\square\)

Proposition 4. We have that
\[
\begin{align*}
    f^+(x) \leq \minimize \quad c x^T + \alpha + \mu 1^T \beta + (\mu^2 + \sigma^2) 1^T \gamma + \mu^2 1^T \Theta 1 \\
    \text{subject to} \quad & \alpha \in \mathbb{R}, \quad \beta \in \mathbb{R}^T, \quad \gamma \in \mathbb{R}^T, \quad \Theta \in \mathbb{S}^T, \quad \kappa \in \mathbb{R}^{2T+1}, \\
    & (\pi_t, \pi_t^\nu, \pi_t^\nu, \Pi_t^w) \in \mathbb{R} \times \mathbb{R}^T \times \mathbb{R}^T \times \mathbb{S}^T \forall t = 1, \ldots, T \\
    & (\pi_t - \kappa_t, \pi_t^\nu, \pi_t^\nu, \Pi_t^w) \geq \kappa_t^\perp 0 \quad \forall t = 1, \ldots, T \\
    & \left(\pi_t + y_0 + \sum_{\tau=1}^{T} x_\tau - \kappa_{T+t}, \pi_t^\nu - (1, \ldots, 1, 0, \ldots, 0)^T, \pi_t^\nu, \Pi_t^w\right) \geq \kappa_t^\perp 0 \\
    & \forall t = 1, \ldots, T \\
    & \left(\alpha - h Ty_0 - h \sum_{t=1}^{T} x_t (T - t + 1) - (b + h) \sum_{t=1}^{T} \pi_t - \kappa_{2T+1}, \\
    & \beta - (b + h) \sum_{t=1}^{T} \pi_t^\nu + h (T, T - 1, \ldots, 1)^T, \\
    & \gamma - (b + h) \sum_{t=1}^{T} \pi_t^\nu, \Theta - (b + h) \sum_{t=1}^{T} \Pi_t^w\right) \geq \kappa_t^\perp 0 \\
    \text{diag}(\Theta) = 0,
\end{align*}
\]
where the vector \((1, \ldots, 1, 0, \ldots, 0)^T\) in the second line of constraints has \(t\) components equal to one and \(T - t\) components equal to zero, for all \(x \in \mathbb{R}^T_+\).

Proof. The proof widely parallels to that of Proposition 2, where the additional decision variables \(\Pi_t^w, t = 1, \ldots, T\), characterize the sensitivity of \(\phi\) in Problem (23) with respect to \(W\). \(\square\)

It could numerically be observed that both the progressive and the L-conservative approximation of \(\min_{x \in X} f^+(x)\) are still close and that \(\min_{x \in X} f^+(x)\) could be significantly smaller than \(\min_{x \in X} f(x)\). Hence, if there is evidence supporting that demands are serially uncorrelated, the inventory manager may wish to adopt these extended models with the cone \(K^\perp\) instead of \(K\) in
order to avoid being overly conservative in his/her ordering policy. Note also that the upper approx-
imation provided in Proposition 4 involves only the primal cone $\mathcal{K}_T$ and not the dual cone $(\mathcal{K}_T^*)^*$.  

6. Numerical experiments

In this section, we will compare the solutions of our distributionally robust inventory model with
those of the state-of-the-art benchmarks, where the latter have access to the demand distribution,
under which each marginal demand follows a two-point or a normal distribution. In addition, we also
assess the quality of our conservative and progressive approximations, both of which are scalable,
by synthetically determining how close they are to the original problem. Finally, we perform a
sensitivity analysis with respect to the cost parameters: $c$, $b$, and $h$.

6.1. Comparison with benchmarking approaches

We compare the exact minimizer $x^*$ of $\min_{x \in \mathcal{X}} f(x)$, which due to Theorem 1 is representable as
a finite second-order cone program, with the solutions of two stochastic inventory problems. First,
we refer to the minimizer of

$$
\begin{align*}
\min_{x \in \mathcal{X}} & \mathbb{E}_{\mathcal{P}} \left[ \sum_{t=1}^{T} \left( cx_t + \max \left\{ h \left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right) \right\} \right] \\
\text{subject to} & x \in \mathcal{X}
\end{align*}
$$

as a stochastic solution and that of

$$
\begin{align*}
\min_{x \in \mathcal{X}} & \mathbb{E}_{\mathcal{P}} \left[ \sum_{t=1}^{T} \left( cx_t(\xi_t^{\epsilon-1}) + \max \left\{ h \left( y_0 + \sum_{\tau=1}^{t} (x_\tau(\xi_t^{\epsilon-1}) - \xi_\tau) \right), -b \left( y_0 + \sum_{\tau=1}^{t} (x_\tau(\xi_t^{\epsilon-1}) - \xi_\tau) \right) \right\} \right] \\
\text{subject to} & \left( x_1, x_2(\xi_1^{\epsilon}), \ldots, x_T(\xi_T^{\epsilon-1}) \right)^\top \in \mathcal{X} \quad \mathbb{P}\text{-a.s.}
\end{align*}
$$

as an adaptive solution. In this experiment, we set aside computational tractability and are only
concerned with the quality of different solutions. Note that tractability will be the focus of our
next experiment concerning the proposed approximations. Throughout, it is assumed that $\mathcal{X} = \mathbb{R}_+^T$
and the demands are serially independent and identically distributed with

$$
\mathbb{P}\left( \xi_t = \xi_t^{\epsilon}\right) = p^L \quad \text{and} \quad \mathbb{P}\left( \xi_t = \xi_t^{\epsilon}\right) = p^H \quad \forall t \in \{1, \ldots, T\}
$$

for some $(p^L, \xi_t^{\epsilon}, p^H, \xi_t^{\epsilon})$ such that $\xi_t^{\epsilon} > \xi_t^{\epsilon}$, $p^L$ and $p^H$ are both non-negative, and they sum up to
one. It is well-known that the adaptive solution follows a base-stock policy (Scarf 1960), whereas
the stochastic solution has been recently studied by Basciftci et al. (2021), for instance. As a result,
both stochastic and adaptive solutions can be characterized by vectors in $\mathbb{R}_+^T$ which we shall denote
by $\tilde{x}^*$ and $S^*$, respectively. When $\mathbb{P}$ is really the true data-generating distribution, the base-stock
policy $S^*$ (despite its impracticality) yields a lower expected total cost than that of $\tilde{x}^*$, and the stochastic solution $\tilde{x}^*$ yields a lower expected total cost than that of the proposed robust solution $x^*$. As this premise however is likely untrue, inspired by Bertsimas et al. (2010), we perform a stress test by determining the expected total cost corresponding to $x^*$ and $\tilde{x}^*$ under the contaminated probability distribution

$$P_\lambda = (1 - \lambda)P + \lambda P_{wc},$$

where $\lambda \in [0, 1]$ represents the contamination level and $P_{wc}$ is the demand distribution that (approximately) attains the worst-case expected cost when $x = \tilde{x}^*$ ($\varepsilon = 10^{-4}$ is used). Note that, unlike $P$, $P_{wc}$ does not necessarily generate serially independent demands.

We consider two plausible scenarios: (i) when it is likely for the demand to be small but there can be a rare surge and (ii) when it is likely for the demand to be large but there might be a rare drop. As a representative of scenario (i), we set $p^L = 0.7$, $p^H = 0.3$, $\xi^L = 30$ and $\xi^H = 70$. Even though it is more likely for the demands to be small, the inventory manager should still prepare for a sudden surge in demand, which may force him or her to backlog the unmet demands. This is particularly crucial when backlogging is expensive, and to investigate this effect we set $b = 3 > 1 = h$ and $c = 8$. As a representative of scenario (ii), we similarly set $p^L = 0.3$, $p^H = 0.7$ and retain $\xi^L = 30$, $\xi^H = 70$. For this case, we are particularly concerned with the possible drop in the demands and high holding cost. We therefore set $h = 3 > 1 = b$ and $c = 3$. As none of these optimization problems are tractable, we choose to work with a relatively small $T = 6$, and we further assume that the initial inventory is $y_0 = 0$. Results for these two scenarios are reported in Figure 2 (left) and Figure 2 (right), respectively, from which it can be observed that the robust policy $x^*$ is more resistant to the misspecification in the demand distribution. Besides, the contamination level at which the robust solution starts to outperform the stochastic solution is given by $\lambda = 11.85\%$ for scenario (i) and by $\lambda = 34.78\%$ for scenario (ii), respectively.

Next, we similarly compare our robust solution with the base-stock policy $S^*$. However, since the base-stock policy is adaptive and thus at an advantageous position as future ordering decisions may depend on the realizations of the previous demands, we execute our robust policy in a shrinking horizon fashion to facilitate a fairer comparison. Put differently, for the first period we solve $\min_{x \in \mathcal{X}} f(x)$ to determine $x^*$ but only implement the here-and-now decision $x^*_1$. For the second period, we resolve the same optimization problem with the updated initial inventory and with the number of periods reduced by one, and we only implement the here-and-now ordering decision again, and so on and so forth; see e.g. Solyali et al. (2016) and Mamani et al. (2017). We evaluate both the adaptive robust and the base-stock policies under different contaminated demand distributions $P_\lambda$, $\lambda \in [0, 1]$, where we substitute for $P_{wc}$ the worst-case demand distribution.
when \( \mathbf{x} = \mathbb{E}_P(\hat{\mathbf{x}}^*) \) with \( \hat{\mathbf{x}}^* \) representing (adaptive) ordering decisions equivalent to \( \mathbf{S}^* \). For both scenarios (i) and (ii), when \( \lambda = 0 \) (and thus \( \mathbb{P}_\lambda = \mathbb{P} \)), the base-stock policy (which is known to be optimal under this distributional setting) is superior to the adaptive robust policy. Nonetheless, as the out-of-sample distribution deviates further from the in-sample distribution, Figure 3 shows that the adaptive robust policy becomes increasingly competitive and eventually outperforms the base-stock policy, which highlights the resilience of the proposed robust policies even in the adaptive setting. We also remark that the differences between the two policies are more prominent in scenario (ii), see Figure 3 (right), where the adaptive robust policy seems to have an inadvertent variance reduction effect for the distribution of the total cost. Finally, we can also relate Figure 3 to its non-adaptive counterpart, \( \text{i.e., Figure 2. While we indeed observe that the adaptive robust and the base-stock solutions yield a smaller expected cost than the robust and the stochastic solutions, respectively, the reduction is not monumental. Thus, when } T \text{ is only moderately large, determining the adaptive robust and/or the base-stock policy may not be a worthwhile pursuit considering the other managerial benefits of agreeing to an advance purchase.}

6.2. Quality of the proposed approximations
To validate the performance of our proposed approximations, we experiment with randomly generated problem instances. Without loss of generality, we set the unit ordering cost \( c \) to be one, and we independently choose \( h \) and \( b \) from a uniform distribution on the unit interval \([0, 1]\). Similarly, we normalize the mean demand \( \mu \) to be one without any loss, and then randomly generate its standard deviation \( \sigma \) from the uniform distribution \( U([0, 2]) \). For each \( T \in \{10, 20, 30, 40, 50\} \), we set \( y_0 = 0 \) and run fifty random experiments in total. Note that, when \( T \geq 20 \), \( \min_{x \in A} f(x) \) cannot be solved exactly within a time limit of 12 hours using MOSEK ApS (2019) and YALMIP interface.
We have three objectives here. Firstly, we would like to numerically show that the Q-conservative approximation that is due to Theorem 5 is (marginally) better than the L-conservative approximation that is due to Proposition 2. To this end, we measure the improvement in conservativeness by

\[
\frac{\overline{f}_L(x^*_L) - \overline{f}_Q(x^*_Q)}{\overline{f}_Q(x^*_Q)},
\]

where \(x^*_L\) and \(x^*_Q\) denote the L- and the Q-conservative solution, respectively. Secondly, we want to quantify how close \(x^*_Q\) is to being optimal in view of the problem \(\min_{x \in X} f(x)\). We define the relative optimality gap as \((f(x^*_Q) - f(x^*)) / f(x^*)\) and propose to upper bound it by

\[
\frac{f(x^*_Q) - f(x^*)}{f(x^*)} \leq \frac{\overline{f}_Q(x^*_0) - f(x^*)}{f(x^*)} \leq \frac{\overline{f}_Q(x^*_0) - f(x^*)}{f(x^*)},
\]

and the rightmost expression, dependent only on the approximate worst-case expected total costs, could be computed rather efficiently even for a large \(T\). Finally and similarly to the above, we can also measure the precision of our Q-conservative solution, \(x^*_Q\), via the following metric

\[
\frac{\overline{f}_Q(x^*_0) - f(x^*_Q)}{f(x^*_Q)}.
\]

Note that by construction, our precision metric is necessarily smaller than the upper bound of the relative optimality gap in (25). These three measurements are computed for each simulation run and reported collectively in Figure 4. On average, the solution \(x^*_Q\) is less conservative than \(x^*_L\), and
it is also close to being robustly optimal. While $\mathcal{F}^*_1$ appears inferior in this experiment, it can cater for additional requirements (see Section 5) more readily and is thus more flexible.

Given the quality of the proposed progressive approximation, we can use the underlying scenario-reduction idea to construct an approximate worst-case distribution by solving a variant of Problem (9) with the exponentially-sized uncertainty set $\mathcal{E}$ replaced by $\tilde{\mathcal{E}} = \{\tilde{\mathcal{E}}^0, \ldots, \tilde{\mathcal{E}}^T\}$. In this case, the construction of the approximate worst-case distribution is dependent on the optimally chosen $\tilde{\alpha}^*: \tilde{\mathcal{E}} \rightarrow \mathbb{R}$, $\tilde{\beta}^*: \tilde{\mathcal{E}} \rightarrow \mathbb{R}^T$, $\tilde{\gamma}^*: \tilde{\mathcal{E}} \rightarrow \mathbb{R}^T$. While the construction of the distribution described in (12) and (13) indicates that demands should be statistically dependent in the worst-case, we provide evidence asserting a stronger statement that they are in fact highly correlated (i.e., they are strongly linearly dependent). As an illustrative example, we set $c = 1$, $h = 1$, $b = 1/3$, $T = 20$, $y_0 = 0$, $\mu = 1$ and consider different $\sigma \in \{\sqrt{0.001}, \sqrt{0.005}, \sqrt{0.025}\}$. For each of the parameter settings, we compute the $\mathcal{F}^*_2$ and the corresponding (approximate) worst-case distribution accordingly to the described procedure. We visualize the resultant correlation matrices with heatmaps in Figure 5 and find that there exists a breakpoint $t^* \in \{1, \ldots, T\}$ such that the demands $\xi_1, \ldots, \xi_{t^*}$ are almost perfectly positively correlated and so are the remaining demands $\xi_{t^*+1}, \ldots, \xi_T$. One could think of $t^*$ as the period with a transitional shift, e.g., a sudden surge or a sudden drop in the demand pattern, which reminds us of several relatable reasons, such as the availability of the substitute products. Recent examples include energy supplies and Covid-19 vaccines. This observation remains unchanged with different parameter settings, and it sheds light on the potential benefit of using the uncorrelated ambiguity set $\mathcal{P}^\perp$ instead of $\mathcal{P}$ when appropriate.

### 6.3. Normality and independence

Here, we revisit the stochastic solution but consider a demand distribution $\mathbb{P}$ which is serially independent and is characterized by $\xi_t \sim \mathcal{N}(\mu, \sigma^2)$, $\forall t$, as opposed to the two-point distribution previously considered. For the ease of exposition, we introduce the cumulative demands $\zeta_t = \sum_{\tau=1}^{t} \xi_\tau$, $\forall t$. It follows that $\zeta_t$ is a normal random variable with the following density function

$$f_t(z) = \frac{1}{\sigma \sqrt{2t\pi}} \exp\left(-\frac{(z-t\mu)^2}{2t\sigma^2}\right).$$

\[\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4}
\caption{Measurements of our approximation quality: (left) improvement in conservativeness of $\mathcal{F}^*_2$ over $\mathcal{F}^*_1$, (middle) our conservative bound on the relative optimality gap of $\mathcal{F}^*_2$, (right) precision of $\mathcal{F}^*_2$.}
\end{figure}\]
Besides, we denote the cumulative distribution function of \( \zeta_t \) by \( F_t \). By its definition, the stochastic solution \( \bar{x}^* \) minimizes an objective \( g(x) \) which equals
\[
\sum_{t=1}^{T} \left\{ c x_t + \int_{-\infty}^{y_0 + x_1 + \ldots + x_T} \left( y_0 + \int_{-\infty}^{y_0 + x_1 + \ldots + x_T} f_t(\zeta_t) \, d\zeta_t - b \int_{y_0 + \sum_{\tau=1}^{t} x_{\tau}}^{+\infty} \left( y_0 + \sum_{\tau=1}^{t} x_{\tau} - \zeta_{\tau} \right) f_t(\zeta_t) \, d\zeta_t \right) \right\}.
\]
Note that, even though the above objective function \( g \) is convex in \( x \), its explicit characterization involves the Gauss error function \( \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{z} \exp(-u^2) \, du \); see e.g. Glaisher (1871). Despite this complication, however, the gradient of \( g \) can be easily determined as
\[
\frac{\partial g}{\partial x_t} = c + h \sum_{t=1}^{T} \frac{\partial}{\partial x_t} \int_{-\infty}^{y_0 + x_1 + \ldots + x_T} \left( y_0 + x_1 + \ldots + x_{\tau} - \zeta_{\tau} \right) f_t(\zeta_t) \, d\zeta_t - b \sum_{t=1}^{T} \frac{\partial}{\partial x_t} \int_{y_0 + x_1 + \ldots + x_{\tau}}^{+\infty} \left( y_0 + x_1 + \ldots + x_{\tau} - \zeta_{\tau} \right) f_t(\zeta_t) \, d\zeta_t
\]
\[
= c + h \sum_{t=1}^{T} \left\{ \frac{\partial}{\partial x_t} \left( y_0 + x_1 + \ldots + x_{\tau} \right) F_t(y_0 + x_1 + \ldots + x_{\tau}) - \frac{\partial}{\partial x_t} \int_{-\infty}^{y_0 + x_1 + \ldots + x_{\tau}} \zeta_t f_t(\zeta_t) \, d\zeta_t \right\} - b \sum_{t=1}^{T} \left\{ \frac{\partial}{\partial x_t} \left( y_0 + x_1 + \ldots + x_{\tau} \right) \left( 1 - F_t(y_0 + x_1 + \ldots + x_{\tau}) \right) - \frac{\partial}{\partial x_t} \int_{y_0 + x_1 + \ldots + x_{\tau}}^{+\infty} \zeta_t f_t(\zeta_t) \, d\zeta_t \right\}
\]
\[
= c + h \sum_{t=1}^{T} F_t(y_0 + x_1 + \ldots + x_{\tau}) - b \sum_{t=1}^{T} \left( 1 - F_t(y_0 + x_1 + \ldots + x_{\tau}) \right) = c - b(T - t + 1) + (h + b) \sum_{t=1}^{T} F_t(y_0 + x_1 + \ldots + x_{\tau}),
\]
where the third equality holds because of the Fundamental Theorem of Calculus. As a result, the stochastic solution can be obtained by using a projected gradient method (Alber et al. 1998), which iteratively constructs a new solution by following the gradient descent direction and projecting it on the feasible set \( X = \mathbb{R}_+^T \) when necessary.

Under this normality and independence assumption, we then proceed to compare the stochastic solution \( \bar{x}^* \) with our conservative solution \( \bar{x}_Q^* \). Figure 6 (left) shows the total order of both solutions, i.e., \( 1^\top \bar{x}^* \) and \( 1^\top \bar{x}_Q^* \) when \( T = 20, c = 1, b = 0.5, y_0 = 0, \mu = 1 \) and \( \sigma = 0.5 \) for varying \( h \in \{0, \ldots, 1\} \).
Figure 6 Total order comparison between our robust and the stochastic solutions under the normality and independence assumption.

Similarly, Figure 6 (right) shows the total order of both solutions for the same set of parameters with the exception that now $h$ is fixed at 0.5 and $b$ is varied in $\{0, \ldots, 1\}$. As expected, when the holding cost increases, both solutions order less, and when it is expensive to backlog, both solutions order more. From the figure, in comparison to the stochastic solution, it can be seen that the robust solution is more receptive to the changes in $h$ and $b$, which we attribute to the fact that the change in $b$ and $h$ has an impact on the worst-case demand distribution but not on the nominal normal distribution. Note that each bar consists of two parts: the darker-shaded part represents the total order made in the first half of the planning horizon ($t \in \{1, \ldots, T/2\}$), whereas the lighter-shaded part represents that made in the second half ($t \in \{T/2 + 1, \ldots, T\}$). For both solutions, at least half of the total order is made in the first half of the planning horizon. Nevertheless, the ratio between the first-half and the second-half order of the robust solution appears significantly greater than that of the stochastic solution.

**Acknowledgements:** This research is supported by the Ministry of Education, Singapore, under its 2019 Academic Research Fund Tier 3 grant call (Award ref: MOE-2019-T3-1-010).

**References**


**Appendix (Proof of Proposition 3)**

The proof of Proposition 3 builds on two technical lemmas, which we present below.

**Lemma 3.** Given $a \in \mathbb{R}_+$ and $b, c \in \mathbb{R}$, it holds that

$$\min_{z \in \mathbb{R}} \{az^2 + bz + c\} = c - \frac{b^2}{4a}.$$ 

**Proof.** When $a = 0$ and $b \neq 0$, the minimization problem is unbounded. When $a = 0$ but $b = 0$, the optimal objective value trivially evaluates to $c$ which coincides with the right-hand side because of our convention that $0/0 = 0$. Finally, when $a > 0$, it is a routine exercise to verify that the minimizer $z^*$ is $-\frac{b}{2a}$ and hence the statement follows again. \hfill $\square$

**Lemma 4.** For $a, q \in \mathbb{R}^n$ such that $1^Tq = 1$, it holds that

$$\left(\sum_{i=1}^n q_i a_i^2\right) - \left(\sum_{i=1}^n q_i a_i\right)^2 = \sum_{1 \leq i < j \leq n} q_i q_j (a_i - a_j)^2.$$
Observe that if \( q \geq 0 \), a direct consequence of Lemma 4 is that \((\sum_{i=1}^{n} q_i a_i^2) - (\sum_{i=1}^{n} q_i) \geq 0\), which coincides with the result of the Cauchy-Schwarz inequality.

**Proof of Lemma 4.** The statement can be verified algebraically as

\[
\left( \sum_{i=1}^{n} q_i a_i^2 \right) - \left( \sum_{i=1}^{n} q_i a_i \right)^2 = \left( \sum_{i=1}^{n} q_i a_i^2 \right) \left( \sum_{i=1}^{n} q_i \right) - \left( \sum_{i=1}^{n} q_i a_i \right)^2 \\
\quad = \left[ \sum_{i=1}^{n} q_i^2 a_i^2 + \sum_{1 \leq i < j \leq n} q_i q_j (a_i^2 + a_j^2) \right] - \left[ \sum_{i=1}^{n} q_i^2 a_i^2 + \sum_{1 \leq i < j \leq n} 2q_i q_j a_i a_j \right] \\
\quad = \sum_{1 \leq i < j \leq n} q_i q_j (a_i^2 + a_j^2 - 2a_i a_j) = \sum_{1 \leq i < j \leq n} q_i q_j (a_i - a_j)^2.
\]

The proof is then completed. \( \square \)

We are now ready to utilize Lemmas 3 and 4 to validate Proposition 3, that is, Problem (18) is dual to Problem (17).

**Proof of Proposition 3.** Let \( q \in \mathbb{R}_+^4 \) be a collection of Lagrange multipliers of the four constraints in Problem (17) respectively. We write down the associated Lagrangian function as

\[
\mathcal{L}(\mathbf{x}, \alpha, \beta, \gamma, q) = c(x_1 + x_2) + \alpha + \mu(\beta_1 + \beta_2) + (\mu^2 + \sigma^2)(\gamma_1 + \gamma_2) - \\
q_1 \left( \alpha - 2hx_1 - hx_2 - (\beta_1 + 2h)^2/4\gamma_1 - (\beta_2 + h)^2/4\gamma_2 \right) - \\
q_2 \left( \alpha + 2hx_1 + hx_2 - (\beta_1 - 2b)^2/4\gamma_1 - (\beta_2 - b)^2/4\gamma_2 \right) - \\
q_3 \left( \alpha + (b - h)x_1 - hx_2 - (\beta_1 + h - b)^2/4\gamma_1 - (\beta_2 + h)^2/4\gamma_2 \right) - \\
q_4 \left( \alpha + (b - h)x_1 + bx_2 - (\beta_1 + h - b)^2/4\gamma_1 - (\beta_2 - b)^2/4\gamma_2 \right).
\]

The dual problem is thus

\[
\max_{q \geq 0} \min_{x_0, y_0, \gamma_0, \alpha, \beta} \mathcal{L}(\mathbf{x}, \alpha, \beta, \gamma, q).
\]

For the inner minimization to be bounded from below with respect to the choice of \( x_1 \), the inequality (18d) must hold. With respect to the choice of \( x_2 \) and \( \alpha \), the inequality (18e) and the equality (18c) must similarly hold, respectively. These three conditions show up explicitly as dual constraints in Problem (18).

After eliminating \( x \) and \( \alpha \), the Lagrangian \( \mathcal{L} \) can be decomposed into two parts:

\[
\frac{1}{4\gamma_1} \left[ 4\mu \beta_1 \gamma_1 + q_1 (\beta_1 + 2h)^2 + q_2 (\beta_1 - 2b)^2 + q_3 (\beta_1 + h - b)^2 + q_4 (\beta_1 + h - b)^2 \right] + (\mu^2 + \sigma^2) \gamma_1,
\]

which depends only on \( \beta_1, \gamma_1, q \) and is thus abbreviated as \( \mathcal{L}_1(\beta_1, \gamma_1, q) \), and

\[
\frac{1}{4\gamma_2} \left[ 4\mu \beta_2 \gamma_2 + q_1 (\beta_2 + h)^2 + q_2 (\beta_2 - b)^2 + q_3 (\beta_2 + h)^2 + q_4 (\beta_2 - b)^2 \right] + (\mu^2 + \sigma^2) \gamma_2,
\]

which depends solely on \( \beta_2, \gamma_2, q \) and we shall abbreviate it as \( \mathcal{L}_2(\beta_2, \gamma_2, q) \).
Next, we invoke Lemma 3 to minimize $L_1(\beta_1, \gamma_1, q)$ over $\beta_1$ and find that the minimum objective is
\[
(\mu^2 + \sigma^2)\gamma_1 + \frac{1}{4\gamma_1} \left[ 4q_1h^2 + 4q_2b^2 + q_3(h-b)^2 + q_4(h-b)^2 - \frac{1}{4} (4\mu\gamma_1 + 4q_1h - 4q_2b + 2q_3(h-b) + 2q_4(h-b))^2 \right].
\] (26)

Utilizing Lemma 3 again to minimize $L_2(\beta_2, \gamma_2, q)$ over $\beta_2$ yields the minimum objective of
\[
(\mu^2 + \sigma^2)\gamma_2 + \frac{1}{4\gamma_2} \left[ q_1h^2 + q_2b^2 + q_3h^2 + q_4b^2 - \frac{1}{4} (4\mu\gamma_2 + 2q_1h - 2q_2b + 2q_3h - 2q_4b)^2 \right].
\] (27)

We subsequently leverage Lemma 4 with $a = (2h, -2b, h-b, h-b)^T$ to simplify $\min_{\beta_1} L_1(\beta_1, \gamma_1, q)$ from (26) further to
\[
\gamma_1\sigma^2 - \mu (2q_1h - 2q_2b + q_3(h-b) + q_4(h-b)) + \frac{1}{4\gamma_1} (h+b)^2 (4q_1q_2 + q_1q_3 + q_1q_4 + q_2q_3 + q_2q_4),
\] (28)

and with $a = (h, -b, h-b, h-b)^T$ to simplify $\min_{\beta_2} L_2(\beta_2, \gamma_2, q)$ from (27) further to
\[
\gamma_2\sigma^2 - \mu (q_1h - q_2b + q_3h - q_4b) + \frac{1}{4\gamma_2} (h+b)^2 (q_1q_2 + q_1q_4 + q_2q_3 + q_3q_4).
\] (29)

From (28) and (29), we respectively use the arithmetic mean–geometric mean inequality to argue that $\min_{\gamma_1 \geq 0, \beta_1} L_1(\beta_1, \gamma_1, q)$ is equal to
\[
\sigma(h+b)\sqrt{4q_1q_2 + q_1q_3 + q_1q_4 + q_2q_3 + q_2q_4 - \mu (2q_1h - 2q_2b + q_3(h-b) + q_4(h-b))}
\]
and that $\min_{\gamma_2 \geq 0, \beta_2} L_2(\beta_2, \gamma_2, q)$ is equal to
\[
\sigma(h+b)\sqrt{q_1q_2 + q_1q_4 + q_2q_3 + q_3q_4 - \mu (q_1h - q_2b + q_3h - q_4b)}.
\]

The summation of $\min_{\gamma_1 \geq 0, \beta_1} L_1(\beta_1, \gamma_1, q)$ and $\min_{\gamma_2 \geq 0, \beta_2} L_2(\beta_2, \gamma_2, q)$ gives rise to the dual objective in Problem (18). The proof is now completed. \[\square\]