A Unifying Framework for the Capacitated Vehicle Routing Problem under Risk and Ambiguity

Shubhechyya Ghosal[1], Chin Pang Ho[2], and Wolfram Wiesemann[1]

[1] *Imperial College Business School, Imperial College London, United Kingdom*

s.ghosal16@imperial.ac.uk, ww@imperial.ac.uk

[2] *School of Data Science, City University of Hong Kong, Hong Kong*

clint.ho@cityu.edu.hk

October 14, 2021

Abstract

We propose a generic model for the capacitated vehicle routing problem (CVRP) under demand uncertainty. By combining risk measures or disutility functions with complete or partial characterizations of the probability distribution governing the demands, our formulation bridges the popular but often independently studied paradigms of stochastic programming and distributionally robust optimization. We characterize when an uncertainty-affected CVRP is (not) amenable to a solution via a popular branch-and-cut scheme, and we elucidate how this solvability relates to the interplay between the employed decision criterion and the available description of the uncertainty. Our framework offers a unified treatment of several CVRP variants from the recent literature, such as formulations that optimize the requirements violation or the essential riskiness indices, while it at the same time allows us to study new problem variants, such as formulations that optimize the worst-case expected disutility over Wasserstein or ϕ-divergence ambiguity sets. All of our formulations can be solved by the same branch-and-cut algorithm with only minimal adaptations, which makes them attractive for practical implementations.

Keywords: Capacitated Vehicle Routing Problem; Stochastic Programming; Distributionally Robust Optimization; Branch-and-Cut.

1 Introduction

The *capacitated vehicle routing problem* (CVRP; Christofides 1976), originally coined the *truck dispatching problem* (Dantzig and Ramser, 1959), asks for the cost-optimal delivery of a single product to geographically dispersed customers through a fleet of homogeneous and capacity-constrained vehicles. It is one of the fundamental problems in logistics, and its variants have found manifold applications, among others, in the delivery and collection of goods and waste, dial-a-ride services as well as the routing of engineers, school buses and snow plow trucks (Toth and Vigo, 2014).

The classical CVRP assumes that all problem parameters, most notably the customer demands and travel times or costs, are known precisely. In many applications, however, the customer demands are unknown for *aleatoric* (*e.g.*, in collection problems, where the waste to be collected is unknown prior to arrival at the customer site) and/or *epistemic* reasons (*e.g.*, in delivery problems, where the actual demand for vehicle space induced by a customer's order differs from the demand predicted by simplified models). Likewise, the travel times (and hence, costs) are typically affected by uncertain traffic conditions. In response to these challenges, and following a wider trend to integrate data into the model building process of operations research, a wide variety of CVRPs under uncertainty have been proposed in recent decades. In this paper, we focus on the CVRP with uncertain customer demands and make the simplifying assumption that the travel times are known. This does not represent a judgment on the relative importance of the two types of uncertainty; it merely facilitates a more concise treatment of what turns out to remain a challenging problem.

Research on the CVRP under uncertainty can be categorized along two dimensions: the available information about the uncertainty and the decision maker's attitude towards the uncertainty. To date, three of the predominant approaches for capturing the available information about the uncertainty are *stochastic programming* (Birge and Louveaux, 2011; Shapiro et al., 2014), which assumes that the uncertain parameters follow a known probability distribution, *robust optimization* (Ben-Tal et al., 2009; Bertsimas et al., 2011), which stipulates that the uncertain parameters are only known to be realized within an uncertainty set, and *distributionally robust optimization* (Delage and Ye, 2010; Wiesemann et al., 2014), which assumes that the probability distribution governing the uncertain parameters is only known to belong to an ambiguity set of rival distributions. The decision maker's attitude towards uncertainty is characterized by the choice of a risk measure (such as the expected value or the value-at-risk) or an expected disutility functional, both of which make random variables comparable in terms of their desirability by mapping them to deterministic quantities. In distributionally robust optimization, where the precise probability distribution is not known, a robust decision is sought that optimizes the worst risk or disutility over all distributions contained in the ambiguity set. The combinations of different informational assumptions and attitudes towards uncertainty have led to a plethora of papers that investigate different variants of the CVRP under uncertainty and propose tailored solution approaches. This wealth of alternative methods can easily overwhelm both researchers and practitioners.

In this paper, we develop a unifying framework for the uncertainty-affected CVRP that bridges the paradigms of stochastic programming and distributionally robust optimization. Our framework combines a versatile ambiguity set with a rich class of risk measures and disutility functions, all combinations of which can be solved by minor variations of a well-known branch-and-cut scheme for the deterministic CVRP that eliminates subtours and capacity violating routes through rounded capacity inequalities (Laporte and Norbert, 1983; Lysgaard et al., 2004). Contrary to the deterministic CVRP, where the right-hand sides of the rounded capacity inequalities constitute cumulative demands over subsets of customers, the right-hand sides in our framework are determined by the optimal values of efficiently solvable optimization problems. As a result, the performance of our branch-and-cut scheme for the stochastic and distributionally robust CVRP is broadly comparable to that of the standard branch-and-cut schemes for the deterministic CVRP. This is in stark contrast to many existing solution approaches for the distributionally robust CVRP, which account for uncertainty via model reformulations that scale primarily to small and medium sized instances.

More specifically, the contributions of the present work can be summarized as follows.

(i) We study which classes of vehicle routing problems are (not) amenable to a solution via a popular branch-and-cut scheme based on rounded capacity inequalities, as well as how the right-hand sides of these inequalities should be selected. This part of our investigation is generic and may find applications in vehicle routing problems other than the CVRP under uncertainty.

(ii) We apply our findings to the CVRP under demand uncertainty. To this end, we consider an ambiguity set that encompasses the stochastic CVRP, certain classes of moment-based distributionally robust CVRPs as well as data-driven CVRPs over Wasserstein and ϕ-divergence based ambiguity sets, and we combine our ambiguity set with rich classes of risk measures and disutility functions. We show how the emerging variants of the uncertainty-affected CVRP

can all be solved by the same branch-and-cut scheme with only minimal adaptations.

(iii) We present numerical results which demonstrate that the considered classes of uncertainty-affected CVRPs possess similar solvability characteristics as those of the deterministic CVRP. The source code of our implementation is made available open source to facilitate reuse in applications, extensions as well as computational comparisons.[1]

To our best knowledge, we propose the first framework for the stochastic and distributionally robust CVRP that combines multiple ambiguity sets with different risk measures and disutility functions. Since we account for uncertainty via adaptations of the rounded capacity inequalities, our formulations also appear to scale more gracefully to larger problem instances. Although we focus on branch-and-cut schemes in the present work, we emphasize that our findings can be employed in branch-and-cut-and-price schemes for the uncertainty-affected CVRP as well.

Our paper relates to the rich and rapidly growing area of vehicle routing under uncertainty. For the sake of brevity, we restrict our review of the related literature to exact approaches for the robust and distributionally robust CVRP; for reviews of the stochastic CVRP as well as heuristic methods, we refer to Gendreau et al. (2014, 2016) and Oyola et al. (2018).

The robust CVRP has first been studied by Sungur and Ordóñez (2008), who assume that the customer demands and travel times are uncertain. The authors determine vehicle routes that satisfy the vehicle capacities and delivery time windows even when all customer demands and travel times can attain their worst-case realizations simultaneously. Under this assumption, the problem simplifies to a deterministic CVRP, which is computationally attractive but may result in overly conservative solutions. Subsequent works have addressed this conservatism by specifying uncertainty sets that preclude such pathological scenarios and solving the resulting robust optimization problems via model reformulations (Ordóñez, 2010; Agra et al., 2012; Gounaris et al., 2013), branch-and-cut schemes (Agra et al., 2013a; Gounaris et al., 2013) as well as branch-and-cut-and-price schemes (Lee et al., 2012; Lu and Gzarao, 2019; Munari et al., 2019; Pessoa et al., 2021). Model reformulations, which are typically based on duality results from classical robust optimization, tend to result in large-scale mixed-integer programs without any readily exploitable problem structure and thus apply primarily to small and medium sized problems. In contrast, the branch-and-cut-and-price schemes currently seem to display the best performance on large in-

[1] The source code is available at: `http://wp.doc.ic.ac.uk/wwiesema/sourcecodes/`.

stances. That said, branch-and-cut-and-price schemes typically require the repeated solution of robust shortest path problems with resource constraints (Pessoa et al., 2015; Pugliese et al., 2019) to generate candidate routes, and efficiently solvable versions of this problem have thus far been identified only for specific classes of uncertainty sets, such as budget uncertainty sets (Bertsimas and Sim, 2004). Branch-and-cut schemes, finally, appear to generalize more easily to larger classes of uncertainty sets. A noteworthy exception to this observation is the recent work of Wang et al. (2021), who combine a branch-and-cut-and-price algorithm for the deterministic CVRP with robust versions of rounded capacity inequalities and thus combine the strengths of the branch-and-cut as well as branch-and-cut-and-price schemes. Lastly, we also note the related works of Agra et al. (2013b) and Eufinger et al. (2020), which study the robust CVRP with deterministic demands but uncertain travel times, as well as Subramanyam et al. (2021), who study a variant of the robust CVRP where the presence of customers is uncertain.

The distributionally robust CVRP appears to have been first studied by Gounaris et al. (2013), who extend their results for the robust CVRP to a distributionally robust chance constrained CVRP over a moment ambiguity set. The authors reformulate the problem as a robust CVRP and solve it approximately using a model reformulation. Results are reported for standard benchmark instances with up to 23 customers. Major progress has been made by Adulyasak and Jaillet (2015), Jaillet et al. (2016) and Zhang et al. (2019, 2021), who study the traveling salesman problem and the CVRP with known demands but uncertain travel times. The authors propose stochastic and distributionally robust models that minimize the violation of pre-specified time windows via a lateness, a requirements violation, a service fulfillment risk and an essential riskiness performance index. The distributionally robust formulations consider all probability distributions characterized by moment conditions or the Wasserstein distance to a reference distribution. The resulting problems are solved via Benders decomposition, branch-and-cut schemes and a variable neighborhood search heuristic. While the reported runtimes are difficult to interpret since the authors propose new instances, it appears that the approaches are mainly suitable for small and medium sized problems. Dinh et al. (2018) study a stochastic and a distributionally robust version of the chance constrained CVRP where the customer demands are uncertain and the vehicles' capacity constraints need to be met with high probability. Their distributionally robust formulation assumes that the unknown true distribution is characterized through the means and the covariances of the customer demands. The authors propose a branch-and-cut-and-price algorithm for both formulations, and they successfully solve a large fraction of the standard benchmark instances. The distributionally robust chance constrained CVRP with uncertain customer demands has also been studied by Ghosal and Wiesemann (2020), who characterize the distribution governing the customer demands through moment conditions and develop a branch-and-cut solution scheme that performs well on the standard benchmark instances. Similar to the present work, the algorithm of Ghosal and Wiesemann (2020) uses adaptations of the rounded capacity inequalities to account for uncertainty. In contrast to this paper, however, Ghosal and Wiesemann (2020) restrict themselves to two closely related moment ambiguity sets, and their approach to compute the right-hand sides of the rounded capacity inequalities does not seem to easily extend to other ambiguity sets or risk measures. We also note the related works of Carlsson and Delage (2013), Carlsson and Behroozi (2017) and Carlsson et al. (2018), which characterize the worst-case distributions of distributionally robust vehicle routing problems where the customer locations are unknown i.i.d. realizations from a distribution that is specified either through moment conditions or the Wasserstein distance to a reference distribution. Finally, Hoogeboom et al. (2021) consider a variant of the CVRP where both routes and time window assignments need to be determined such that the expected travel times and the risk of violating the time windows are minimized simultaneously.

The remainder of the paper unfolds as follows. After defining the problem of interest in Section 2, Sections 3 and 4 investigate which VRP variants are amenable to a solution with a branch-and-cut scheme based on rounded capacity inequalities. Section 5 specializes our findings to the CVRP under uncertainty, Section 6 reports numerical results, and we offer concluding remarks in Section 7.

Notation. We denote by $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ the sets of real numbers, non-negative real numbers as well as strictly positive real numbers, respectively. For $S \subseteq \{1, \ldots, n\}$ we denote by $\mathbf{1}_S \in \{0, 1\}^n$ the vector that satisfies $(\mathbf{1}_S)_i = 1$ if $i \in S$ and $(\mathbf{1}_S)_i = 0$ otherwise. Moreover, $\mathbf{e}$ is the vector of all ones, and $\mathbf{e}_i$ is the i-th canonical basis vector; in both cases, the dimension will be clear from the context. The p-norm of a vector, $p \geqslant 1$, is denoted by $\|\cdot\|_p$, and we use $\|\cdot\|_\infty$ to denote the infinity (maximum) norm. We denote by $[\cdot]_+ = \max\{\cdot, 0\}$ the non-negative part of a scalar, which we also apply to vectors in a component-wise fashion.

2 Problem Formulation

Consider a complete, directed and weighted graph $G = (V, A, c)$ with nodes $V = \{0, \ldots, n\}$, arcs $A = \{(i, j) \in V \times V : i \neq j\}$ and transportation costs $c : A \to \mathbb{R}_+$. Here, 0 is the depot node and $V_C = \{1, \ldots, n\}$ represents the set of customer nodes. The vehicle routing problem we wish to study asks for a cost optimal route plan for a set $K = \{1, 2, \ldots, m\}$ of vehicles starting and ending at the depot node 0 such that a given set of constraints is met. Firstly, we require the route plan to form an m-partition of the customer set V_C, that is, the route plan $\mathbf{R}$ has to belong to the set

$$\mathfrak{P}(V_C, m) = \Bigg\{ \mathbf{R} = \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m\} : \boldsymbol{R}_k = (R_{k,1}, \ldots, R_{k,n_k}) \text{ with } n_k \geqslant 1 \text{ and } R_{k,i} \in V_C \;\; \forall k, i, \\ R_{k,i} \neq R_{l,j} \;\; \forall (k, i) \neq (l, j), \;\; \bigcup_{k \in K} \boldsymbol{R}_k = V_C \Bigg\}.$$

Each route plan $\mathbf{R}$ is a set of m routes $\boldsymbol{R}_k$, which are themselves nonempty ordered lists of customers that the vehicles visit sequentially. Here and in the following, we apply set operations to lists whenever their interpretation is clear. In particular, intersections and unions of ordered lists are interpreted as the application of the respective operators on the sets formed from the involved lists.

In addition to the aforementioned partition requirement, we assume that each route $\boldsymbol{R}_k$ of the route plan $\mathbf{R}$ has to satisfy some (technological, economic, ecological, quality-related or other) *intra-route* constraints (Irnich et al., 2014, §1.3.3), which we describe by the set

$$\mathcal{C} \subseteq \{\boldsymbol{R} = (R_1, \ldots, R_\nu) : \nu \geqslant 1 \text{ and } R_i \in V_C \;\; \forall i = 1, \ldots, \nu\}.$$

To be feasible, a route plan has to reside in the set $\mathfrak{P}(V_C, m) \cap \mathcal{C}_m$, where $\mathcal{C}_m = \{\mathbf{R} = \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m\} : \boldsymbol{R}_k \in \mathcal{C} \;\; \forall k\}$. Note that we do not consider *inter-route* (or global) constraints (Irnich et al., 2014, §1.3.5) that tie the feasibility of a route to the characteristics of other routes (as is the case, *e.g.*, in the presence of globally constrained resources or fairness considerations).

With the above notation, we are interested in solving the problem

$$\begin{aligned} &\text{minimize} && \sum_{k \in K} \sum_{l=0}^{n_k} c(R_{k,l}, R_{k,l+1}) \\ &\text{subject to} && \mathbf{R} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m. \end{aligned} \qquad (\text{VRP}(\mathcal{C}))$$

Here we use the convention that $R_{k,0} = R_{k,n_k+1} = 0$, which ensures that each vehicle starts and ends at the depot. To avoid trivially infeasible problem instances, we assume throughout the paper that for all customers $i \in V_C$, there is $\boldsymbol{R} \in \mathcal{C}$ such that $i \in \boldsymbol{R}$. In other words, every customer's demand can principally be served by a single vehicle in every problem instance.

For the results in this paper, we will typically impose the following two assumptions:

(D) $\mathcal{C}$ is *downward closed*, that is, if $\boldsymbol{R} \in \mathcal{C}$ for $\boldsymbol{R} = (R_1, \ldots, R_\nu)$, then $\boldsymbol{S} \in \mathcal{C}$ for all $\boldsymbol{S} = (R_{i_1}, \ldots, R_{i_\sigma})$ with $1 \leqslant \sigma \leqslant \nu$ and $1 \leqslant i_1 < i_2 < \ldots < i_\sigma \leqslant \nu$.

(P) $\mathcal{C}$ is *permutation invariant*, that is, if $\boldsymbol{R} \in \mathcal{C}$ then $\boldsymbol{S} \in \mathcal{C}$ for all permutations $\boldsymbol{S}$ of $\boldsymbol{R}$.

Condition **(D)** implies that we cannot model problems that disallow routes in which vehicles serve "too few" customers, since a subset of the customers of a feasible route can always be served as well (modulo the requirement imposed by $\mathfrak{P}(V_C, m)$ that the omitted customers need to be served by the other vehicles). Condition **(P)** implies that the order of customers within a route does not matter for its feasibility (but it will normally still matter in terms of its optimality). Here and in the following, we say that a set S is contained in a set of lists $\mathcal{S}$ if and only if every permutation of S, expressed as a list, is contained in $\mathcal{S}$. Thus, condition **(P)** is equivalent to requiring that $\boldsymbol{S} \in \mathcal{C}$ only if $S \in \mathcal{C}$ for the set S formed from the elements of $\boldsymbol{S}$.

Example 1 (Instances of VRP($\mathcal{C}$))**.** VRP($\mathcal{C}$) *recovers the* classical CVRP *if we set*

$$\mathcal{C} = \left\{ \boldsymbol{R} = (R_1, \ldots, R_\nu) \;:\; \sum_{i \in \boldsymbol{R}} q_i \leqslant Q \right\}, \tag{1}$$

where q_i is the demand of customer i and Q is the capacity of each vehicle. The set $\mathcal{C}$ satisfies **(P)** *by definition, and it satisfies* **(D)** *whenever the customer demands $\boldsymbol{q}$ are nonnegative. More generally, we obtain a variant of the* VRP with compartments *if we set*

$$\mathcal{C} = \left\{ \boldsymbol{R} = (R_1, \ldots, R_\nu) \;:\; \sum_{i \in \boldsymbol{R}} q_{ip} \leqslant Q_p \;\; \forall p = 1, \ldots, P \right\}, \tag{2}$$

where q_{ip} now denotes the demand of customer i for space in compartment p and Q_p is the capacity of compartment p in each vehicle. Again, both **(D)** *and* **(P)** *are satisfied as long as $\boldsymbol{q}$ is nonnegative.*

We recover the chance constrained CVRP *if we set*

$$\mathcal{C} = \left\{ \boldsymbol{R} = (R_1, \ldots, R_\nu) \;:\; \mathbb{P}\left[\sum_{i \in \boldsymbol{R}} \tilde{q}_i \leqslant Q \right] \geqslant 1 - \epsilon \right\},$$

where we assume that the customer demands $\tilde{q}_i$ are random variables that are governed by the probability distribution $\mathbb{P}$, and where ϵ is a risk threshold selected by the decision maker. Both the *chance constrained CVRP and its extension to multiple compartments satisfy the assumptions* **(D)** *and* **(P)** *as long as the customer demands satisfy* $\tilde{\boldsymbol{q}} \geqslant \mathbf{0}$ $\mathbb{P}$*-almost surely.*

While Example 1 shows that the CVRP and some of its variants satisfy the assumptions **(D)** and **(P)**, it is worth pointing out that several important VRP variants do *not* fall under our framework. The distance constrained CVRP, for example, imposes the constraints $\sum_{l=0}^{\nu} t(R_l, R_{l+1}) \leqslant T$ for some distance function t, and these constraints violate **(P)** since different permutations of the customers along a route lead to different route lengths in general. For the same reason, the CVRP with time windows, which requires each customer $i \in V_C$ to be visited at some time $t_i \in [\underline{t}_i, \bar{t}_i]$, violates **(P)**, and it additionally violates **(D)** if we do not permit idle times. As we will see in Section 5, however, the assumptions **(D)** and **(P)** are satisfied for a broad range of stochastic and distributionally robust formulations of the CVRP, which form the focus of this paper.

To solve VRP($\mathcal{C}$) numerically, we consider its reformulation as the well-known two-index vehicle flow model (Laporte and Norbert, 1983; Lysgaard et al., 2004)

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{(i,j)\in A} c(i,j)\, x_{ij} & \\
\text{subject to} & \displaystyle\sum_{\substack{j\in V:\\(i,j)\in A}} x_{ij} = \sum_{\substack{j\in V:\\(j,i)\in A}} x_{ji} = \delta_i & \forall i \in V \\
& \displaystyle\sum_{i\in V\setminus S}\sum_{j\in S} x_{ij} \geqslant d(S) & \forall \emptyset \neq S \subseteq V_C \\
& x_{ij} \in \{0,1\} & \forall (i,j) \in A,
\end{array}
\qquad (2\text{VF}(d))
$$

where $\delta_i = 1$ for $i \in V_C$ and $\delta_0 = m$. We call the function $d : 2^{V_C} \to \mathbb{R}_+$ the *demand estimator*, and the set of constraints involving d are called the *capacity constraints*. By writing the set S in regular (non-bold) font, we emphasize that S is unordered (as opposed to the ordered list $\boldsymbol{R}_k$, for example). We assume that $d(S) = 0 \Leftrightarrow S = \emptyset$. Note that the value of $d(\emptyset)$ can be chosen freely as it does not affect the formulation. Moreover, the choice $d(S) > 0$ for $S \neq \emptyset$ ensures that route plans containing short cycles are excluded from the feasible region of 2VF(d).

Solving VRP($\mathcal{C}$) via 2VF(d) enjoys several potential advantages. Firstly, mature (and open source) solvers are available to solve 2VF(d), see, *e.g.*, Lysgaard et al. (2004) and Semet et al. (2014). These algorithms introduce the capacity constraints iteratively as part of a branch-and-cut algorithm. Thus, if we can show that VRP($\mathcal{C}$) is equivalent to 2VF(d) for some demand estimator d, then we can solve VRP($\mathcal{C}$) as long as we can evaluate d efficiently. Secondly, 2VF(d) offers a unified solution framework for different problem variants where only the demand estimator d needs to be adapted. In other words, minor variations of the same branch-and-cut algorithm can be employed to solve different variants of the problem. This is an important consideration for adoption in practice, where it is unreasonable to expect that fundamentally different algorithms will be developed and maintained to solve different variants of the same problem. Finally, the capacity constraints of 2VF(d) constitute an important building block in modern branch-and-cut-and-price algorithms, and efficiently separable cuts for 2VF(d) can be applied to that algorithm class as well.

We want to investigate when VRP($\mathcal{C}$) is equivalent to 2VF(d), which is amenable to a solution via standard branch-and-cut algorithms. To this end, we first formalize our notion of equivalence.

Equivalence. VRP($\mathcal{C}$) and 2VF(d) are said to be *equivalent* whenever they satisfy:

(a) Every feasible route plan $\mathbf{R}$ in VRP($\mathcal{C}$) induces a feasible solution $\boldsymbol{x}$ in 2VF(d) via

$$x_{ij} = 1 \iff \exists k \in K,\ \exists l \in \{0, \dots, n_k\} \,:\, (i, j) = (R_{k,l}, R_{k,l+1}). \tag{3}$$

(b) Every feasible solution $\boldsymbol{x}$ in 2VF(d) induces a feasible route plan $\mathbf{R}$ in VRP($\mathcal{C}$) via (3).

Note that if VRP($\mathcal{C}$) and 2VF(d) are equivalent, then any feasible route plan $\mathbf{R}$ in VRP($\mathcal{C}$) induces a *unique* feasible solution $\boldsymbol{x}$ in 2VF(d) via (3) and vice versa. In the remainder of the paper, we refer to these unique solutions as $\boldsymbol{x}(\mathbf{R})$ and $\mathbf{R}(\boldsymbol{x})$, respectively. Note also that the objective functions of VRP($\mathcal{C}$) and 2VF(d) coincide, which justifies our notion of equivalence.

3 Equivalence of VRP($\mathcal{C}$) and 2VF(d)

We first show that the assumptions **(D)** and **(P)** are sufficient for VRP($\mathcal{C}$) and 2VF(d) to be equivalent under a range of demand estimators d, which we characterize explicitly. We then demonstrate that the assumptions **(D)** and **(P)** are tight in the sense that there are VRP($\mathcal{C}$) instances violating either assumption for which no demand estimator d results in an equivalent 2VF(d) instance.

A seemingly natural choice for the demand estimator d in 2VF(d) is

$$\overline{d}^m(S) = \inf\left\{ J \in \mathbb{N} \,:\, S \subseteq \bigcup_{k=1,\dots,J} \boldsymbol{R}_k \ \text{ for } \{\boldsymbol{R}_1, \dots, \boldsymbol{R}_J, \dots, \boldsymbol{R}_m\} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m \right\}$$

for $\varnothing \neq S \subseteq V_C$, as well as $\overline{d}^m(\varnothing) = 0$. This demand estimator records the minimum number of vehicles required to serve the customers in S in any feasible route plan $\mathbf{R} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m$.

Note that $\overline{d}^m(S) = \infty$ is possible if the problem instance is infeasible, which motivates our use of the infimum operator. The capacity constraints under the demand estimator $\overline{d}^m$ are commonly referred to as *generalized capacity constraints*. Since $\overline{d}^m$ is difficult to compute even for simple sets $\mathcal{C}$, however, it is not typically used in practice. Instead, research has focused on relaxations (*i.e.*, lower bounds) of this demand estimator that are easier to calculate while still tight enough to establish an equivalence between VRP($\mathcal{C}$) and 2VF(d). One such demand estimator is

$$\overline{d}^1(S) = \min\left\{ I \in \mathbb{N} \,:\, S \subseteq \bigcup_{k=1,\ldots,I} \boldsymbol{R}_k \text{ for some } \boldsymbol{R}_1, \ldots, \boldsymbol{R}_I \in \mathcal{C} \right\}$$

for $\emptyset \neq S \subseteq V_C$, as well as $\overline{d}^1(\emptyset) = 0$. This demand estimator determines the minimum number of vehicles required to serve the customers in $S \subseteq V_C$, but—in contrast to $\overline{d}^m$—it ignores the customers in $V_C \backslash S$. Note that $\overline{d}^1(S)$ is always finite by our earlier assumption that $i \in \boldsymbol{R}$ for some $\boldsymbol{R} \in \mathcal{C}$, $i \in V_C$, and thus the use of the minimum operator is justified. The capacity constraints under the demand estimator $\overline{d}^1$ are commonly referred to as *weak capacity constraints*. Although $\overline{d}^1$ tends to be easier to calculate than $\overline{d}^m$, its computation is still NP-hard for most commonly employed sets $\mathcal{C}$, and thus it is not normally used to identify violated capacity constraints in a branch-and-cut scheme. On the other end of the spectrum, we have the naive demand estimator

$$\underline{d}(S) = \begin{cases} 0 & \text{if } S = \emptyset, \\ 1 & \text{if } \emptyset \neq S \in \mathcal{C}, \\ 2 & \text{otherwise.} \end{cases}$$

Remember that $S \in \mathcal{C}$ if and only if $\boldsymbol{S} \in \mathcal{C}$ for every list $\boldsymbol{S}$ that can be formed from the elements of S, and under **(P)** we have $\boldsymbol{S} \in \mathcal{C}$ if and only if $S \in \mathcal{C}$. While the demand estimator $\underline{d}$ is typically easy to compute, the resulting capacity constraints are weak and thus slow down the branch-and-cut scheme significantly. In the remainder of this section, we will see that the above three demand estimators characterize the range of demand estimators under which VRP($\mathcal{C}$) and 2VF(d) are equivalent; in the next section, we will discuss two demand estimators within this range that are preferable to $\overline{d}^m$, $\overline{d}^1$ and $\underline{d}$ due to their favourable tightness-tractability trade-off.

Under the assumptions **(D)** and **(P)**, the three demand estimators form a natural order.

Proposition 1. *Assume that* **(D)** *and* **(P)** *are satisfied. Then for any* $S \subseteq V_C$*, we have*

$$\underline{d}(S) \leqslant \overline{d}^1(S) \leqslant \overline{d}^m(S).$$

Here **(D)** *and* **(P)** *are necessary and sufficient for* $\underline{d} \leqslant \overline{d}^1$*, whereas* $\overline{d}^1 \leqslant \overline{d}^m$ *holds by construction.*

It is easy to construct instances where the three demand estimators in Proposition 1 produce the same values for all $S \subseteq V_C$. The following example is inspired by Cornuejols and Harche (1993) and shows that the inequalities in Proposition 1 can also all be strict.

Example 2. *Consider the* VRP($\mathcal{C}$) *instance with* $n = 5$ *customers,* $m = 5$ *vehicles and* $\mathcal{C} = \{(1), \ldots, (5), (1,5)\}$*. For* $S = \{1, 2, 3, 5\}$*, we have* $\underline{d}(S) = 2$ *since* $S \notin \mathcal{C}$*,* $\overline{d}^1(S) = 3$ *since* S *is covered by the routes* (2)*,* (3) *and* (1, 5)*, and* $\overline{d}^m(S) = 4$ *since no route plan can serve customers* 1 *and* 5 *in the same route and at the same time utilize all* 5 *vehicles.*

The natural ordering from Proposition 1 typically ceases to hold when the assumptions **(D)** and **(P)** are violated. We now show that under the assumptions **(D)** and **(P)**, VRP($\mathcal{C}$) and 2VF(d) are equivalent *essentially* if and only if the demand estimator d satisfies $\underline{d} \leqslant d \leqslant \overline{d}^m$. We qualify this equivalence with 'essentially' as there are pathological cases in which demand estimators $d \ngeqslant \underline{d}$ also result in equivalent formulations, as we will discuss further below in Proposition 2.

Theorem 1. VRP($\mathcal{C}$) *is equivalent to* 2VF(d) *for any* d *satisfying* $\underline{d} \leqslant d \leqslant \overline{d}^m$.

Note that while the assumptions **(D)** and **(P)** are not required for the statement of Theorem 1, they are typically required for the function interval $[\underline{d}, \overline{d}^m]$ to be nonempty (*cf.* Proposition 1).

Proposition 2. *Fix any feasible* VRP($\mathcal{C}$) *instance satisfying* **(D)** *and* **(P)**.

(i) If $\mathfrak{P}(V_C, m) \subseteq \mathcal{C}_m$ *and* $d \leqslant \overline{d}^m$*, then* VRP($\mathcal{C}$) *is equivalent to* 2VF(d) *even if* $d \ngeqslant \underline{d}$.

(ii) If $\mathfrak{P}(V_C, m) \nsubseteq \mathcal{C}_m$ *and* $d \leqslant \overline{d}^m$*, then there are* $d \ngeqslant \underline{d}$ *such that* VRP($\mathcal{C}$) *and* 2VF(d) *are equivalent, but there are also* $d \ngeqslant \underline{d}$ *such that* VRP($\mathcal{C}$) *and* 2VF(d) *are not equivalent.*

(iii) VRP($\mathcal{C}$) *fails to be equivalent to* 2VF(d) *for every* $d \nleqslant \overline{d}^m$.

From Theorem 1 and Proposition 2 we conclude that under the assumptions **(D)**, **(P)** and $d \geqslant \underline{d}$, the requirement $d \leqslant \overline{d}^m$ is necessary and sufficient for the equivalence of VRP($\mathcal{C}$) and 2VF(d). In contrast, under the assumptions **(D)**, **(P)** and $d \leqslant \overline{d}^m$, the requirement $d \geqslant \underline{d}$ is sufficient but not necessary for the equivalence of the two formulations.

We close this section by showing that there are VRP($\mathcal{C}$) instances violating either **(D)** or **(P)** for which no demand estimator d results in an equivalent 2VF(d) instance. This establishes that the assumptions **(D)** and **(P)** are not only sufficient, but also (in the aforementioned sense) tight.

Theorem 2. *There are* VRP($\mathcal{C}$) *instances violating either of the assumptions* **(D)** *or* **(P)** *that have no equivalent* 2VF(d) *instances.*

On the flipside, however, there are VRP($\mathcal{C}$) instances violating both **(D)** and **(P)** for which there still exist demand estimators d under which VRP($\mathcal{C}$) and 2VF(d) are equivalent.

In summary, we have shown that under **(D)** and **(P)**, we have $\underline{d} \leqslant \overline{d}^1 \leqslant \overline{d}^m$ (*cf.* Proposition 1), and any demand estimator $d \in [\underline{d}, \overline{d}^m]$ makes VRP($\mathcal{C}$) and 2VF(d) equivalent (*cf.* Theorem 1). In contrast, if a VRP($\mathcal{C}$) instance violates either **(D)** or **(P)**, then there may not be any demand estimator d that leads to an equivalent 2VF(d) formulation (*cf.* Theorem 2). In the remainder of this paper, we will focus on VRP($\mathcal{C}$) instances that satisfy both assumptions **(D)** and **(P)**.

4 Demand Estimators for 2VF(d)

In this section, we represent the intra-route constraints as

$$\mathcal{C} = \Big\{ \boldsymbol{R} = (R_1, \ldots, R_\nu) \,:\, \nu \geqslant 1 \text{ and } R_i \in V_C \;\; \forall i, \;\; \varphi\left(\mathbf{1}_{\boldsymbol{R}}\right) \leqslant B \Big\}, \tag{4}$$

where $\varphi : [0,1]^n \to \mathbb{R}$. To recover the classical CVRP, for example, we can choose $\varphi(\boldsymbol{x}) = \sum_{i \in V_C} q_i x_i$ and $B = Q$. Note that any class of intra-route constraints from Section 2 that satisfies **(P)** admits a representation of the form (4), for example by selecting $B = 0$ and $\varphi(\boldsymbol{x}) = 0$ if $\boldsymbol{x} = \mathbf{1}_{\boldsymbol{R}}$ for some $\boldsymbol{R} \in \mathcal{C}$, $\varphi(\boldsymbol{x}) = 1$ otherwise. However, we will be particularly interested in sets $\mathcal{C}$ and functions φ that satisfy certain properties. First and foremost, the assumptions **(D)** and **(P)** should be satisfied in order to ensure that the VRP($\mathcal{C}$) instance has an equivalent 2VF(d) instance.

Proposition 3. *A* VRP($\mathcal{C}$) *instance with intra-route constraints expressible in the form of* (4) *satisfies* **(P)** *by construction, and it satisfies* **(D)** *whenever* φ *is monotone.*

Recall that φ is monotone if $\varphi(\boldsymbol{x}) \leqslant \varphi(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ satisfying $\boldsymbol{x} \leqslant \boldsymbol{y}$.

We now consider two demand estimators that turn out to be of special interest due to their tractability as well as their versatility. The *summation demand estimator* d^{S} is defined as

$$d^{\mathrm{S}}(S) = \max\left\{1, \lceil \varphi\left(\mathbf{1}_S\right)/B \rceil\right\} \qquad \forall \varnothing \neq S \subseteq V_C,$$

as well as $d^{\mathrm{S}}(\varnothing) = 0$. For the classical CVRP with $\varphi(\boldsymbol{x}) = \sum_{i \in V_C} q_i x_i$ and $B = Q$, the use of the summation demand estimator d^{S} in 2VF(d) reduces to the well-known *rounded capacity inequalities.*

Recently, Ghosal and Wiesemann (2020) have used d^{S} with $\varphi(\boldsymbol{x}) = \text{WC-VaR}(\tilde{\boldsymbol{q}}^\top \boldsymbol{x})$, the worst-case value-at-risk of the customer demands, to solve a 2VF(d) formulation of the distributionally robust chance constrained CVRP. The *packing demand estimator* d^{P} is defined as

$$d^{\mathrm{P}}(S) = \min\left\{I \in \mathbb{N} : \exists \boldsymbol{X} \in [0,1]^{n \times I} \text{ such that } \boldsymbol{X}\mathbf{e} = \mathbf{1}_S,\ \ \varphi(\boldsymbol{x}_k) \leqslant B\ \ \forall k = 1, \ldots, I\right\}$$

for all $\varnothing \neq S \subseteq V_C$, as well as $d^{\mathrm{P}}(\varnothing) = 0$. Here, $\boldsymbol{x}_k \in \mathbb{R}^n$ is the k-th column of the matrix $\boldsymbol{X}$, $k = 1, \ldots, I$. To our best knowledge, the packing demand estimator d^{P} has not been studied previously. It can be interpreted as the optimal value of a fractional bin packing problem; this interpretation is formalized in the following proposition.

Proposition 4. *If φ is monotone and we restrict ourselves to binary assignment matrices $\boldsymbol{X} \in \{0,1\}^{n \times I}$ in d^{P}, then d^{P} coincides with the demand estimator $\overline{d}^1$ defined in Section 3.*

The evaluation of the packing demand estimator d^{P} requires the solution of an assignment problem, which can become computationally prohibitive if d^{P} has to be evaluated frequently. It turns out, however, that d^{P} admits a closed-form solution when φ is convex.

Proposition 5. *If φ is convex, then the packing demand estimator d^{P} evaluates to*

$$d^{\mathrm{P}}(S) = \min\left\{I \in \mathbb{N} : \varphi(\mathbf{1}_S / I) \leqslant B\right\} \qquad \forall \varnothing \neq S \subseteq V_C.$$

One can construct counterexamples which show that the statement of Proposition 5 ceases to hold when φ is not convex. In summary, the summation demand estimator d^{S} requires a single evaluation of φ. Assuming that φ is convex, the packing demand estimator d^{P} requires $\mathcal{O}(\log m)$ evaluations of φ since the the minimizer $I^\star$ in Proposition 5 can be determined via a binary (if φ is also monotone) or trisection search. Thus, both demand estimators can be computed efficiently whenever φ allows for an efficient evaluation. As we will see in the next section, this is the case for a broad range of CVRP variants under stochastic and distributionally robust descriptions of the uncertainty governing the customer demands.

We now study the applicability of the two demand estimators d^{S} and d^{P}.

Theorem 3. *Assume that φ is monotone.*

(i) If φ is subadditive, *we have $\underline{d} \leqslant d^{\mathrm{S}} \leqslant d^{\mathrm{P}} \leqslant \overline{d}^m$. If φ is also* positive homogeneous *then $d^{\mathrm{S}} = d^{\mathrm{P}}$; otherwise, $d^{\mathrm{S}} = d^{\mathrm{P}}$ does not hold in general.*

(ii) If φ is additive*, we have $\underline{d} \leqslant d^{\mathrm{S}} = d^{\mathrm{P}} \leqslant \overline{d}^m$. Furthermore,* VRP($\mathcal{C}$) *can be reformulated as a deterministic CVRP instance, and every deterministic CVRP instance can be reformulated as a* VRP($\mathcal{C}$) *instance with additive φ.*

(iii) If φ is not subadditive*, then $\underline{d} \leqslant d^{\mathrm{P}} \leqslant \overline{d}^m$, whereas $d^{\mathrm{S}} \leqslant \overline{d}^m$ does not hold in general.*

Recall that φ is subadditive whenever $\varphi(\boldsymbol{x} + \boldsymbol{y}) \leqslant \varphi(\boldsymbol{x}) + \varphi(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ satisfying $\boldsymbol{x} + \boldsymbol{y} \in [0,1]^n$, and that φ is additive if the inequality holds as equality. Likewise, φ is positive homogeneous if $\varphi(\lambda \boldsymbol{x}) = \lambda \varphi(\boldsymbol{x})$ for all $\lambda > 0$ and all $\boldsymbol{x} \in [0,1]^n$ satisfying $\lambda \boldsymbol{x} \in [0,1]^n$. A subadditive and positive homogeneous function is also called sublinear.

Theorem 3 shows that for a subadditive and positive homogeneous function φ, the summation and packing demand estimators coincide, and we should use the summation demand estimator due to its favorable complexity. We will see in the next section that examples of subadditive and positive homogeneous φ include all coherent risk measures (such as the conditional value-at-risk and expectile risk measures) as well as the underperformance risk index under a stochastic as well as a distributionally robust description of the uncertainty. If, on the other hand, φ is subadditive but not positive homogeneous, then the packing demand estimator can result in tighter capacity constraints. An example of a subadditive function φ that fails to be positive homogeneous is the ramp disutility function (discussed in the next section) under a stochastic as well as a distributionally robust description of the uncertainty. Figure 1 (left) illustrates how the packing demand estimator can yield tighter capacity constraints than the summation demand estimator for this risk measure. An example of an additive function φ is the expected loss over a stochastic description of the uncertainty. Examples of functions φ that fail to be subadditive include, as the next section shows, the expected disutility, entropic risk measures, the essential riskiness index, the service fulfillment risk index and the requirements violation index. Figure 1 (right) illustrates that in this case, we have to use the packing demand estimator as the summation demand estimator may fall outside the interval $[\underline{d}, \overline{d}^m]$ and thus cut off feasible route plans. Since all of the aforementioned risk measures are convex, the packing demand estimator can be computed efficiently for all of them.

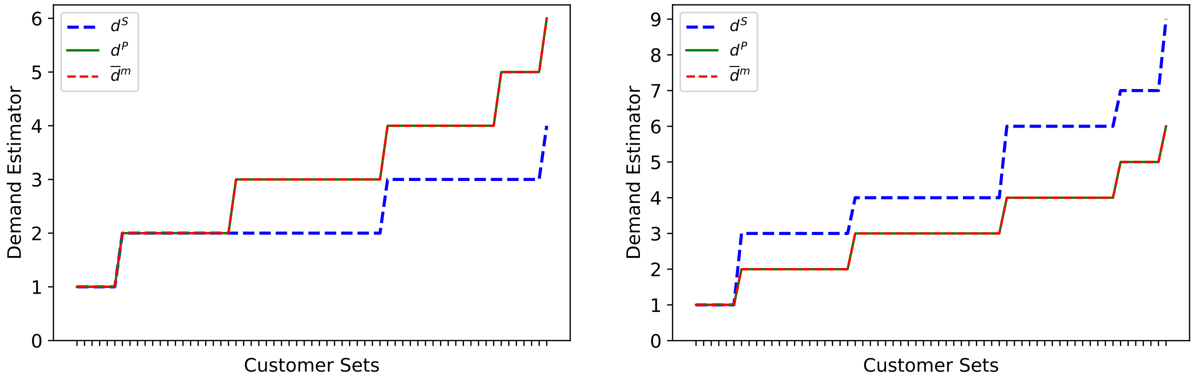

Figure 1. Demand estimators for a stochastic CVRP instance with $n = 6$ customers and the demand distribution $\mathbb{P}[\tilde{\boldsymbol{q}} = 5\mathbf{e}] = 0.05$, $\mathbb{P}[\tilde{\boldsymbol{q}} = 16\mathbf{e}] = 0.9$ and $\mathbb{P}[\tilde{\boldsymbol{q}} = 30\mathbf{e}] = 0.05$ using (a) the expected ramp disutility $\mathbb{E}_{\mathbb{P}}[\max\{\sum_{i \in S} \tilde{q}_i, 30\}]$ with $B = 30$ (left) and (b) the entropic risk $10 \log \mathbb{E}_{\mathbb{P}}[\exp(0.1 \sum_{i \in S} \tilde{q}_i)]$ with $B = 17.2$ (right).

5 VRP($\mathcal{C}$) under Risk and Ambiguity

From now on, we focus on the intra-route constraints of the distributionally robust CVRP, where the uncertain customer demands $\tilde{\boldsymbol{q}}$ can be governed by any distribution $\mathbb{P}$ from the ambiguity set $\mathcal{P}$, and the feasibility of a route depends on its worst-case risk over all distributions $\mathbb{P} \in \mathcal{P}$:

$$\mathcal{C} = \left\{ \boldsymbol{R} = (R_1, \ldots, R_\nu) \;:\; \nu \geqslant 1 \text{ and } R_i \in V_C \;\; \forall i, \;\; \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\sum_{i \in \boldsymbol{R}} \tilde{q}_i \right] \leqslant Q \right\} \tag{5}$$

We call the collection $\rho = \{\rho_{\mathbb{P}}\}_{\mathbb{P} \in \mathcal{P}}$ of risk measures an *ambiguous risk measure*, and we define $\varphi(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}}\right]$ for $\boldsymbol{x} \in [0,1]^n$ as the *worst-case risk measure*. Each individual risk measure $\rho_{\mathbb{P}}$ maps scalar random variables to real numbers with the interpretation that larger numbers correspond to greater risks, and the ambiguous risk measure ρ allows us to quantify the worst-case risk over all distributions $\mathbb{P} \in \mathcal{P}$. The upper bound Q represents either the homogeneous capacity of all vehicles (if the risk measure maps to quantities that have the same unit as the customer demands, such as the worst-case expectation or the worst-case (conditional) value-at-risk) or more generally a bound on the acceptable risk (*e.g.*, if the risk measure corresponds to the expected disutility). The intra-route constraints (5) are readily recognized as a special case of the intra-route constraints (4) studied in Section 4. Note that the intra-route constraints (5) of the distributionally robust CVRP generalize those of the stochastic CVRP, which correspond to instances of (5) with singleton ambiguity sets, as well as those of the robust CVRP, which emerge if the ambiguity set $\mathcal{P}$ contains all Dirac distributions supported on a subset of $\mathbb{R}^n_+$ (the *uncertainty set*).

We assume that $\tilde{\boldsymbol{q}} \geqslant \mathbf{0}$ $\mathbb{P}$-almost surely for all $\mathbb{P} \in \mathcal{P}$ and that each individual risk measure $\rho_{\mathbb{P}}$, $\mathbb{P} \in \mathcal{P}$, is monotone. This implies that the worst-case risk measure φ is monotone, that is, $\varphi(\boldsymbol{x}) \leqslant \varphi(\boldsymbol{y})$ whenever $\boldsymbol{x} \leqslant \boldsymbol{y}$, and VRP($\mathcal{C}$) satisfies the assumptions **(D)** and **(P)** due to Proposition 3. Additionally, we will be interested in cases where the worst-case risk measure is subadditive and/or convex so that we can apply the demand estimators d^{S} and d^{P} from Section 4 to solve the corresponding instance of 2VF(d). Finally, we will be interested in worst-case risk measures that can be evaluated quickly so that the resulting 2VF(d) instances can be solved efficiently.

Throughout the remainder of the paper, we consider the scenario-wise first-order ambiguity set

$$\mathcal{P} \;=\; \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n_+ \times \mathcal{W}) : \exists \boldsymbol{s} \in \mathcal{S} \text{ such that } \begin{bmatrix} \mathbb{P}\left[\underline{\boldsymbol{q}}^w \leqslant \tilde{\boldsymbol{q}} \leqslant \overline{\boldsymbol{q}}^w \mid \tilde{w} = w\right] = 1 \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{q}} \mid \tilde{w} = w\right] = \boldsymbol{\mu}^w \\ \mathbb{E}_{\mathbb{P}}\left[|\tilde{\boldsymbol{q}} - \boldsymbol{\mu}^w| \mid \tilde{w} = w\right] \leqslant \boldsymbol{\nu}^w \\ \mathbb{P}\left[\tilde{w} = w\right] = s_w \end{bmatrix} \forall w \in \mathcal{W} \right\} \tag{6}$$

proposed by Chen et al. (2019) and Long et al. (2020). Here, $\tilde{w}$ is a random scenario supported on the set $\mathcal{W} = \{1, \ldots, W\}$, $[\underline{\boldsymbol{q}}^w, \overline{\boldsymbol{q}}^w]$ is the support of the uncertain demands $\tilde{\boldsymbol{q}}$ under scenario $w \in \mathcal{W}$, $\boldsymbol{\mu}^w \in (\underline{\boldsymbol{q}}^w, \overline{\boldsymbol{q}}^w)$ and $\boldsymbol{\nu}^w > \mathbf{0}$ represent the expectation and the mean absolute deviation of the demand vector under scenario w, respectively, $\boldsymbol{s}$ denotes the scenario probabilities that are only known to be contained in the subset $\mathcal{S}$ of the probability simplex in $\mathbb{R}^W$, and $\mathcal{P}_0(\mathbb{R}^n_+ \times \mathcal{W})$ is the set of all probability distributions supported on $\mathbb{R}^n_+ \times \mathcal{W}$. We allow for the mean and mean absolute deviation conditions to be absent in (6), in which case some of the computations considered below simplify. All of our results also extend to ambiguity sets in which the mean absolute deviation is replaced by the expectation of a piecewise affine convex function (*cf.* Long et al. 2020), which allows us to stipulate, among others, approximate upper bounds on the marginal variances or the Huber losses of the customer demands (Wiesemann et al., 2014).

As we show in the following, the ambiguity set (6) is very versatile and allows us to model a range of well-known ambiguity sets from the literature.

Example 3 (Ambiguity Set $\mathcal{P}$)**.** *The ambiguity set* (6) *recovers a* stochastic CVRP

$$\mathcal{P} \;=\; \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n_+) \,:\, \mathbb{P}\left[\boldsymbol{q} = \hat{\boldsymbol{q}}^w\right] = \hat{s}_w \;\; \forall w \in \mathcal{W} \right\}$$

if we set $\underline{\boldsymbol{q}}^w = \overline{\boldsymbol{q}}^w = \hat{\boldsymbol{q}}^w$, $w \in \mathcal{W}$, $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$ *and disregard the expectation and mean absolute deviation constraints. Likewise, we obtain a distributionally robust CVRP over the* marginalized moment ambiguity set *(Ghosal and Wiesemann, 2020)*

$$\mathcal{P} \;=\; \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n_+) \,:\, \mathbb{P}\left[\underline{\boldsymbol{q}} \leqslant \tilde{\boldsymbol{q}} \leqslant \overline{\boldsymbol{q}}\right] = 1,\;\; \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{q}}\right] = \boldsymbol{\mu},\;\; \mathbb{E}_{\mathbb{P}}\left[|\tilde{\boldsymbol{q}} - \boldsymbol{\mu}|\right] \leqslant \boldsymbol{\nu} \right\}$$

if we set $W = 1$ *and* $\mathcal{S} = \{1\}$*. We recover a distributionally robust CVRP over the* type-∞ Wasserstein ambiguity set *(Mohajerin Esfahani and Kuhn, 2018; Kuhn et al., 2019; Bertsimas et al., 2021)*

$$\mathcal{P} \;=\; \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n_+) \,:\, \mathsf{d}^{\mathrm{W}}_{\infty}\left(\mathbb{P},\, \frac{1}{W} \sum_{w \in \mathcal{W}} \delta_{\hat{\boldsymbol{q}}^w} \right) \leqslant \theta \right\},$$

where $\frac{1}{W}\sum_{w\in\mathcal{W}} \delta_{\hat{\boldsymbol{q}}^w}$ *is the empirical distribution over the historical demands* $\hat{\boldsymbol{q}}^1, \ldots, \hat{\boldsymbol{q}}^W$ *and*

$$\mathsf{d}^{\mathrm{W}}_{\infty}(\mathbb{P}, \mathbb{Q}) \;=\; \inf \left\{ \Pi\text{-ess sup} \left\| \tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\xi}}' \right\|_{\infty} \,:\, \left[\begin{array}{c} \Pi \text{ is a joint distribution over } \tilde{\boldsymbol{\xi}} \text{ and } \tilde{\boldsymbol{\xi}}' \\ \text{with marginals } \mathbb{P} \text{ and } \mathbb{Q} \end{array} \right] \right\}$$

is the type-∞ Wasserstein distance with the ∞-norm as ground metric, by setting $\underline{\boldsymbol{q}} = [\hat{\boldsymbol{q}}^w - \theta \cdot \mathbf{e}]_+$ *and* $\overline{\boldsymbol{q}} = \hat{\boldsymbol{q}}^w + \theta \cdot \mathbf{e}$ *for all* $w \in \mathcal{W}$, $\mathcal{S} = \left\{\frac{1}{W} \cdot \mathbf{e}\right\}$ *and disregarding the expectation and mean absolute deviation constraints, see Proposition 3 of Bertsimas et al. (2021). A distributionally robust CVRP over the* Kullback-Leibler (KL) divergence ambiguity set *(Bayraksan and Love, 2015, §3.1)*

$$\mathcal{P} \;=\; \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n_+) \,:\, \mathrm{supp}(\mathbb{P}) = \left\{\hat{\boldsymbol{q}}^1, \ldots, \hat{\boldsymbol{q}}^W\right\},\;\; \mathsf{d}^{\mathrm{KL}}\left(\mathbb{P},\, \frac{1}{W} \sum_{w \in \mathcal{W}} \delta_{\hat{\boldsymbol{q}}^w} \right) \leqslant \theta \right\},$$

where supp($\mathbb{P}$) *denotes the support of the distribution* $\mathbb{P}$ *and*

$$\mathsf{d}^{\mathrm{KL}}\left(\sum_{w \in \mathcal{W}} p_w \cdot \delta_{\hat{\boldsymbol{q}}^w},\; \sum_{w \in \mathcal{W}} q_w \cdot \delta_{\hat{\boldsymbol{q}}^w} \right) \;=\; \sum_{w \in \mathcal{W}} p_w \log\left(\frac{p_w}{q_w}\right)$$

is the KL divergence between two discrete distributions over the common support $\left\{\hat{\boldsymbol{q}}^1, \ldots, \hat{\boldsymbol{q}}^W\right\}$*, is recovered if we fix* $\underline{\boldsymbol{q}}^w = \overline{\boldsymbol{q}}^w = \hat{\boldsymbol{q}}^w$ *for all* $w \in \mathcal{W}$, $\mathcal{S} = \left\{\boldsymbol{s} \in \mathbb{R}^W_+ \,:\, \sum_{w\in\mathcal{W}} s_w \log\left(s_w W\right) \leqslant \theta,\; \mathbf{e}^\top \boldsymbol{s} = 1\right\}$ *and disregard the expectation and mean absolute deviation constraints. We recover a distributionally robust CVRP over the* total variation ambiguity set *(Bayraksan and Love, 2015, §3.1)*

$$\mathcal{P} \;=\; \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n_+) \,:\, \mathrm{supp}(\mathbb{P}) = \left\{\hat{\boldsymbol{q}}^1, \ldots, \hat{\boldsymbol{q}}^W\right\},\;\; \mathsf{d}^{\mathrm{TV}}\left(\mathbb{P},\, \frac{1}{W} \sum_{w \in \mathcal{W}} \delta_{\hat{\boldsymbol{q}}^w} \right) \leqslant \theta \right\},$$

where $\mathrm{supp}(\mathbb{P})$ *denotes the support of the distribution* $\mathbb{P}$ *and*

$$\mathsf{d}^{\mathrm{TV}}\left(\sum_{w\in\mathcal{W}} p_w \cdot \delta_{\hat{\boldsymbol{q}}^w},\ \sum_{w\in\mathcal{W}} q_w \cdot \delta_{\hat{\boldsymbol{q}}^w}\right) \;=\; \sum_{w\in\mathcal{W}} q_w \cdot \left|\frac{p_w}{q_w} - 1\right|$$

is the total variation between two discrete distributions over the common support $\{\hat{\boldsymbol{q}}^1, \ldots, \hat{\boldsymbol{q}}^W\}$*, finally, if we fix* $\underline{\boldsymbol{q}}^w = \overline{\boldsymbol{q}}^w = \hat{\boldsymbol{q}}^w$ *for all* $w \in \mathcal{W}$*,* $\mathcal{S} = \left\{\boldsymbol{s} \in \mathbb{R}^W_+ \,:\, \left\|\boldsymbol{s} - \frac{1}{W}\cdot \mathbf{e}\right\|_1 \leqslant \theta,\ \mathbf{e}^\top \boldsymbol{s} = 1\right\}$ *and disregard the expectation and mean absolute deviation constraints.*

Real-life logistics problems—especially those of consumer-facing businesses—tend to be large in scale, and they are typically solved in two stages: In a first stage, a districting problem partitions the overall problem into several smaller problems based on geographic proximity. Subsequently, each smaller problem—which may contain of the order of 50-100 customers—is solved as a CVRP. Even for such CVRP instances, however, a model-free characterization of the underlying probability distribution using scenarios, be it in a stochastic programming or a data-driven framework (as in the Wasserstein, KL-divergence and total variation ambiguity sets above), would require an unrealistically large number of scenarios due to the curse of dimensionality. Instead, we propose to use the scenarios $w \in \mathcal{W}$ in our ambiguity set (6) to model macroscopic effects, such as demand shocks that affect all customers in a particular region, and to characterize the residual variability of the individual customer demands using the supports $[\underline{\boldsymbol{q}}^w, \overline{\boldsymbol{q}}^w]$ and the first-order information $(\boldsymbol{\mu}^w, \boldsymbol{\nu}^w)$. The resulting instances of the ambiguity set (6) may then contain of the order of tens of scenarios, and our numerical experiments in Section 6 will show that such problems can be solved in runtimes that are comparable to those of the corresponding deterministic CVRP instances.

Section 5.1 discusses how $\mathrm{VRP}(\mathcal{C})$ with the intra-route constraints (5) can be solved via its reformulation $2\mathrm{VF}(d)$ when the set of scenario probabilities $\mathcal{S}$ in the ambiguity set (6) is a singleton, that is, when $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$. Section 5.2 extends our results to the more general setting where $\mathcal{S}$ is a convex subset of the probability simplex in $\mathbb{R}^W$. Section 5.3, finally, discussed the chance constrained CVRP, whose underlying risk measure is non-convex and thus requires a special treatment.

5.1 Ambiguity Sets With Known Scenario Probabilities

Long et al. (2020) optimize the worst-case expectation in two-stage distributionally robust optimization problems where the uncertain parameters affect the constraint right-hand sides of the second-stage problem. They show that for ambiguity sets of the form (6) with known scenario probabilities $\hat{\boldsymbol{s}}$, the worst-case expectation is attained by a discrete distribution that does not depend on the first-stage decisions, and thus the two-stage distributionally robust optimization problem reduces to a two-stage stochastic program. Our setting differs from theirs in the following aspects: *(i)* we consider the random quantity $\boldsymbol{x}^\top \tilde{\boldsymbol{q}}$ that is parametric in the weights $\boldsymbol{x}$, rather than the optimal value of a second-stage problem that is parametric in the first-stage decisions; *(ii)* we consider a broad range of risk measures, whereas Long et al. (2020) focus on the expected value, the expected disutility and the optimized certainty equivalent; and *(iii)* the random vector $\tilde{\boldsymbol{q}}$ multiplies the parameters $\boldsymbol{x}$ in our context, whereas it is isolated on the constraint right-hand sides in their setting. Nevertheless, we can adapt the arguments of Long et al. (2020) to show that the worst-case risk $\varphi(\boldsymbol{x})$ is attained by a finite demand distribution that is independent of ρ and $\boldsymbol{x}$.

Theorem 4 (Long et al. (2020))**.** *Fix an ambiguity set $\mathcal{P}$ of the form* (6) *where $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$, and assume that $\varphi(\boldsymbol{x})$ can be represented as a worst-case expectation $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \right]$ of a convex function $f : \mathbb{R} \to \mathbb{R}$. Then Algorithm 1 in Appendix B identifies a $W(2n+1)$-point worst-case distribution $\mathbb{P}^\star = \sum_{w \in \mathcal{W}} \sum_{j=1}^{2n+1} \hat{s}_w p^\star_{wj} \cdot \delta_{\boldsymbol{q}^\star_{wj}} \in \mathcal{P}$ such that $\varphi(\boldsymbol{x}) = \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})]$ for all $\boldsymbol{x} \in [0,1]^n$. Moreover, the parameters $p^\star_{wj}$ and $\boldsymbol{q}^\star_{wj}$ characterizing $\mathbb{P}^\star$ do not depend on f and $\boldsymbol{x}$.*

The intuition underlying Theorem 4 is as follows. If we condition on the event $\tilde{w} = w$, then the resulting ambiguity set $\mathcal{P}^w$ becomes rectangular in the customers $i \in V_C$ in the sense that

$$\begin{aligned} \mathcal{P}^w &= \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n_+) \ : \ \exists \mathbb{Q} \in \mathcal{P} \text{ such that } \mathbb{P}\left[\cdot\right] = \mathbb{Q}\left[\cdot \mid \tilde{w} = w\right] \right\} \\ &= \bigtimes_{i \in V_C} \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}) : \mathbb{P}\left[\underline{q}^w_i \leqslant \tilde{q}_i \leqslant \overline{q}^w_i \right] = 1, \ \ \mathbb{E}_{\mathbb{P}}\left[\tilde{q}_i\right] = \mu^w_i, \ \ \mathbb{E}_{\mathbb{P}}\left[|\tilde{q}_i - \mu^w_i| \right] \leqslant \nu^w_i \right\}. \end{aligned} \tag{7}$$

One can then verify that for a convex function f, $\sup_{\mathbb{P} \in \mathcal{P}^w} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \right]$ is attained by a distribution $\mathbb{P}^\star$ that only places positive probability on demand realizations $\boldsymbol{q} \in \bigtimes_{i \in V_C} \left\{ \underline{q}^w_i, \, \mu^w_i, \, \overline{q}^w_i \right\}$, and that these probabilities do not depend on f or $\boldsymbol{x}$. This, however, only allows us to conclude that there is a worst-case distribution with an exponentially large number 3^n of realizations. Next, fix any worst-case distribution $\mathbb{P}^\star$ supported on the demands $\boldsymbol{q} \in \bigtimes_{i \in V_C} \left\{ \underline{q}^w_i, \, \mu^w_i, \, \overline{q}^w_i \right\}$, and assume that $\mathbb{P}^\star[\boldsymbol{q}], \mathbb{P}^\star[\boldsymbol{q}'] > 0$ for an unordered pair of demands $\boldsymbol{q}$ and $\boldsymbol{q}'$, that is, $\boldsymbol{q}$ and $\boldsymbol{q}'$ satisfying neither $\boldsymbol{q} \leqslant \boldsymbol{q}'$ nor $\boldsymbol{q} \geqslant \boldsymbol{q}'$. In that case, we can move equal amounts of probability mass from the demand realizations $\boldsymbol{q}$ and $\boldsymbol{q}'$ to their join $\max\{\boldsymbol{q}, \, \boldsymbol{q}'\}$ and meet $\min\{\boldsymbol{q}, \, \boldsymbol{q}'\}$ without affecting the marginal distributions of $\mathbb{P}^\star$ and thus guaranteeing, by the rectangularity of $\mathcal{P}^w$, that the new distribution is also in $\mathcal{P}^w$. On the other hand, one can show that the new distribution has a weakly larger expected value since $f(\boldsymbol{x}^\top \max\{\boldsymbol{q}, \boldsymbol{q}'\}) + f(\boldsymbol{x}^\top \min\{\boldsymbol{q}, \boldsymbol{q}'\}) \geqslant f(\boldsymbol{x}^\top \boldsymbol{q}) + f(\boldsymbol{x}^\top \boldsymbol{q}')$. We can repeat this procedure iteratively until $\mathbb{P}^\star$ no longer places positive probability on any unordered pairs, in which case all probability is concentrated on at most $2n+1$ demand realizations. Of course, this iterative mass transportation procedure is impractical as it may require exponentially many iterations depending on the initial distribution. Instead, Algorithm 1 in Appendix B computes a worst-case probability distribution over $\mathcal{P}^w$ in $\mathcal{O}(n)$ iterations. Applying the same principle to each marginal ambiguity set $\mathcal{P}^w$, $w \in \mathcal{W}$, we obtain in $\mathcal{O}(Wn)$ iterations a $W(2n+1)$-point distribution $\mathbb{P}^\star$ that maximizes the expectation of $f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})$ over all $\mathbb{P} \in \mathcal{P}$. Since $\mathbb{P}^\star$ does not depend on $\boldsymbol{x}$, it only needs to be computed once for each 2VF(d) instance.

Theorem 4 implies that for suitable ambiguous risk measures ρ, the distributionally robust CVRP over the ambiguity set (6) with known scenario probabilities $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$ reduces to a stochastic CVRP over a probability distribution that does not depend on ρ or $\boldsymbol{x}$. Note, however, that the risk itself depends on the choice of ρ and $\boldsymbol{x}$, and hence the feasible region of the distributionally robust CVRP varies for different risk measures ρ.

In the remainder of this section, we review a number of popular risk measures, we show how their worst-case risk can be computed efficiently, and we discuss which of the demand estimators d^{S} and d^{P} can be employed in their associated reformulations 2VF(d).

Theorem 5 (Expected Disutility-Based Risk Measures)**.** *Fix an ambiguity set $\mathcal{P}$ of the form* (6) *with $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$.*

1. Expected Disutility. *The worst-case expected disutility $\varphi_{\mathrm{ED}}(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-ED}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})$ with*

$$\mathbb{P}\text{-ED}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \mathbb{E}_{\mathbb{P}}\left[U(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})\right],$$

where the disutility function U is monotonically non-decreasing and convex with $U(0) \geqslant 0$, affords a $W(2n+1)$-point worst-case distribution that can be computed with Algorithm 1 and that is independent of $\boldsymbol{x}$. Moreover, φ_{ED} is monotone, convex and not *subadditive.*

2. Essential Riskiness Index (Zhang et al., 2019). *The essential riskiness index φ_{ERI} with*

$$\varphi_{\mathrm{ERI}}(\boldsymbol{x}) \;=\; \inf\left\{\alpha \geqslant 0 \;:\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\max\left\{\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho},\; -\alpha\right\}\right] \leqslant 0\right\},$$

where $\overline{\rho}$ is the acceptable demand threshold, can be computed in time $\mathcal{O}(n^2 W + nW \log nW)$. Moreover, φ_{ERI} is monotone, convex and not *subadditive.*

3. Expectiles. *The worst-case expectile risk measure* φ_{E} *with*

$$\varphi_{\mathrm{E}}(\boldsymbol{x}) \;=\; \operatorname*{arg\,min}_{u \in \mathbb{R}} \left\{ \alpha \cdot \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - u \right]_+^2 \right] + (1-\alpha) \cdot \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left[u - \boldsymbol{x}^\top \tilde{\boldsymbol{q}} \right]_+^2 \right] \right\},$$

where $\alpha \in [1/2, 1)$*, can be computed in time* $\mathcal{O}(n^2 W + nW \log nW)$*. Moreover,* φ_{E} *is monotone, convex and subadditive.*

4. Entropic Risk. *The worst-case entropic risk* $\varphi_{\mathrm{ent}}(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-ent}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})$ *with*

$$\mathbb{P}\text{-ent}\big(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}\big) \;=\; \frac{1}{\theta} \log \mathbb{E}_{\mathbb{P}} \left[\exp\left(\theta \cdot \boldsymbol{x}^\top \tilde{\boldsymbol{q}} \right) \right],$$

where $\theta > 0$*, affords a* $W(2n+1)$*-point worst-case distribution that can be computed with Algorithm 1 and that is independent of* $\boldsymbol{x}$*. Moreover,* φ_{ent} *is monotone, convex and* not *subadditive.*

5. Requirements Violation Index (Jaillet et al., 2016) *The requirements violation index* φ_{RV} *with*

$$\varphi_{\mathrm{RV}}(\boldsymbol{x}) \;=\; \inf \left\{ \alpha \geqslant 0 \,:\, C_\alpha(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \leqslant \overline{\rho} \right\},$$

where C_α *is the worst-case certainty equivalent under an exponential disutility,*

$$C_\alpha(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) = \begin{cases} \displaystyle\sup_{\mathbb{P} \in \mathcal{P}} \alpha \log \mathbb{E}_{\mathbb{P}} \left[\exp\left(\frac{\boldsymbol{x}^\top \tilde{\boldsymbol{q}}}{\alpha} \right) \right] & \text{if } \alpha > 0 \\ \displaystyle\lim_{\gamma \to 0} C_\gamma(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) & \text{if } \alpha = 0, \end{cases}$$

and $\overline{\rho}$ *is the acceptable demand threshold, can be computed to* ϵ*-accuracy via bisection search. Moreover,* φ_{RV} *is monotone, convex and* not *subadditive.*

Since the worst-case expectiles are subadditive as well as positive homogeneous (Bellini and Bignozzi, 2015, Theorem 4.9(b)), Theorem 3 *(i)* implies that both demand estimators d^{S} and d^{P} can be applied and yield the same results. We thus prefer d^{S} for its ease of computation. In contrast, the other risk measures of Theorem 5 fail to be subadditive, and Theorem 3 *(iii)* implies that we have to use the demand estimator d^{P}. Fortunately, since all of these risk measures are convex, d^{P} can be computed efficiently thanks to Proposition 5.

Two commonly used risk measures are variants of the worst-case expected disutility. The worst-case *expected demand* $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\cdot]$ emerges as a special case of the worst-case expected disutility if we set $U(x) = x$. Since the worst-case distribution $\mathbb{P}^\star$ from Theorem 4 does not depend on $\boldsymbol{x}$ and the expectation $\mathbb{E}_{\mathbb{P}^\star}[\boldsymbol{x}^\top \tilde{\boldsymbol{q}}]$ is linear in $\boldsymbol{x}$, Theorem 3 *(ii)* implies that the corresponding worst-case distributionally robust CVRP instance reduces to a deterministic CVRP. The worst-case *expected ramp disutility* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[\max\{\cdot,\, \tau\}]$, where $\tau \in \mathbb{R}_+$ is a parameter, is monotone, subadditive and *not* positive homogeneous. While both demand estimators d^{S} and d^{P} are applicable in this case, we prefer to use d^{P} as it can offer tighter bounds, see Theorem 3 *(i)* as well as Figure 1 (left). The convexity of the expected ramp disutility allows us to evaluate d^{P} efficiently.

Theorem 6 (CVaR-Based Risk Measures)**.** *Fix an ambiguity set $\mathcal{P}$ of the form* (6) *with $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$.*

1. Conditional Value-at-Risk. *The worst-case conditional value-at-risk (CVaR) at level $1-\epsilon$, $\varphi_{\mathrm{CVaR}}(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_{1-\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})$ with*

$$\mathbb{P}\text{-CVaR}_{1-\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \inf_{u\in\mathbb{R}}\; u + \frac{1}{1-\epsilon}\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - u\right]_+ ,$$

   where $\epsilon \in [0,1)$, can be computed in time $\mathcal{O}(n^2W + nW\log nW)$. Moreover, φ_{CVaR} is monotone, convex and subadditive.

2. Service Fulfillment Risk Index (Zhang et al., 2021). *The service fulfillment risk index φ_{SRI} with*

$$\varphi_{\mathrm{SRI}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \inf\left\{\alpha \geqslant 0 \,:\, \varphi_{\mathrm{CVaR}}\left(\max\left\{\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho},\, -\alpha\right\}\right) \leqslant 0\right\},$$

   where $\overline{\rho}$ is the acceptable demand threshold and the worst-case CVaR is evaluated at level $1-\gamma$ with γ being the the service level, can be computed in time $\mathcal{O}(n^2W + nW\log nW)$. Moreover, φ_{SRI} is monotone, convex and not *subadditive.*

Since the worst-case CVaR is subadditive and positive homogeneous (Rockafeller and Uryasev, 2002, Corollary 12), Theorem 3 *(i)* implies that we can use either d^{P} or d^{S}, and the values of both demand estimators coincide. We thus prefer d^{S} as it is easier to evaluate. In contrast, the service fulfilment risk index is not subadditive, and Theorem 3 *(iii)* implies that we have to use d^{P}. Fortunately, d^{P} can be evaluated efficiently since the service fulfilment risk index is convex.

Theorem 7 (Other Risk Measures)**.** *Fix an ambiguity set $\mathcal{P}$ of the form* (6) *with $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$. The* underperformance risk index φ_{URI} *(Hall et al., 2015) with*

$$\varphi_{\mathrm{URI}}(\boldsymbol{x}) \;=\; \inf\left\{\frac{1}{\alpha} \,:\, \sup_{\mathbb{P}\in\mathcal{P}} \psi_{\mathbb{P}}\left(\alpha\left(\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho}\right)\right) \leqslant 0,\;\; \alpha > 0\right\},$$

where $\psi_{\mathbb{P}}$ is a monotone, translation invariant and convex risk measure satisfying $\psi_{\mathbb{P}}(0) = 0$ that can be expressed as the expectation of a convex function and $\overline{\rho}$ is the acceptable demand threshold, can be evaluated to ϵ-accuracy via bisection search. Moreover, φ_{URI} is monotone, convex and subadditive.

Since the underperformance risk index in Theorem 7 is subadditive and positive homogeneous (*cf.* Hall et al. 2015, Definition 3), Theorem 3 *(i)* implies that d^{S} and d^{P} are both applicable and yield the same results. We thus prefer d^{S} for its ease of computation.

One readily verifies that when the expectation and mean absolute deviation conditions in the ambiguity set (6) are absent, all the worst-case risk measures in this section are optimized by the W-point distribution $\mathbb{P}^{\star} = \sum_{w \in \mathcal{W}} \hat{s}_w \cdot \delta_{\overline{\boldsymbol{q}}^w}$ under which the customer-wise largest demands are attained almost surely in each scenario $w \in \mathcal{W}$. Since this worst-case distribution is supported on W instead of $W(2n+1)$ points (*cf.* Theorem 4), the computational complexity of evaluating the worst-case risks in Theorems 5–7 reduces accordingly in this case.

Remark 1 (Incremental Evaluation of Risk Measures)**.** *In branch-and-cut implementations, $\varphi(\boldsymbol{x})$ rarely needs to be computed from scratch; instead, it is computed iteratively for vectors $\boldsymbol{x}$ that differ in one or a few components. In this case, incremental evaluations of the worst-case risks in Theorems 5–7 can reduce the runtime for the cut evaluation by up to a factor of n.*

5.2 Ambiguity Sets With Ambiguous Scenario Probabilities

We now consider a more general setting where the set $\mathcal{S}$ of scenario probabilities in the ambiguity set (6) is one of the following convex subsets of the probability simplex $\Delta_W = \{\boldsymbol{s} \in \mathbb{R}^W_+ \, : \, \mathbf{e}^\top \boldsymbol{s} = 1\}$:

1. *1-Norm Ambiguity Set.* $\mathcal{S} = \{\boldsymbol{s} \in \Delta_W \, : \, \|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_1 \leqslant \theta\}$ with $\theta \in \mathbb{R}_+$ and $\hat{\boldsymbol{s}} \in \Delta_W$.

2. *∞-Norm Ambiguity Set.* $\mathcal{S} = \{\boldsymbol{s} \in \Delta_W \, : \, \|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_\infty \leqslant \theta\}$ with $\theta \in \mathbb{R}_+$ and $\hat{\boldsymbol{s}} \in \Delta_W$.

3. *Ellipsoidal Ambiguity Set.* $\mathcal{S} = \{\boldsymbol{s} \in \Delta_W \, : \, (\boldsymbol{s} - \hat{\boldsymbol{s}})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{s} - \hat{\boldsymbol{s}}) \leqslant \theta\}$ with $\theta \in \mathbb{R}_{++}$, $\boldsymbol{\Sigma} \succ \mathbf{0}$ and $\hat{\boldsymbol{s}} \in \Delta_W$.

4. *Axis-Parallel Ellipsoidal Ambiguity Set.* Ellipsoidal ambiguity set with $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_W)$.

5. *Entropy Ambiguity Set.* $\mathcal{S} = \{\boldsymbol{s} \in \Delta_W \, : \, \sum_{w \in \mathcal{W}} s_w \log(s_w / \hat{s}_w) \leqslant \theta\}$ with $\theta \in \mathbb{R}_{++}$ and $\hat{\boldsymbol{s}} \in \Delta_W$.

Norm-based and ellipsoidal ambiguity sets are used extensively to characterize the scenario probabilities in robust Markov decision processes (Iyengar, 2005; Nilim and El Ghaoui, 2005; Wiesemann et al., 2013) and, more broadly, distributionally robust optimization (Erdoğan and Iyengar, 2006; Zhu and Fukushima, 2009). Norm-based and entropy-based ambiguity sets are frequently used to characterize distances between probability distributions in data-driven optimization (Ben-Tal et al., 2013; Bayraksan and Love, 2015). Indeed, the total variation and KL divergence ambiguity sets from Example 3 are special cases of the 1-norm and the entropy ambiguity sets, respectively, if we set $\underline{\boldsymbol{q}}^w = \overline{\boldsymbol{q}}^w = \hat{\boldsymbol{q}}^w$, $w \in \mathcal{W}$, $\hat{\boldsymbol{s}} = \frac{1}{W} \cdot \mathbf{e}$ and disregard the expectation and mean absolute deviation constraints. Finally, ellipsoidal uncertainty sets are used characterize the uncertain customer demands (as opposed to their probabilities) in the classical robust CVRP, see Gounaris et al. (2013), Subramanyam et al. (2020) and Wang et al. (2021).

Proposition 6. *Let $\rho = \{\rho_{\mathbb{P}}\}_{\mathbb{P} \in \mathcal{P}}$ be an ambiguous risk measure whose worst-case risk φ satisfies*

$$\varphi(\boldsymbol{x}) \;=\; \max_{\boldsymbol{s} \in \mathcal{S}} \Big/ \min_{\boldsymbol{s} \in \mathcal{S}} \; f\left(\boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})\right) \qquad \forall \boldsymbol{x} \in [0,1]^n$$

for $f : \mathbb{R} \to \mathbb{R}$ monotonically increasing and $\{\boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}) : \boldsymbol{s} \in \mathcal{S}\} \subseteq dom(f)$. If $\boldsymbol{\pi}(\boldsymbol{x}) \in \mathbb{R}^W$ is known, then $\varphi(\boldsymbol{x})$ can be computed:

1. *in time $\mathcal{O}(W \log W)$ for the 1-norm and the ∞-norm ambiguity set;*
2. *to ϵ-accuracy in time $\mathcal{O}(W \log[\overline{\pi}/\epsilon])$ for the entropy ambiguity set, where $\overline{\pi} = \max\{\pi_w(\boldsymbol{x}) : w \in \mathcal{W}\}$;*
3. *to ϵ-accuracy in time $\mathcal{O}(W \log W \cdot \log \epsilon^{-1})$ for the axis-parallel ellipsoidal ambiguity set;*
4. *to ϵ-accuracy in polynomial time via the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) for the ellipsoidal ambiguity set.*

Corollary 1. *The expected disutility and the entropic risk satisfy the conditions of Proposition 6.*

The proof of Corollary 1 shows that the components $\pi_w(\boldsymbol{x})$ in Proposition 6 relate to the worst-case risk over the marginal ambiguity sets $\mathcal{P}^w$, $w \in \mathcal{W}$, and those quantities can be computed from Theorems 5–7. In practical applications (*cf.* Section 6) we expect the number W of scenarios to be small, in which case the overhead caused by the incorporation of ambiguous scenario probabilities can be considered to be a constant factor for the expected disutility and the entropic risk.

Proposition 7. *Let $\rho = \{\rho_{\mathbb{P}}\}_{\mathbb{P}\in\mathcal{P}}$ be an ambiguous risk measure whose worst-case risk φ satisfies*

$$\varphi(\boldsymbol{x}) \;=\; \inf_{u\in\mathcal{U}} \Big/ \operatorname*{arg\,min}_{u\in\mathcal{U}} \left\{ f_0(u) + \sum_{\ell=1}^{L} \max_{\boldsymbol{s}\in\mathcal{S}} f_\ell\left(\boldsymbol{s}^\top \boldsymbol{\pi}_\ell(\boldsymbol{x},u),u\right) \right\} \qquad \forall \boldsymbol{x}\in[0,1]^n$$

or

$$\varphi(\boldsymbol{x}) \;=\; \inf_{u\in\mathcal{U}} \Big/ \operatorname*{arg\,min}_{u\in\mathcal{U}} \left\{ g(u) \;:\; f_0(u) + \sum_{\ell=1}^{L} \max_{\boldsymbol{s}\in\mathcal{S}} f_\ell\left(\boldsymbol{s}^\top \boldsymbol{\pi}_\ell(\boldsymbol{x},u),u\right) \leqslant 0 \right\} \qquad \forall \boldsymbol{x}\in[0,1]^n,$$

where $\mathcal{U}\subseteq\mathbb{R}$ is a left-bounded, right-bounded or unbounded interval, $u\mapsto g(u)$ is a monotonic mapping, and $u\mapsto f_0(u)$ and $u\mapsto f_\ell(\boldsymbol{s}^\top\boldsymbol{\pi}_\ell(\boldsymbol{x},u),u)$, $\ell=1,\ldots,L$, are convex mappings. Assume further that the embedded maximization problems over $\boldsymbol{s}\in\mathcal{S}$ can be solved in time $\mathcal{O}(T)$ for any fixed values of $\boldsymbol{x}$ and u. Then $\varphi(\boldsymbol{x})$ can be computed to ϵ-accuracy in time $\mathcal{O}(LT\log\epsilon^{-1})$.

Sufficient conditions for $u\mapsto f_\ell(\boldsymbol{s}^\top\boldsymbol{\pi}_\ell(\boldsymbol{x},u),u)$ to be convex are that *(i)* each f_ℓ is jointly convex and non-decreasing and π_ℓ is convex in u; *(ii)* each f_ℓ is jointly convex and non-increasing and π_ℓ is concave in u, see (Boyd and Vandenberghe, 2004, Page 86).

Corollary 2. *The essential riskiness index, the expectiles, the requirements violation index, the CVaR, the service fulfilment index and the underperformance risk index satisfy the conditions of Proposition 7.*

For the risk measures considered in Corollary 2, the representation in Proposition 7 satisfies $L\in\{1,2\}$. The computational overhead caused by the incorporation of ambiguous scenario probabilities thus amounts to a multiplicative factor of $\mathcal{O}(\log\epsilon^{-1})$ in the compuation of the demand estimators.

5.3 The Chance Constrained CVRP

In this section, we consider the ambiguous chance constrained CVRP with technology sets

$$\mathcal{C}_{\text{CC}} = \left\{ \boldsymbol{R} = (R_1,\ldots,R_\nu) \;:\; \nu\geqslant 1 \text{ and } R_i\in V_C \;\; \forall i,\;\; \mathbb{P}\left[\sum_{i\in\boldsymbol{R}} \tilde{q}_i \leqslant B\right] \geqslant 1-\epsilon \;\; \forall\mathbb{P}\in\mathcal{P} \right\},$$

where the ambiguity set $\mathcal{P}$ is of the form (6), B denotes the vehicles' capacities and $\epsilon\in(0,1)$ is a risk threshold selected by the decision maker. Throughout the following, we assume that $\epsilon < 1/m$.

Observation 1. *For any $S\subseteq V_C$, we have*

$$\mathbb{P}\left[\sum_{i\in\boldsymbol{R}} \tilde{q}_i \leqslant B\right] \geqslant 1-\epsilon \;\; \forall\mathbb{P}\in\mathcal{P} \quad\iff\quad \varphi_{\text{VaR}}(\mathbf{1}_{\boldsymbol{R}}) \leqslant B,$$

where $\varphi_{\text{VaR}}(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}\left[\boldsymbol{x}^\top\tilde{\boldsymbol{q}}\right]$ and $\mathbb{P}\text{-VaR}_{1-\epsilon}$ denotes the $(1-\epsilon)$-value-at-risk (VaR).

Observation 1, whose statement is well known and immediately follows from the properties of the value-at-risk, allows us to use technology sets of the form (5) with $\varphi = \varphi_{\text{VaR}}$ to model the ambiguous chance constrained CVRP. Recall that in order to solve the corresponding 2VF(d) formulation, φ_{VaR} has to satisfy certain properties as outlined in Section 4. We examine this next.

Observation 2. *For an ambiguity set of the form* (6)*,* φ_{VaR} *is monotone and positive homogeneous, but it is neither subadditive nor convex in general.*

The discussion surrounding Theorem 3 implies that we cannot use the demand estimator d^{S} in conjunction with φ_{VaR}, and the demand estimator d^{P} is difficult to evaluate.

Example 4. *Consider an ambiguous chance constrained CVRP instance with* $n = 2$ *customers and* $m = 2$ *vehicles. Fix an ambiguity set of the form* (6) *with* $W = 3$*,* $\hat{\boldsymbol{s}} = \frac{1}{3}\mathbf{e}$*, expectation and mean absolute deviation constraints disregarded as well as* $\underline{\boldsymbol{q}}^1 = \overline{\boldsymbol{q}}^1 = (1,3)^\top$*,* $\underline{\boldsymbol{q}}^2 = \overline{\boldsymbol{q}}^2 = (8,3)^\top$ *and* $\underline{\boldsymbol{q}}^3 = \overline{\boldsymbol{q}}^3 = (1,11)^\top$*. For* $\epsilon = 0.3$ *and* $B = 3$*, we have* $\mathcal{C}_{\text{CC}} = \{(1),(2)\}$ *and thus* $\overline{d}^m(V_C) = 2$*, while the technology sets* (5) *with* $\varphi = \varphi_{\text{VaR}}$ *yield* $d^{\text{S}}(V_C) = \max\{1, \lceil 11/3 \rceil\} = 4 > \overline{d}^m(V_C)$*. On the other hand, note that* $\boldsymbol{Y}$ *with* $\boldsymbol{y}_1 = (1,0)^\top$ *and* $\boldsymbol{y}_2 = (0,1)^\top$ *is a feasible solution for* d^{P}*, which implies that* $d^{\text{P}}(V_C) \leqslant 2 = \overline{d}^m(V_C)$*.*

Laporte et al. (1989) use the demand estimator d^{S} in conjunction with φ_{VaR} for chance constrained CVRPs where the customer demands follow independent normal distributions. This observation has later been generalized to normally distributed customer demands (that are not necessarily independent) by Dinh et al. (2018). Theorem 3 reveals why this is possible: one readily verifies that in these specific cases, φ_{VaR} is subadditive, and part *(i)* of the theorem shows that d^{S} is indeed admissible. Ghosal and Wiesemann (2020) combine d^{S} with φ_{VaR} to solve ambiguous chance constrained CVRPs over moment ambiguity sets. Again one can show that in this special case, φ_{VaR} is subadditive, and Theorem 3 *(i)* offers a justification for the use of d^{S}. Our theory from Sections 3 and 4 thus provides a theoretical understanding why the value-at-risk is applicable under these specific circumstances. In contrast, our scenario-wise ambiguity sets (6) require a different approach, which we develop in the following.

Following Dinh et al. (2018), we define the modified VaR $\varphi_{\text{mVaR}} : [0,1]^n \to \mathbb{R}$ as

$$\varphi_{\text{mVaR}}(\boldsymbol{x}) = B \cdot \max_{k \in K} \; a(\boldsymbol{x}, k)$$

where $a(\boldsymbol{x}, 1) = 1$ and $a(\boldsymbol{x}, k) = \min\{k, \lceil \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})/B \rceil\}$ otherwise, as well as

$$\mathcal{C}_{\text{mVaR}} = \{\boldsymbol{R} = (R_1, \ldots, R_\nu) \,:\, \nu \geqslant 1 \text{ and } R_i \in V_C \ \ \forall i, \ \ \varphi_{\text{mVaR}}(\boldsymbol{x}) \leqslant B\}\,.$$

Although we have $\varphi_{\text{mVaR}} \neq \varphi_{\text{VaR}}$, it turns out that both worst-case risk measures lead to the same technology sets.

Proposition 8. *We have* $\mathcal{C}_{\text{mVaR}} = \mathcal{C}_{\text{CC}}$.

In contrast to φ_{VaR}, however, φ_{mVaR} has desirable features in view of our demand estimators.

Proposition 9. *The function* φ_{mVaR} *is monotone and subadditive, but it is neither convex nor positive homogeneous in general.*

The monotonicity of φ_{mVaR} guarantees via Proposition 3 that the technology sets (5) with $\varphi = \varphi_{\text{mVaR}}$ satisfy **(D)** and **(P)**. Since φ_{mVaR} is subadditive but not positive homogeneous, Theorem 3 implies that $d^{\text{S}} \leqslant d^{\text{P}}$ and that $d^{\text{S}} = d^{\text{P}}$ does not hold in general. However, d^{P} appears difficult to evaluate due to the non-convexity of φ_{mVaR}, and we thus prefer to use d^{S}.

Example 4 (cont'd)**.** *We have* $a(\mathbf{1}_{V_C}, 1) = 1$ *by definition and* $a(\mathbf{1}_{V_C}, 2) = \min\{2, 4\} = 2$*, implying that* $\varphi_{\text{mVaR}}(\mathbf{1}_{V_C}) = 2B = 6$ *and thus* $d^{\text{S}}(\mathbf{1}_{V_C}) = 2 = \overline{d}^m$.

Dinh et al. (2018) introduce φ_{mVaR} as a valid lower bound for the minimum number of vehicles required to serve a set of customers in the context of stochastic chance constrained CVRPs with normally distributed customer demands as well as ambiguous chance constrained CVRPs over moment-based ambiguity sets where the chance constraints admit deterministic representations as individual second-order cone constraints. Their justification of φ_{mVaR} is derived from first principles, whereas our derivations in this section leverage Theorem 3 to apply φ_{mVaR} to a broader class of ambiguity sets that comprise, among others, scenario-based as well as Wasserstein and ϕ-divergence based representations of the ambiguous demand distribution.

In the remainder, we discuss how d^{S} can be evaluated efficiently for the technology set $\mathcal{C}_{\text{mVaR}}$.

Theorem 8. *For an ambiguity set of the form* (6) *with* $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$, $\varphi_{\text{mVaR}}(\boldsymbol{x})$ *can be evaluated to* κ*-accuracy in time* $\mathcal{O}(Wn \cdot \log m \cdot \log n \cdot \log \kappa^{-1})$.

We note that the accuracy κ in Theorem 8 is measured *relative* to the maximally possible demand; to obtain a complexity estimate for an *absolute* accuracy, the term $\log \kappa^{-1}$ has to be increased to $\log(\kappa^{-1} \cdot \max_{w \in \mathcal{W}} \mathbf{e}^\top \overline{\boldsymbol{q}}^w)$. We close this section with the computation of φ_{mVaR} for instances of the ambiguity set (6) where the scenario probabilities are ambiguous.

Proposition 10. *For an ambiguity set of the form* (6), $\varphi_{\text{mVaR}}(\boldsymbol{x})$ *can be computed to* κ*-accuracy:*

1. *in time* $\mathcal{O}((Wn\log n + W\log W) \cdot \log\kappa^{-1} \cdot \log m)$ *for the 1-norm and the* ∞*-norm ambiguity set;*

2. *in time* $\mathcal{O}((Wn\log n + W\log[\overline{\pi}/\kappa]) \cdot \log\kappa^{-1} \cdot \log m)$ *for the entropy ambiguity set, where* $\overline{\pi} = \max\{\pi_w(\boldsymbol{x}) \, : \, w \in \mathcal{W}\}$*;*

3. *in time* $\mathcal{O}((Wn\log n + W\log W \cdot \log\kappa^{-1}) \cdot \log\kappa^{-1} \cdot \log m)$ *for the axis-parallel ellipsoidal ambiguity set;*

4. *in polynomial time via FISTA for the ellipsoidal ambiguity set.*

For ease of notation, Proposition 10 again uses a relative accuracy κ.

6 Numerical Experiments

Our numerical experiments use the standard CVRP benchmark instances compiled by Díaz (2006). Each instance label 'X-nY-kZ' indicates the literature source X of the instance, the number Y of nodes (including the depot) as well as the number Z of vehicles. Since our ambiguity set construction below is based on geographic information, we disregard instances that do not provide Euclidean coordinates for the nodes. Following the literature convention, we identify the transportation costs c_{ij} with the 2-norm distance between i and j, rounded to the nearest integer number.

The customer demands in the CVRP benchmark problems are deterministic. To construct stochastic demands whose distribution is characterized by an ambiguity set of the form (6), we subdivide each instance into 4 quadrants (northwest, northeast, southwest and southeast) according to the nodal coordinates. We then create $W = 4$ scenarios with equal probabilities $\hat{\boldsymbol{s}} = \mathbf{e}/4$, each of which is associated with one of the quadrants. In each scenario we set the expected demands of the customers in the associated quadrant to 110%, of the customers in the horizontally or vertically adjacent quadrants to 100%, and of the customers in the diagonally opposite quadrant to 90% of the nominal demands from the deterministic instance. The lower and upper demand bounds

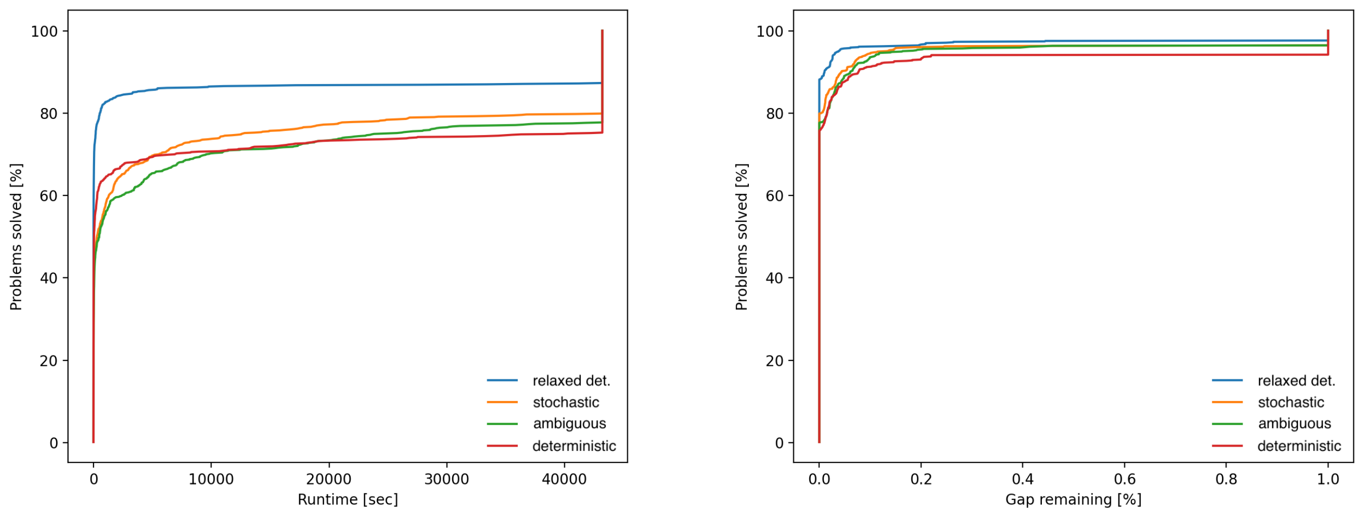

Figure 2. Runtimes and optimality gaps for our branch-and-cut schemes. Shown are the runtimes (left graph) and optimality gaps after 12 hours (right graph) for the deterministic branch-and-cut scheme (color) as well as the distributionally robust branch-and-cut schemes with known (color) and unknown (color) scenario probabilities.

undercut and exceed these expected demands in each scenario by 10% of the nominal demands. The mean absolute deviations of the customer demands are set to those of a Normal distribution that is centered at the mean demands and that places 90% of its probability mass onto the demand interval. Since the CVRP benchmark instances tend to have little slack in the vehicle capacities, we follow Gounaris et al. (2013) and Ghosal and Wiesemann (2020) and increase the vehicle capacities Q by 20% to ensure that the distributionally robust instances are feasible.

We implemented a 'vanilla' CVRP solution scheme that augments the branch-and-cut capacity of CPLEX Studio 20.1.0 with an RCI cut separation procedure that follows the tabu search algorithm outlined by Augerat et al. (1998). Our method is implemented in C++, and the source code is available on the authors' webpages (see Footnote 1). All problems are solved in single-core mode on an Intel Xeon 2.66GHz processor with 8GB main memory and a runtime limit of 12 hours.

6.1 Runtime Comparison

In our first experiment, we compare the runtimes and optimality gaps of our branch-and-cut algorithm for the deterministic CVRP with those of our algorithm for the distributionally robust CVRP under the 90%-CVaR risk measure. To this end, we consider two variants of the deterministic CVRP: one ('deterministic') where the original vehicle capacities are employed, and another one ('relaxed deterministic') where the vehicle capacities are increased by 20% as in the uncertainty-affected CVRP. We also consider two variants of our ambiguity set (6): in the 'stochastic' case, the scenario probabilities are known to be $\hat{\boldsymbol{s}} = \mathrm{e}/4$, whereas in the 'ambiguous' case these probabilities are only known to be contained in a 1-norm ambiguity set of radius $\theta = 0.1$ that is centered at the nomial probabilities $\hat{\boldsymbol{s}}$. The ambiguous setup thus models a total variation ambiguity set with a uniform empirical distribution (or 'prior'), see Example 3. The results are summarized in Figure 2 and presented in further detail in Table 1 in Appendix C.

The results show that, as expected, increasing the vehicles' capacity by 20% in the deterministic CVRP substantially simplifies the problem instances. If the price to paid by accounting for uncertainty was to be small, we would expect the runtimes and optimality gaps for the stochastic and ambiguous instances to be contained in the interval spanned by the deterministic and relaxed deterministic instances. In fact, although the stochastic and ambiguous instances enjoy a capacity increase of 20% (akin to the relaxed deterministic instances), the incorporation of demand variability and distributional ambiguity as well as risk and ambiguity aversion reduce the factually available vehicle capacity. On the other hand, our construction of the stochastic and ambiguous instances guarantees that the uncertain customer demands never exceed 20% of their nominal values from the deterministic instances. The results show that, broadly, the runtimes and optimality gaps for the stochastic and ambiguous instances are upper and lower bounded by those of the deterministic and the relaxed deterministic instances, which indicates that the computational price to be paid is mainly determined by the slack in the vehicle capacities and less so by the incorporation of risk and ambiguity. We thus conclude that the same branch-and-cut algorithm can solve all three problem classes in runtimes and optimality gaps that are of the same order of magnitude.

6.2 The Impact of Risk Aversion

In our second experiment, we focus on the benchmark instance A-n32-k5 and solve the distributionally robust CVRP associated with the family of exponential disutility functions $U_a(q) = (\exp(aq) - 1)/a$, $a > 0$, and $U_0(q) = q$. The scalar parameter $a \in \mathbb{R}_+$ controls the risk aversion of the decision maker: $a = 0$ reflects a neutral stance towards demand variability, whereas larger values correspond to an increasing risk aversion. For every value of a, we set the budget B in (5) to

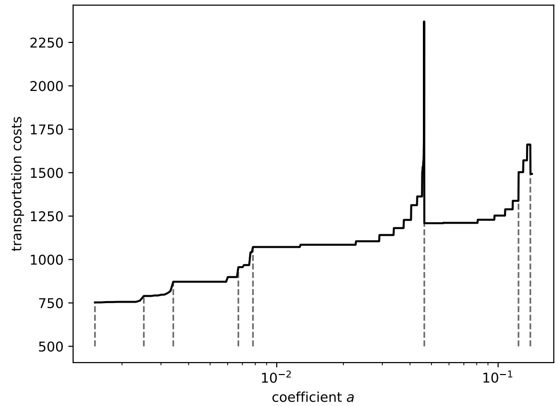

Figure 3. Minimum number of vehicles and optimal transportation costs for A-n32-k5 with an exponential class of disutility functions parameterized by a. The vertical lines indicate the parameter ranges covered by 5, 6, . . . , 11 vehicles (from left to right).

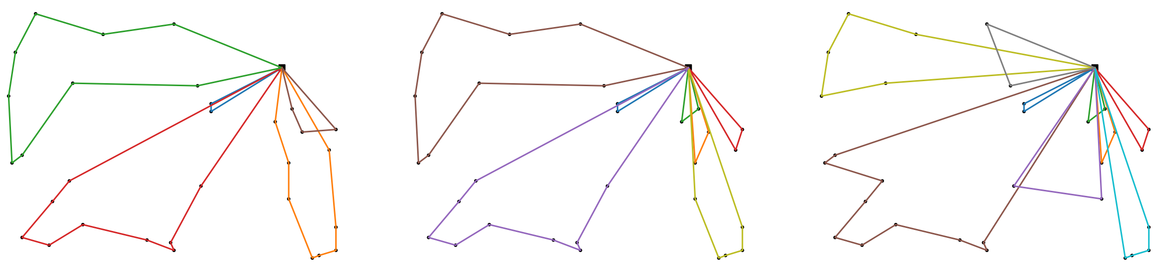

Figure 4. Optimal route plans for A-n32-k5 with exponential disutilities $a = 0$ (left; 5 vehicles), $a = 3.41$E-3 (middle; 7 vehicles) and $a = 7.81$E-3 (right; 9 vehicles).

$B = U_a(1.2Q)$, where Q is the nominal vehicle capacity from the deterministic CVRP instance and the factor 1.2 corresponds to the 20% capacity increase described earlier. This choice ensures that the feasibility of route plans for deterministic demands is unaffected by the choice of the risk aversion a and coincides with that of the deterministic instance (apart from the 20% capacity increase). Figure 3 visualizes how the minimum number of vehicles required to serve the customer demands, as well as the resulting transportation costs, vary as a function of the risk aversion a. Moreover, Figure 4 illustrates the optimal route plans for three different choices of a. We observe that higher degrees of risk aversion require larger numbers of vehicles to serve the customer demands, which in turn tends to increase the transportation costs (apart from two dips where the necessity to increase the number of vehicles results in smaller overall costs).

7 Conclusions

The use of ambiguity sets and risk measures to reflect different degrees of knowledge and attitudes towards ambiguity and risk is well established in stochastic programming and distributionally robust optimization. In this paper, we propose a framework that studies a broad variety of ambiguity sets and risk measures for the CVRP. An attractive feature of our framework is that all emerging combinations can be solved with minimal adaptations of the same branch-and-cut scheme, and the resulting algorithms perform broadly on par with those for the deterministic CVRP, thus allowing practitioners to incorporate uncertainty without incurring an excessive computational burden.

In the stochastic and distributionally robust CVRP, uncertain demands are *qua definitione* high-dimensional, and therefore the standard model-free characterizations of the underlying probability distribution result in optimization problems that can be overly conservative (*e.g.*, if pure moment-based descriptions are being employed) or computationally prohibitive (*e.g.*, if pure data-driven characterizations are being used). Our framework attempts to alleviate this issue by combining a scenario-based description (which characterizes systematic effects that affect multiple customers) with moment information (that describes the idiosyncratic variability of individual demands). A promising avenue for future research, in our view, is the study and comparison of alternative model-based ambiguity sets for the uncertainty-affected CVRP that offer realistic descriptions of the uncertain customer demands while avoiding the curse of dimensionality that plagues direct characterizations of high-dimensional probability distributions.

References

Y. Adulyasak and P. Jaillet. Models and algorithms for stochastic and robust vehicle routing with deadlines. *Transportation Science*, 50(2):608–626, 2015.

A. Agra, M. Christiansen, R. Figueiredo, L. M. Hvattum, M. Poss, and C. Requejo. Layered formulation for the robust vehicle routing problem with time windows. In A. Ridha Mahjoub, V. Markakis, I. Milis, and V. T. Paschos, editors, *Combinatorial Optimization: Second International Symposium (ISCO)*, chapter 7, pages 249–260. Springer, 2012.

A. Agra, M. Christiansen, R. Figueiredo, L. M. Hvattum, M. Poss, and C. Requejo. The robust vehicle routing problem with time windows. *Computers & Operations Research*, 40(3):856–866, 2013a.

A. Agra, M. Christiansen, L. M. Hvattum, and F. Rodrigues. Robust optimization for a maritime inventory routing problem. *Transportation Science*, 52(3):509–525, 2013b.

P. Augerat, J. M. Belenguer, E. Benavent, A. Corberán, and D. Naddef. Separating capacity constraints in the CVRP using tabu search. *European Journal of Operational Research*, 106(2–3):546–557, 1998.

G. Bayraksan and D. K. Love. Data-driven stochastic programming using phi-divergences. In D. M. Aleman and A. C. Thiele, editors, *INFORMS TutORials in Operations Research*, pages 1–19. INFORMS, 2015.

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

A. Beck and M. Teboulle. Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization*, 22(2):557–580, 2012.

F. Bellini and V. Bignozzi. On elicitable risk measures. *Quantitative Finance*, 15(5):725–733, 2015.

F. Bellini, B. Klaar, A. Muller, and E. R. Giannin. Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54:41–48, 2014.

A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

A. Ben-Tal, D. den Hertog, A. de Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

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

D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

D. Bertsimas, S. Shtern, and B. Sturt. A data-driven approach to multi-stage stochastic linear optimization. Forthcoming in Management Science, 2021.

J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 2nd edition, 2011.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

J. G. Carlsson and M. Behroozi. Worst-case demand distributions in vehicle routing. *European Journal of Operational Research*, 256(2):462–472, 2017.

J. G. Carlsson and E. Delage. Robust partitioning for stochastic multivehicle routing. *Operations Research*, 61(3):546–557, 2013.

J. G. Carlsson, M. Behroozi, and K. Mihic. Wasserstein distance and the distributionally robust TSP. *Operations Research*, 66(6):1603–1624, 2018.

Z. Chen, M. Sim, and H. Xu. Distributionally robust optimization with infinitely constrained ambiguity sets. *Operations Research*, 67(5):1328–1344, 2019.

N. Christofides. The vehicle routing problem. *RAIRO (Recherche Opérationnelle)*, 10(1):55–70, 1976.

G. Cornuejols and F. Harche. Polyhedral study of the capacitated vehicle routing problem. *Mathematical Programming*, 60(1):21–52, 1993.

G. B. Dantzig and J. H. Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):596–612, 2010.

B. D. Díaz. The VRP web. `http://www.bernabe.dorronsoro.es/vrp/`, 2006. Online; accessed October 2021.

T. Dinh, R. Fukasawa, and J. Luedtke. Exact algorithms for the chance-constrained vehicle routing problem. *Mathematical Programming*, 172(1–2):105–138, 2018.

E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1–2):37–61, 2006.

L. Eufinger, J. Kurtz, C. Buchheim, and U. Clausen. A robust approach to the capacitated vehicle routing problem with uncertain costs. *INFORMS Journal on Optimization*, 2(2):79–95, 2020.

H. Föllmer and T. Knispel. Entropic risk measures: Coherence vs convexity, model ambiguity and robust large deviations. *Stochastics and Dynamics*, 11(2–3):1–19, 2011.

H. Föllmer and A. Schied. Convex and coherent risk measures. In R. Cont, editor, *Encyclopedia of Quantitative Finance*. John Wiley and Sons, 2010.

M. Gendreau, O. Jabali, and W. Rei. Stochastic vehicle routing problems. In P. Toth and D. Vigo, editors, *Vehicle Routing: Problems, Methods, and Applications*, chapter 8, pages 213–239. MOS-SIAM, 2nd edition, 2014.

M. Gendreau, O. Jabali, and W. Rei. Future research directions in stochastic vehicle routing. *Transportation Science*, 50(4):1163–1173, 2016.

S. Ghosal and W. Wiesemann. The distributionally robust chance constrained vehicle routing problem. *Operations Research*, 68(3):716–732, 2020.

C. E. Gounaris, W. Wiesemann, and C. A. Floudas. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research*, 61(3):677–693, 2013.

N. G. Hall, D. Z. Long, J. Qi, and M. Sim. Managing underperformance risk in project portfolio selection. *Operations Research*, 63(3):660–675, 2015.

M. Hoogeboom, Y. Adulyasak, W. Dullaert, and P. Jaillet. The robust vehicle routing problem with time window assignments. *Transportation Science*, 55(2):395–413, 2021.

S. Irnich, P. Toth, and D. Vigo. The family of vehicle routing problems. In P. Toth and D. Vigo, editors, *Vehicle Routing: Problems, Methods, and Applications*, chapter 1, pages 1–33. MOS-SIAM, 2nd edition, 2014.

G. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

P. Jaillet, J. Qi, and M. Sim. Routing optimization under uncertainty. *Operations Research*, 64(1):186–200, 2016.

D. Kuhn, P. Mohajerin Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In S. Netessine, editor, *INFORMS TutORials in Operations Research*, pages 130–166. INFORMS, 2019.

G. Laporte and Y. Norbert. A branch and bound algorithm for the capacitated vehicle routing problem. *Operations Research Spektrum*, 5(2):77–85, 1983.

G. Laporte, F. Louveaux, and H. Mercure. The vehicle routing problem with stochastic travel times. *European Journal on Operational Research*, 39(1):71–78, 1989.

C. Lee, K. Lee, and S. Park. Robust vehicle routing problem with deadlines and travel time/demand uncertainty. *Journal of the Operational Research Society*, 63(9):1294–1306, 2012.

D. Z. Long, J. Qi, and A. Zhang. Supermodularity in two-stage distributionally robust optimization. Available on Optimization Online, 2020.

D. Lu and F. Gzarao. The robust vehicle routing problem with time windows: Solution by branch and price and cut. *European Journal on Operational Research*, 275(3):925–938, 2019.

J. Lysgaard, A. N. Letchford, and R. W. Eglese. A new branch-and-cut algorithm for the capacitated vehicle routing problem. *Mathematical Programming*, 100(2):423–445, 2004.

N. Megiddo and T. Ichimori. A two-resource allocation problem solvable in linear time. *Mathematics of Operations Research*, 10(1):7–16, 1985.

P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1–2):115–166, 2018.

P. Munari, A. Moreno, J. De La Vega, D. Alem, J. Gondzio, and R. Morabito. The robust vehicle routing problem with time windows: Compact formulation and branch-price-and-cut method. *Transportation Science*, 53(4):1043–1066, 2019.

A. Nilim and L. El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

B. O'Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15(3):715–732, 2015.

F. Ordóñez. Robust vehicle routing. In J. J. Hasenbein, editor, *INFORMS TutORials in Operations Research*, chapter 7, pages 153–178. INFORMS, 2010.

J. Oyola, H. Arntzen, and D. L. Woodruff. The stochastic vehicle routing problem, a literature review, part I: Models. *EURO Journal on Transportation and Logistics*, 7(3):193–221, 2018.

A. A. Pessoa and M. Poss. Robust network design with uncertain outsourcing cost. *INFORMS Journal on Computing*, 27(3):507–524, 2015.

A. A. Pessoa, L. Di Puglia Pugliese, F. Guerriero, and M. Poss. Robust constrained shortest path problems under budgeted uncertainty. *Networks*, 66(2):98–111, 2015.

A. A. Pessoa, M. Poss, R. Sadykov, and F. Vanderbeck. Branch-cut-and-price for the robust capacitated vehicle routing problem with knapsack uncertainty. *Operations Research*, 69(3):739–754, 2021.

M. Petrik and D. Subramanian. RAAM: The benefits of robustness in approximating aggregated MDPs in reinforcement learning. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, volume 2, 2014.

L. Di Puglia Pugliese, F. Guerriero, and M. Poss. The resource constrained shortest path problem with uncertain data: A robust formulation and optimal solution approach. *Computers & Operations Research*, 107:140–155, 2019.

R. T. Rockafeller and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002.

F. Semet, P. Toth, and D. Vigo. Classical exact algorithms for the capacitated vehicle routing problem. In P. Toth and D. Vigo, editors, *Vehicle Routing: Problems, Methods, and Applications*, chapter 2, pages 37–57. MOS-SIAM, 2nd edition, 2014.

A. Shapiro and A. J. Kleywegt. Minmax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542, 2002.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2nd edition, 2014.

D. Simchi-Levi, X. Chen, and J. Bramel. *The Logic of Logistics*. Springer, 3rd edition, 2005.

A. Subramanyam, P. P. Repoussis, and C. E. Gounaris. Robust optimization of a broad class of heterogeneous vehicle routing problems under demand uncertainty. *INFORMS Journal on Computing*, 32(3):661–681, 2020.

A. Subramanyam, F. Mufalli, J. M. Laínez-Aguirre, J. M. Pinto, and C. E. Gounaris. Robust multiperiod vehicle routing under customer order uncertainty. *Operations Research*, 69(1):30–60, 2021.

I. Sungur and F. Ordóñez. A robust optimization approach for the capacitated vehicle routing problem with demand uncertainty. *IIE Transactions*, 40(5):509–523, 2008.

P. Toth and D. Vigo, editors. *The Vehicle Routing Problem*. SIAM, 2nd edition, 2014.

A. Wang, A. Subramanyam, and C. E. Gounaris. Robust vehicle routing under uncertainty via branch-price-and-cut. *Optimization and Engineering*, Available Online First, 2021.

W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2013.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

Y. Zhang, R. Baldacci, M. Sim, and J. Tang. Routing optimization with time windows under uncertainty. *Mathematical Programming*, 175(1):263–305, 2019.

Y. Zhang, Z. Zhang, A. Lim, and M. Sim. Robust data-driven vehicle routing with time windows. *Operations Research*, 69(2):469–485, 2021.

S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168, 2009.

Appendix A: Proofs

Proof of Proposition 1. In view of the first inequality, we note that $\underline{d}(\emptyset) = 0 = \overline{d}^1(\emptyset)$ by definition and $\underline{d}(S) = 1 \leqslant \overline{d}^1(S)$ for any $\emptyset \neq S \in \mathcal{C}$ since at least one route $\boldsymbol{R} \in \mathcal{C}$ is required in $\overline{d}^1$ to cover a non-empty set S. To see that $\underline{d}(S) = 2 \leqslant \overline{d}^1(S)$ for $\emptyset \neq S \notin \mathcal{C}$, we observe that $S \notin \mathcal{C}$ implies that there is at least one list $\boldsymbol{S}$ formed from the elements of S such that $\boldsymbol{S} \notin \mathcal{C}$. Assumptions **(D)** and **(P)** then imply that $\mathcal{C}$ cannot contain any route $\boldsymbol{R}$ that contains all of the customers of $\boldsymbol{S}$ (in any order). We thus conclude that $S \nsubseteq \boldsymbol{R}$ for any $\boldsymbol{R} \in \mathcal{C}$ and thus $\overline{d}^1(S) \geqslant 2$.

To show that **(D)** and **(P)** are necessary for the first inequality, assume first that assumption **(D)** is violated. Then there is $\boldsymbol{R} \in \mathcal{C}$ such that $\boldsymbol{S} \notin \mathcal{C}$ for some subsequence $\boldsymbol{S}$ of $\boldsymbol{R}$. One readily verifies that $\underline{d}(S) = 2$ but $\overline{d}^1(S) = 1$ for the set S formed from the elements of $\boldsymbol{S}$; the latter holds since the route covering $\boldsymbol{R}$ in the definition of $\overline{d}^1$ also covers $\boldsymbol{S}$. If assumption **(P)** is violated, on the other hand, then there is $\boldsymbol{R} \in \mathcal{C}$ such that $\boldsymbol{S} \notin \mathcal{C}$ for some permutation $\boldsymbol{S}$ of $\boldsymbol{R}$. One again verifies that $\underline{d}(S) = 2$ yet $\overline{d}^1(S) = 1$ for the set S formed from the elements of $\boldsymbol{S}$; the latter holds since the route covering $\boldsymbol{R}$ in the definition of $\overline{d}^1$ also covers any permutation of $\boldsymbol{R}$.

As for the second inequality, assume that $\overline{d}^m(S) = \theta$ for some $\emptyset \neq S \subseteq V_C$; the case $S = \emptyset$ is trivial. The definition of $\overline{d}^m$ implies that there is $\{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_\theta, \ldots \boldsymbol{R}_m\} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m$ such that $S \subseteq \boldsymbol{R}_1 \cup \ldots \cup \boldsymbol{R}_\theta$. Since $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_\theta \in \mathcal{C}$ by definition of $\mathcal{C}_m$, we have $\overline{d}^1(S) \leqslant \theta$ as desired. $\square$

Proof of Theorem 1. Fix any demand estimator d. The statement follows if we show that:

(i) Any $\mathbf{R} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m$ induces a solution $\boldsymbol{x}(\mathbf{R})$ feasible in 2VF(d) if $d \leqslant \overline{d}^m$.

(ii) Any solution $\boldsymbol{x}$ feasible in 2VF(d) induces $\mathbf{R}(\boldsymbol{x}) \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m$ if $d \geqslant \underline{d}$.

In view of *(i)*, fix any $\mathbf{R} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m$. The definition of $\boldsymbol{x}(\mathbf{R})$ in (3) implies that $\boldsymbol{x}(\mathbf{R})$ satisfies the binarity and degree constraints of 2VF(d). To see that $\boldsymbol{x}(\mathbf{R})$ satisfies the capacity constraints of 2VF(d), we note that for any $\emptyset \neq S \subseteq V_C$, we have that

$$\begin{aligned} d(S) \;\leqslant\; \overline{d}^m(S) &= \inf\left\{ J \in \mathbb{N} \,:\, S \subseteq \bigcup_{k=1,\ldots,J} \boldsymbol{R}'_k \text{ for } \{\boldsymbol{R}'_1, \ldots, \boldsymbol{R}'_J, \ldots, \boldsymbol{R}'_m\} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m \right\} \\ &\leqslant \inf\left\{ J \in \mathbb{N} \,:\, S \subseteq \bigcup_{k=1,\ldots,J} \boldsymbol{R}_{j_k} \text{ for } j_1, \ldots, j_J \in \{1, \ldots, m\} \right\} \\ &= |k \in K \,:\, \boldsymbol{R}_k \cap S \neq \emptyset| \;\leqslant\; \sum_{i \in V \setminus S} \sum_{j \in S} x_{ij}(\mathbf{R}), \end{aligned}$$

where the second inequality holds since $\mathbf{R} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m$. The second equality is due to the fact that the minimum number of routes in $\mathbf{R}$ required to cover S is precisely the number of routes $\boldsymbol{R}_k$, $k = 1, \ldots, m$, that have a nonempty intersection with S. In view of the last inequality, finally, fix $k \in K$ with $\boldsymbol{R}_k \cap S \neq \emptyset$ and let $j_k \in \boldsymbol{R}_k \cap S$ be the first customer on the route $\boldsymbol{R}_k$ that is contained in S. By definition of j_k, we have $\sum_{i \in V \setminus S} x_{ij_k}(\mathbf{R}) = 1$. The inequality now follows from the fact that there are $|k \in K : \boldsymbol{R}_k \cap S \neq \emptyset|$ different customer nodes j_k with this property.[2] In summary, we have shown that $\boldsymbol{x}(\mathbf{R})$ satisfies the capacity constraints—and thus all constraints—of 2VF(d).

As for *(ii)*, fix any feasible solution $\boldsymbol{x}$ in 2VF(d). We construct a route plan $\mathbf{R}(\boldsymbol{x})$, in the following abbreviated as $\mathbf{R}$, satisfying (3) as follows. Since $\sum_{j \in V_C} x_{0j} = m$, there is $j_1, \ldots, j_m \in V_C$, $j_1 < \ldots < j_m$, such that $x_{0,j_1} = \ldots = x_{0,j_m} = 1$. For each route $\boldsymbol{R}_k$, $k \in K$, we set $R_{k,1} \leftarrow j_k$ and $n_k \leftarrow 1$. Since $\sum_{j \in V} x_{R_{k,n_k},j} = 1$, we either have $x_{R_{k,n_k},j} = 1$ for some $j \in V_C$ or $x_{R_{k,n_k},0} = 1$. In the former case, we extend route $\boldsymbol{R}_k$ by the customer $R_{k,n_k+1} \leftarrow j$, we set $n_k \leftarrow n_k + 1$ and we continue the procedure with customer j. In the latter case, we have completed the route $\boldsymbol{R}_k$. By construction, the route plan $\mathbf{R}$ thus created satisfies (3).

We show that $\mathbf{R} \in \mathfrak{P}(V_C, m)$. Note that $n_k \geqslant 1$ due to the existence of the customers $j_1, \ldots, j_m$. The degree constraints in 2VF(d) ensure that $R_{k,i} \neq R_{l,j}$ for all $(k,i) \neq (l,j)$. It remains to be shown that $\bigcup_k \boldsymbol{R}_k = V_C$. Imagine, to the contrary, that there is a customer $j \in V_C$ such that $j \notin \bigcup_k \boldsymbol{R}_k$. By construction of the above algorithm, j must lie on a short cycle $S \subset V_C$ that is not connected to the depot node 0. Since $S \neq \emptyset$, its associated capacity constraint would require that $\sum_{i \in V \setminus S} \sum_{j \in S} x_{ij} \geqslant d(S) \geqslant \underline{d}(S) \geqslant 1$. However, $\sum_{i \in V \setminus S} \sum_{j \in S} x_{ij} = 0$ because S is a short cycle not connected to the depot node 0. Thus, the capacity constraint associated with S would be violated.

We finally show that $\{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m\} \in \mathcal{C}_m$ as well. We have $\sum_{i \in V \setminus \boldsymbol{R}_k} \sum_{j \in \boldsymbol{R}_k} x_{ij} = 1 \geqslant d(R_k) \geqslant \underline{d}(R_k)$ for all $k \in K$, where R_k is the set formed from the customers in $\boldsymbol{R}_k$. Here, the equality follows from the construction of the routes $\boldsymbol{R}_k$, and the two inequalities hold due to the feasibility of $\boldsymbol{x}$ in 2VF(d) and the fact that $d \geqslant \underline{d}$, respectively. Since $\boldsymbol{R}_k \neq \emptyset$, we thus conclude that $\underline{d}(R_k) = 1$, that is, $\boldsymbol{R}_k \in \mathcal{C}$, for all $k \in K$. This implies that $\{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m\} \in \mathcal{C}_m$, and consequently we have $\mathbf{R} \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m$ as desired. □

Proof of Proposition 2. In view of assertion *(i)*, fix any VRP($\mathcal{C}$) instance and demand estimator

[2] Note that the same vehicle may enter and leave the set S multiple times, hence we cannot strengthen the inequality to an equality in general.

d as described in the statement. The first part of the proof of Theorem 1 implies that any route plan $\mathbf{R}$ feasible in VRP($\mathcal{C}$) induces a solution $\boldsymbol{x}(\mathbf{R})$ that is feasible in 2VF(d). Thus, we only need to show that any solution $\boldsymbol{x}$ feasible in 2VF(d) also induces a route plan $\mathbf{R}(\boldsymbol{x})$ that is feasible in VRP($\mathcal{C}$). Indeed, the route plan $\mathbf{R}(\boldsymbol{x})$ considered in the second part of the proof of Theorem 1 satisfies $n_k \geqslant 1$ due to the degree constraint for the depot node and $R_{k,i}(\boldsymbol{x}) \neq R_{l,j}(\boldsymbol{x})$ for all $(k,i) \neq (l,j)$ by virtue of the degree constraints for the customer nodes, respectively. Moreover, we have $\bigcup_k \boldsymbol{R}_k(\boldsymbol{x}) = V_C$ since our earlier assumption that $d(S) > 0$ for all nonempty $S \subseteq V_C$ disallows any short cycles in $\boldsymbol{x}$. We thus conclude that $\mathbf{R}(\boldsymbol{x}) \in \mathfrak{P}(V_C, m)$. Since $\mathfrak{P}(V_C, m) \subseteq \mathcal{C}_m$, it follows that $\mathbf{R}(\boldsymbol{x}) \in \mathcal{C}_m$ as well.

As for assertion *(ii)*, we first show that for any VRP($\mathcal{C}$) instance with $\mathfrak{P}(V_C, m) \nsubseteq \mathcal{C}_m$ the demand estimator d defined through $d(S) = 0$ if $S = \emptyset$ and $d(S) = 1$ otherwise satisfies $d \ngeqslant \underline{d}$ and implies that VRP($\mathcal{C}$) and 2VF(d) are *not* equivalent. To see that $d \ngeqslant \underline{d}$, we note that $\underline{d}(V_C) > 1$ since otherwise $V_C \in \mathcal{C}$, which would in turn imply by **(D)** that $S \in \mathcal{C}$ for all $S \subseteq V_C$ and hence $\mathfrak{P}(V_C, m) \subseteq \mathcal{C}_m$ in contradiction to our assumption. To see that VRP($\mathcal{C}$) and 2VF(d) are not equivalent, fix any $\mathbf{R} \in \mathfrak{P}(V_C, m) \backslash \mathcal{C}_m$. We show that the solution $\boldsymbol{x}(\mathbf{R})$ defined through (3) is feasible in 2VF(d), which implies that VRP($\mathcal{C}$) and 2VF(d) are not equivalent. Indeed, $\boldsymbol{x}(\mathbf{R})$ satisfies the binarity and degree constraints in 2VF(d) by construction, and it satisfies all capacity constraints since $\sum_{i \in V \backslash S} \sum_{j \in S} x_{ij}(\mathbf{R}) \geqslant 1 = d(S)$ for all nonempty $S \subseteq V_C$.

We now show that for any VRP($\mathcal{C}$) instance with $\mathfrak{P}(V_C, m) \nsubseteq \mathcal{C}_m$ the demand estimator d defined through $d(S) = 1$ if $S = V_C$ and $d(S) = \underline{d}(S)$ otherwise satisfies $d \ngeqslant \underline{d}$ and makes VRP($\mathcal{C}$) and 2VF(d) equivalent. To see that $d \ngeqslant \underline{d}$, we note that $d(V_C) = 1$ whereas $\underline{d}(V_C) = 2$ according to our discussion from the previous paragraph. To see that VRP($\mathcal{C}$) and 2VF(d) are equivalent under d, the first part of the proof of Theorem 1 implies that we only need to show that any solution $\boldsymbol{x}$ feasible in 2VF(d) induces a route plan $\mathbf{R}(\boldsymbol{x})$ that is feasible in VRP($\mathcal{C}$). The route plan $\mathbf{R}(\boldsymbol{x})$ constructed in the second part of the proof of Theorem 1 satisfies $n_k \geqslant 1$, $R_{k,i}(\boldsymbol{x}) \neq R_{l,j}(\boldsymbol{x})$ for all $(k,i) \neq (l,j)$ and $\bigcup_k \boldsymbol{R}_k(\boldsymbol{x}) = V_C$. Thus, we have $\mathbf{R}(\boldsymbol{x}) \in \mathfrak{P}(V_C, m)$. To see that $\mathbf{R}(\boldsymbol{x}) \in \mathcal{C}_m$, we first note that no single route $\boldsymbol{R}_k(\boldsymbol{x})$ can contain all customers since otherwise $m = 1$ and the assumption that $\mathfrak{P}(V_C, 1) \nsubseteq \mathcal{C}_1$ implies that the VRP($\mathcal{C}$) instance is infeasible, which contradicts the assumptions of the theorem. Next, we note that $\boldsymbol{R}_k(\boldsymbol{x}) \in \mathcal{C}$ for all $k \in K$ as the capacity constraints $\sum_{i \in V \backslash \boldsymbol{R}_k(\boldsymbol{x})} \sum_{j \in \boldsymbol{R}_k(\boldsymbol{x})} x_{ij} = 1 \geqslant d(\boldsymbol{R}_k(\boldsymbol{x})) = \underline{d}(\boldsymbol{R}_k(\boldsymbol{x}))$ are satisfied; the last equality follows from the fact that $\boldsymbol{R}_k(\boldsymbol{x})$ does not contain all customers. By definition of $\mathcal{C}_m$, we have $\mathbf{R}(\boldsymbol{x}) \in \mathcal{C}_m$, that is, VRP($\mathcal{C}$) and 2VF($d$) are indeed equivalent.

In view of assertion *(iii)*, fix any VRP($\mathcal{C}$) instance and demand estimator d as described in the statement. We prove the assertion by constructing a route plan $\mathbf{R}'$ feasible in VRP($\mathcal{C}$) such that the associated solution $\boldsymbol{x}(\mathbf{R}')$ is not feasible in 2VF(d). To this end, fix $S \subseteq V_C$ such that $d(S) > \overline{d}^m(S)$, and let $\mathbf{R} = \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m\}$ be such that $S \subseteq \bigcup_{k=1,\ldots,\overline{d}^m(S)} \boldsymbol{R}_k$. Such a route plan exists since the feasibility of VRP($\mathcal{C}$) implies that $\overline{d}^m(S) \neq \infty$. We now construct the desired route plan $\mathbf{R}' = \{\boldsymbol{R}'_1, \ldots, \boldsymbol{R}'_m\}$ from $\mathbf{R}$ as follows. We set $\boldsymbol{R}'_k = \boldsymbol{R}_k$ for any route k satisfying $\boldsymbol{R}_k \cap S = \emptyset$. For the other routes $\boldsymbol{R}_k$, we obtain $\boldsymbol{R}'_k$ by reordering the customers in $\boldsymbol{R}_k$ such that those in $\boldsymbol{R}_k \cap S$ appear first (in any order). The assumption **(P)** implies that $\mathbf{R}' \in \mathfrak{P}(V_C, m) \cap \mathcal{C}_m$ as well. For the solution $\boldsymbol{x}(\mathbf{R}')$ constructed from (3), however, we observe that

$$\begin{aligned}\sum_{i \in V \setminus S} \sum_{j \in S} x_{ij}(\mathbf{R}') &= \sum_{i \in V \setminus S} \sum_{\substack{k \in K: \\ S \cap \boldsymbol{R}'_k \neq \emptyset}} \sum_{j \in S \cap \boldsymbol{R}'_k} x_{ij}(\mathbf{R}') = \sum_{\substack{k \in K: \\ S \cap \boldsymbol{R}'_k \neq \emptyset}} \sum_{i \in V \setminus S} \sum_{j \in S \cap \boldsymbol{R}'_k} x_{ij}(\mathbf{R}') \\ &= \sum_{\substack{k \in K: \\ S \cap \boldsymbol{R}'_k \neq \emptyset}} 1 = |k \in K \, : \, S \cap \boldsymbol{R}'_k \neq \emptyset| = \overline{d}^m(S),\end{aligned}$$

where the third equality follows from the reordering of the customers in $\mathbf{R}'$. Since $d(S) > \overline{d}^m(S)$, the solution $\boldsymbol{x}(\mathbf{R}')$ is infeasible in 2VF(d) even though $\mathbf{R}'$ is feasible in VRP($\mathcal{C}$). □

We split the proof of Theorem 2 into the following two lemmas.

Lemma 1. *There exist* VRP($\mathcal{C}$) *instances violating* **(D)** *but satisfying* **(P)** *such that* VRP($\mathcal{C}$) *and* 2VF(d) *are not equivalent under any demand estimator d.*

Proof. Consider the VRP($\mathcal{C}$) instance with $n = 4$ customers, $m = 2$ vehicles and $\mathcal{C}$ consisting of all routes that comprise 1, 3 or 4 customers. This instance satisfies the assumption **(P)**, but it violates the assumption **(D)** since, for example, $(1, 2) \notin \mathcal{C}$ even though $(1, 2, 3) \in \mathcal{C}$. The feasible route plans of VRP($\mathcal{C}$) are all partitions in $\mathfrak{P}(V_C, 2)$ where one vehicle serves one customer and the other vehicle serves the remaining three customers.

We claim that there is no demand estimator d such that 2VF(d) has the same set of feasible solutions. Indeed, note that any admissible d must satisfy $d(S) \leqslant 2$ for all $S \subseteq V_C$ in order to result in a feasible 2VF(d) instance. Moreover, to allow for the feasible solutions $\boldsymbol{x}(\{(1, 2, 3), (4)\})$ and $\boldsymbol{x}(\{(1), (2, 3, 4)\})$, any admissible d must satisfy $d(S) \leqslant 1$ for all nonempty subsets of $\{1, 2, 3\}$

and $\{2,3,4\}$. This implies, however, that any admissible demand estimator must result in a 2VF(d) instance that also allows for the infeasible solution $\boldsymbol{x}(\{(1,2),(3,4)\})$. $\square$

Lemma 2. *There exist* VRP($\mathcal{C}$) *instances violating* **(P)** *but satisfying* **(D)** *such that* VRP($\mathcal{C}$) *and* 2VF(d) *are not equivalent under any demand estimator d.*

Proof. Consider the VRP($\mathcal{C}$) instance with $n = 2$ customers, $m = 1$ vehicle and $\mathcal{C} = \{(1), (2), (1,2)\}$. This instance satisfies **(D)**, but it violates **(P)** since $(2,1) \notin \mathcal{C}$ even though $(1,2) \in \mathcal{C}$. The only feasible route plan for VRP($\mathcal{C}$) is $\{(1,2)\}$.

We claim that there is no demand estimator d such that 2VF(d) has the same set of feasible solutions. Indeed, for the solution $\boldsymbol{x}(\{(1,2)\})$ to be feasible in 2VF(d), any admissible demand estimator d must satisfy $d(\{1\})$, $d(\{2\})$, $d(\{1,2\}) \leqslant 1$. However, any such demand estimator d would then also allow the infeasible route plan $\boldsymbol{x}(\{(2,1)\})$. $\square$

Proof of Theorem 2. The proof follows immediately from Lemmas 1 and 2. $\square$

Proof of Proposition 3. For any permutation $\boldsymbol{S}$ of $\boldsymbol{R} \in \mathcal{C}$, we have $\mathbf{1}_{\boldsymbol{S}} = \mathbf{1}_{\boldsymbol{R}}$ and thus $\varphi(\mathbf{1}_{\boldsymbol{S}}) = \varphi(\mathbf{1}_{\boldsymbol{R}}) \leqslant B$, implying that $\boldsymbol{S} \in \mathcal{C}$, that is, assumption **(P)** is satisfied. To prove that $\mathcal{C}$ satisfies **(D)** whenever φ is monotone, consider any $\boldsymbol{R} = (R_1, \ldots, R_\nu) \in \mathcal{C}$ and $\boldsymbol{S} = (R_{i_1}, \ldots, R_{i_\sigma})$ such that $1 \leqslant \sigma \leqslant \nu$ and $1 \leqslant i_1 < i_2 < \ldots < i_\sigma \leqslant \nu$. We then have $\mathbf{1}_{\boldsymbol{S}} \leqslant \mathbf{1}_{\boldsymbol{R}}$, and the monotonicity of φ implies that $\varphi\left(\mathbf{1}_{\boldsymbol{S}}\right) \leqslant \varphi\left(\mathbf{1}_{\boldsymbol{R}}\right) \leqslant B$. Thus, $\boldsymbol{S} \in \mathcal{C}$, and assumption **(D)** holds. $\square$

Proof of Proposition 4. We have $d^{\mathrm{P}}(\varnothing) = \overline{d}^1(\varnothing) = 0$, and any $\varnothing \neq S \subseteq V_C$ satisfies

$$
\begin{aligned}
\overline{d}^1(S) \;&=\; \min\left\{ I \in \mathbb{N} \,:\, S \subseteq \bigcup_{k=1,\ldots,I} \boldsymbol{R}_k \;\text{ for some } \boldsymbol{R}_1, \ldots, \boldsymbol{R}_I \in \mathcal{C} \right\} \\
&=\; \min\left\{ I \in \mathbb{N} \,:\, S \subseteq \bigcup_{k=1,\ldots,I} \boldsymbol{R}_k \;\text{ such that } \varphi\left(\mathbf{1}_{\boldsymbol{R}_k}\right) \leqslant B \text{ for all } k = 1, \ldots, I \right\} \\
&=\; \min\left\{ I \in \mathbb{N} : \exists \boldsymbol{X} \in \{0,1\}^{n \times I} \;\text{ such that } \boldsymbol{X}\mathbf{e} \geqslant \mathbf{1}_S,\;\; \varphi(\boldsymbol{x}_k) \leqslant B \;\; \forall k = 1, \ldots, I \right\} \\
&=\; \min\left\{ I \in \mathbb{N} : \exists \boldsymbol{X} \in \{0,1\}^{n \times I} \;\text{ such that } \boldsymbol{X}\mathbf{e} = \mathbf{1}_S,\;\; \varphi(\boldsymbol{x}_k) \leqslant B \;\; \forall k = 1, \ldots, I \right\} \\
&=\; d^{\mathrm{P}}(S),
\end{aligned}
$$

where the second identity follows from the definition of $\mathcal{C}$ in (4). The union on the right-hand side of the second identity corresponds to the constraint $\boldsymbol{X}\mathbf{e} \geqslant \mathbf{1}_S$ where $\boldsymbol{x}_k \in \{0,1\}^n$ for $k = 1, \ldots, I$

so that a customer can be assigned to more than one route. Choose any $\boldsymbol{X}$ that is feasible in the right-hand side of the third identity. Since φ is monotone, for each $i \in S$, we can arbitrarily choose one of the k's for which $x_{ik} = 1$ and set $x_{ik'} = 0$ for all $k' \neq k$. The monotonicity of φ guarantees that this new solution $\boldsymbol{x}_k$, $k = 1, \ldots, I$, is also feasible, which leads to the fourth identity. $\square$

Proof of Proposition 5. We denote the two expressions for the packing estimator d^{P} as

$$d_1(S) = \min \left\{ I \in \mathbb{N} : \exists \boldsymbol{X} \in [0,1]^{n \times I} \text{ such that } \boldsymbol{X}\mathbf{e} = \mathbf{1}_S, \;\; \varphi(\boldsymbol{x}_k) \leqslant B \;\; \forall k = 1, \ldots, I \right\}$$

and

$$d_2(S) = \min \left\{ I \in \mathbb{N} : \varphi(\mathbf{1}_S / I) \leqslant B \right\},$$

where $\varnothing \neq S \subseteq V_C$. We want to show that $d_1(S) = d_2(S)$ for all $\varnothing \neq S \subseteq V_C$. One readily verifies that $d_1(S) \leqslant d_2(S)$ since for any $I \in \mathbb{N}$ feasible in the minimization problem that defines $d_2(S)$, $(I', \boldsymbol{X}') = (I, \mathbf{1}_S \, \mathbf{e}^\top / I)$ is feasible in the minimization problem that defines $d_1(S)$ and achieves the same objective value I. To see that $d_1(S) \geqslant d_2(S)$, fix any solution $(I, \boldsymbol{X})$ that is feasible in the minimization problem that defines $d_1(S)$. In the following, we prove that $\varphi(\mathbf{1}_S / I) \leqslant B$, which shows that I is also feasible in the minimization problem that defines $d_2(S)$.

Let Π be the group of all permutations $\pi : \{1, \ldots, I\} \to \{1, \ldots, I\}$ of the set $\{1, \ldots, I\}$, and define $\pi(\boldsymbol{X}) = (\boldsymbol{x}_{\pi(1)}, \ldots, \boldsymbol{x}_{\pi(I)})$ for $\pi \in \Pi$. By construction, $(I, \pi(\boldsymbol{X}))$ is feasible in $d_1(S)$ for any $\pi \in \Pi$. Moreover, since for any fixed $I \in \mathbb{N}$ the projection of the feasible region of $d_1(S)$ onto $\boldsymbol{X}$ is convex by assumption, $(I, \boldsymbol{X}')$ with

$$\boldsymbol{X}' \; = \; \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \pi(\boldsymbol{X})$$

is also feasible in $d_1(S)$. However, the k-th column of $\boldsymbol{X}'$ satisfies

$$\boldsymbol{x}'_k \; = \; \frac{1}{I!} \cdot \sum_{\pi \in \Pi} \boldsymbol{x}_{\pi(k)} \; = \; \frac{1}{I!} \cdot \sum_{l=1}^{I} \sum_{\substack{\pi \in \Pi: \\ \pi(k) = l}} \boldsymbol{x}_{\pi(k)} \; = \; \frac{1}{I!} \cdot \sum_{l=1}^{I} (I-1)! \cdot \boldsymbol{x}_l \; = \; \frac{\mathbf{1}_S}{I},$$

where the first and penultimate equalities follow from the fact that a set with ℓ elements admits $\ell!$ permutations, and the last identify holds since $\boldsymbol{X}\mathbf{e} = \mathbf{1}_S$ as $\boldsymbol{X}$ is feasible in $d_1(S)$. $\square$

Proof of Theorem 3. We first show that $\underline{d} \leqslant d^{\mathrm{P}} \leqslant \overline{d}^m$, irrespective of whether φ is sub- or superadditive. The fact that $d^{\mathrm{P}} \leqslant \overline{d}^m$ follows from Proposition 4, which implies that $d^{\mathrm{P}} \leqslant \overline{d}^1$, and

Proposition 1, which shows that $\overline{d}^1 \leqslant \overline{d}^m$. To see that $d^{\mathrm{P}} \geqslant \underline{d}$, we note that $d^{\mathrm{P}}(S) \geqslant 1 = \underline{d}(S)$ for all $\emptyset \neq S \in \mathcal{C}$ by construction, while $d^{\mathrm{P}}(S) \geqslant 2 = \underline{d}(S)$ for all $\emptyset \neq S \notin \mathcal{C}$ since $\varphi(\mathbf{1}_S) > B$.

As for statement *(i)*, we note that $d^{\mathrm{S}} \geqslant \underline{d}$ by construction. To see that $d^{\mathrm{S}} \leqslant d^{\mathrm{P}}$ when φ is subadditive, we observe that for any $\emptyset \neq S \subseteq V_C$ and any $I \in \mathbb{N}$, we have

$$\begin{aligned} I \geqslant d^{\mathrm{P}}(S) \quad &\Longleftrightarrow \quad \exists \boldsymbol{X} \in [0,1]^{n \times I} \text{ such that } \boldsymbol{X}\mathbf{e} = \mathbf{1}_S, \;\; \varphi(\boldsymbol{x}_k) \leqslant B \;\; \forall k = 1, \ldots, I \\ &\Longrightarrow \quad \exists \boldsymbol{X} \in [0,1]^{n \times I} \text{ such that } \boldsymbol{X}\mathbf{e} = \mathbf{1}_S, \;\; \varphi(\boldsymbol{x}_1) + \ldots + \varphi(\boldsymbol{x}_I) \leqslant I \cdot B \\ &\Longrightarrow \quad \varphi(\mathbf{1}_S) \leqslant I \cdot B \\ &\Longleftrightarrow \quad I \geqslant d^{\mathrm{S}}(S). \end{aligned}$$

Here, the first line holds by construction of d^{P}. The third line follows from the subadditivity of φ and the fact that $\boldsymbol{X}\mathbf{e} = \mathbf{1}_S$, and the last line holds by definition of d^{S} and since $I \in \mathbb{N}$.

If φ is subadditive and positive homogeneous, then it is indeed convex, and for any $\emptyset \neq S \subseteq V_C$ and $I \in \mathbb{N}$, Proposition 5 implies that $I \geqslant d^{\mathrm{P}}(S)$ if and only if $\varphi(\mathbf{1}_S/I) \leqslant B$, that is, $\varphi(\mathbf{1}_S) \leqslant I \cdot B$. By definition of d^{S} and since $I \in \mathbb{N}$, we thus have $I \geqslant d^{\mathrm{P}}(S)$ if and only if $I \geqslant d^{\mathrm{S}}(S)$.

To see that $d^{\mathrm{P}} = d^{\mathrm{S}}$ does not hold in general when φ is subadditive but not positive homogeneous, consider the VRP($\mathcal{C}$) instance with $n = 3$ customers, $m = 3$ vehicles and a set $\mathcal{C}$ of the form (4) with $\varphi(x_1, x_2, x_3) = \sqrt{x_1 + x_2 + x_3}$ as well as $B = 1$. Note that φ is subadditive but not positive homogeneous. One readily verifies that $d^{\mathrm{P}}(V_C) = 3$ but $d^{\mathrm{S}}(V_C) = 2$.

In view of statement *(ii)*, we first show that $d^{\mathrm{S}} = d^{\mathrm{P}}$ whenever φ is additive. We know from statement *(i)* that $d^{\mathrm{S}} \leqslant d^{\mathrm{P}}$ in this setting, so we only need to show that $d^{\mathrm{S}} \geqslant d^{\mathrm{P}}$ as well. Imagine, to the contrary, that $d^{\mathrm{S}}(S) = I' < d^{\mathrm{P}}(S)$ for some $\emptyset \neq S \subseteq V_C$. In that case, we have $\varphi(\mathbf{1}_S)/B \leqslant I'$. The additivity of φ implies that $I' \cdot \varphi(\mathbf{1}_S/I') = \varphi(\mathbf{1}_S) \leqslant I' \cdot B$, however, and the solution $(I, \boldsymbol{X}) = (I', \mathbf{1}_S \mathbf{e}^\top / I')$ is feasible in the minimization problem that defines d^{P}. We thus have $d^{\mathrm{P}}(S) \leqslant I'$, which contradicts the assumption that $d^{\mathrm{S}}(S) < d^{\mathrm{P}}(S)$.

When φ is additive, we have $\varphi(\mathbf{1}_S) = \sum_{i \in S} \varphi(\mathbf{e}_i)$. Thus, any VRP($\mathcal{C}$) instance with additive φ can be reformulated as a deterministic CVRP instance with customer demands $q_i = \varphi(\mathbf{e}_i)$ and vehicle capacity $Q = B$. Likewise, one readily verifies that any deterministic CVRP instance can be formulated as an instance of VRP($\mathcal{C}$) with $\varphi(\boldsymbol{x}) = \boldsymbol{q}^\top \boldsymbol{x}$ and $B = Q$.

As for statement *(iii)*, we only need to show that $d^{\mathrm{S}} \leqslant \overline{d}^m$ does not hold in general when φ is not subadditive. Indeed, consider the VRP($\mathcal{C}$) instance with $n = 3$ customers, $m = 3$ vehicles and the set $\mathcal{C}$ of the form (4) with $\varphi(x_1, x_2, x_3) = \exp(x_1 + x_2 + x_3) - 1$ as well as $B = 2$. Note that φ is not subadditive. One readily verifies that $d^{\mathrm{S}}(V_C) = 10$ but $\overline{d}^m(V_C) = 3$. $\square$

Proof of Theorem 4. Fix any $\boldsymbol{x} \in [0,1]^n$. Since f is convex and $\boldsymbol{x} \geqslant \mathbf{0}$, it follows from Theorem 2.2.6(a) of Simchi-Levi et al. (2005) that the mapping $\boldsymbol{q} \mapsto f(\boldsymbol{x}^\top \boldsymbol{q})$ is supermodular. The rectangularity of $\mathcal{P}$ then allows us to re-express the worst-case expectation as

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \right] \;=\; \sup_{\substack{\mathbb{P}_w \in \mathcal{P}^w: \\ w \in \mathcal{W}}} \sum_{w \in \mathcal{W}} \hat{s}_w \cdot \mathbb{E}_{\mathbb{P}_w} \left[f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \right] \;=\; \sum_{w \in \mathcal{W}} \hat{s}_w \cdot \sup_{\mathbb{P}_w \in \mathcal{P}^w} \mathbb{E}_{\mathbb{P}_w} \left[f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \right], \tag{8}$$

where $\mathcal{P}^w$ is defined in (7) and the first identity holds because $\mathcal{S}$ is a singleton set.

We can then apply Proposition 3 of Long et al. (2020) to evaluate $\sup_{\mathbb{P}_w \in \mathcal{P}^w} \mathbb{E}_{\mathbb{P}_w} \left[f\left(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}\right) \right]$ for each $w \in \mathcal{W}$. Note that this proposition assumes that the function inside the worst-case expectation constitutes the second-stage cost of a two-stage distributionally robust optimization problem; since the proof of that result only makes use of the supermodularity of the second-stage cost function, however, the proposition extends to our setting. We thus conclude that $\sup_{\mathbb{P}_w \in \mathcal{P}^w} \mathbb{E}_{\mathbb{P}_w} \left[f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \right] = \sum_{j=1}^{2n+1} p^\star_{wj} \cdot f(\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj})$, where $p^\star_{wj}, \boldsymbol{q}^\star_{wj}$ are obtained from Algorithm 1, $j = 1, \ldots, 2n+1$ and $w \in \mathcal{W}$. Combining this with (8), we obtain

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \right] \;=\; \sum_{w \in \mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot f(\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj}),$$

which implies the statement of the theorem. $\square$

Many of the results from Section 5.1 rely on properties that the worst-case risk measure φ inherits from its constituent risk measures $\rho_{\mathbb{P}}$, $\mathbb{P} \in \mathcal{P}$. We summarize those findings next.

Lemma 3. *The worst-case risk measure φ is* (i) *monotonic,* (ii) *positive homogeneous,* (iii) *subadditive or* (iv) *convex whenever each of its constituent risk measures $\rho_{\mathbb{P}}$, $\mathbb{P} \in \mathcal{P}$, is.*

Proof. In view of *(i)*, we note that any $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ with $\boldsymbol{x} \leqslant \boldsymbol{y}$ satisfy

$$\begin{aligned}
\rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \;\; \forall \mathbb{P} \in \mathcal{P} \quad &\Longrightarrow \quad \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \;\; \forall \mathbb{P} \in \mathcal{P} \\
&\Longleftrightarrow \quad \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \quad \Longleftrightarrow \quad \varphi(\boldsymbol{x}) \;\leqslant\; \varphi(\boldsymbol{y}),
\end{aligned}$$

where the first inequality follows from the monotonicity of $\rho_{\mathbb{P}}$ for all $\mathbb{P} \in \mathcal{P}$.

As for *(ii)*, we observe that any $\lambda \geqslant 0$ and $\boldsymbol{x} \in [0,1]^n$ satisfy

$$\varphi(\lambda \boldsymbol{x}) \;=\; \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\lambda \cdot \boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \lambda \cdot \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \lambda \cdot \varphi(\boldsymbol{x}),$$

where the second identity follows from the positive homogeneity of $\rho_{\mathbb{P}}$ for all $\mathbb{P} \in \mathcal{P}$.

In view of *(iii)*, we observe that any $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ satisfy

$$\begin{aligned}
& \rho_{\mathbb{P}}((\boldsymbol{x}+\boldsymbol{y})^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) + \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) && \forall \mathbb{P} \in \mathcal{P} \\
\Longrightarrow\quad & \rho_{\mathbb{P}}((\boldsymbol{x}+\boldsymbol{y})^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \sup_{\mathbb{P} \in \mathcal{P}} \left\{ \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}} + \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \right\} && \forall \mathbb{P} \in \mathcal{P} \\
\Longrightarrow\quad & \rho_{\mathbb{P}}((\boldsymbol{x}+\boldsymbol{y})^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) + \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) && \forall \mathbb{P} \in \mathcal{P} \\
\Longleftrightarrow\quad & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}((\boldsymbol{x}+\boldsymbol{y})^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) + \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \\
\Longleftrightarrow\quad & \varphi(\boldsymbol{x}+\boldsymbol{y}) \;\leqslant\; \varphi(\boldsymbol{x}) + \varphi(\boldsymbol{y}),
\end{aligned}$$

where the first inequality follows from the subadditivity of $\rho_{\mathbb{P}}$ for all $\mathbb{P} \in \mathcal{P}$ and the second implication is due to the subadditivity of the supremum operator, respectively.

As for *(iv)*, finally, we note that any $\lambda \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ satisfy

$$\begin{aligned}
& \rho_{\mathbb{P}}(\lambda \boldsymbol{x}^\top \tilde{\boldsymbol{q}} + (1-\lambda) \boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \lambda \cdot \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) + (1-\lambda) \cdot \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) && \forall \mathbb{P} \in \mathcal{P} \\
\Longrightarrow\quad & \rho_{\mathbb{P}}(\lambda \boldsymbol{x}^\top \tilde{\boldsymbol{q}} + (1-\lambda) \boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \sup_{\mathbb{P} \in \mathcal{P}} \left\{ \lambda \cdot \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) + (1-\lambda) \cdot \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \right\} && \forall \mathbb{P} \in \mathcal{P} \\
\Longrightarrow\quad & \rho_{\mathbb{P}}(\lambda \boldsymbol{x}^\top \tilde{\boldsymbol{q}} + (1-\lambda) \boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \lambda \cdot \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) + (1-\lambda) \cdot \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) && \forall \mathbb{P} \in \mathcal{P} \\
\Longleftrightarrow\quad & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\lambda \boldsymbol{x}^\top \tilde{\boldsymbol{q}} + (1-\lambda) \boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \;\leqslant\; \lambda \cdot \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) + (1-\lambda) \cdot \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}}) \\
\Longleftrightarrow\quad & \varphi(\lambda \boldsymbol{x} + (1-\lambda) \boldsymbol{y}) \;\leqslant\; \lambda \cdot \varphi(\boldsymbol{x}) + (1-\lambda) \cdot \varphi(\boldsymbol{y}),
\end{aligned}$$

where the first inequality follows from the convexity of $\rho_{\mathbb{P}}$ for all $\mathbb{P} \in \mathcal{P}$ and the second implication is due to the subadditivity of the supremum operator, respectively. □

We split the proof of Theorem 5 into the following five lemmas.

Lemma 4 (Expected Disutility)**.** *Fix an ambiguity set $\mathcal{P}$ of the form* (6) *with $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$. The worst-case expected disutility $\varphi_{\text{ED}}(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-ED}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})$ with*

$$\mathbb{P}\text{-ED}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \mathbb{E}_{\mathbb{P}} \left[U(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \right],$$

where the disutility function U is monotonically non-decreasing and convex with $U(0) \geqslant 0$, affords a $W(2n+1)$-point worst-case distribution that can be computed with Algorithm 1 and that is independent of $\boldsymbol{x}$. Moreover, φ_{ED} is monotone, convex and not *subadditive.*

Proof. The first part of the statement directly follows from Theorem 4, which applies since U is convex. In view of the second part, Lemma 3 implies that φ inherits monotonicity and convexity from $\mathbb{E}_{\mathbb{P}}[U(\cdot)]$, $\mathbb{P} \in \mathcal{P}$. To see that φ_{ED} is not subadditive, consider an ambiguity set of the form (6) with $W = 2$, $\hat{\boldsymbol{s}} = (0.3, 0.7)^\top$, $\underline{\boldsymbol{q}}^1 = \overline{\boldsymbol{q}}^1 = (5, 7)^\top$ and $\underline{\boldsymbol{q}}^2 = \overline{\boldsymbol{q}}^2 = (6, 3)^\top$, and assume for ease of exposition that the expectation and mean absolute deviation constraints in the definition of $\mathcal{P}$ are absent. For $U(x) = x^2$, $\boldsymbol{x} = (1, 0)^\top$ and $\boldsymbol{y} = (0, 1)^\top$, we have

$$\underbrace{\varphi_{\text{ED}}(\boldsymbol{x})}_{=\,32.7} + \underbrace{\varphi_{\text{ED}}(\boldsymbol{y})}_{=\,21} \;<\; \underbrace{\varphi_{\text{ED}}(\boldsymbol{x}+\boldsymbol{y})}_{=\,99.9},$$

which shows that φ_{ED} is indeed not subadditive. □

Lemma 5 (Essential Riskiness Index)**.** *Fix an ambiguity set $\mathcal{P}$ of the form* (6) *with $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$. The essential riskiness index φ_{ERI} with*

$$\varphi_{\text{ERI}}(\boldsymbol{x}) \;=\; \inf\left\{\alpha \geqslant 0 \;:\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\max\left\{\boldsymbol{x}^\top\tilde{\boldsymbol{q}} - \overline{\rho},\; -\alpha\right\}\right] \leqslant 0\right\},$$

where $\overline{\rho}$ is the acceptable demand threshold, can be computed in time $\mathcal{O}(n^2W + nW\log NW)$. Moreover, φ_{ERI} is monotone, convex and not *subadditive.*

Proof. Note that the worst-case expectation embedded in the expression for φ_{ERI} satisfies the conditions of Theorem 4 since the mapping $x \mapsto \max\{x, -\alpha\}$ is convex. We can thus express the essential riskiness index as

$$\varphi_{\text{ERI}}(\boldsymbol{x}^\top\tilde{\boldsymbol{q}}) \;=\; \inf\left\{\alpha \geqslant 0 \;:\; \sum_{w\in\mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot \max\left\{\boldsymbol{x}^\top\boldsymbol{q}^\star_{wj} - \overline{\rho},\; -\alpha\right\} \;\leqslant\; 0\right\},$$

where $p^\star_{wj}$ and $\boldsymbol{q}^\star_{wj}$, $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$, do not depend on α or $\boldsymbol{x}$. The expression

$$\sum_{w\in\mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot \max\left\{\boldsymbol{x}^\top\boldsymbol{q}^\star_{wj} - \overline{\rho},\; -\alpha\right\} \tag{9}$$

is piecewise affine and monotonically non-increasing in α with breakpoints $\overline{\rho} - \boldsymbol{x}^\top\boldsymbol{q}^\star_{wj}$, $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$. We can calculate these breakpoints in time $\mathcal{O}(n^2W)$, sort them in time $\mathcal{O}(nW\log nW)$ and conduct a binary search over them to determine the smallest root of the expression (9). The binary search requires $\mathcal{O}(\log nW)$ iterations, and the evaluation of (9) in each iteration requires time $\mathcal{O}(nW)$. Note that since the worst-case distribution $\mathbb{P}^\star$ is independent of α and $\boldsymbol{x}$, the parameters $p^\star_{wj}$ and $\boldsymbol{q}^\star_{wj}$ can be determined once per 2VF(d) instance.

Convexity of φ_{ERI} follows from Proposition 3 of Zhang et al. (2019). One readily verifies that φ_{ERI} is monotone since $\tilde{\boldsymbol{q}} \geqslant \mathbf{0}$ $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}$. To see that φ_{ERI} is not subadditive, finally, consider an ambiguity set of the form (6) with $W = 2$, $\hat{\boldsymbol{s}} = (0.2, 0.8)^\top$, $\underline{\boldsymbol{q}}^1 = \overline{\boldsymbol{q}}^1 = (5, 10)^\top$ and $\underline{\boldsymbol{q}}^2 = \overline{\boldsymbol{q}}^2 = (10, 5)^\top$, and assume for ease of exposition that the expectation and mean absolute deviation constraints in the definition of $\mathcal{P}$ are absent. For $\overline{\rho} = 12$, $\boldsymbol{x} = (1, 0)^\top$ and $\boldsymbol{y} = (0, 1)^\top$, we have

$$\underbrace{\varphi_{\text{ERI}}(\boldsymbol{x})}_{=0} + \underbrace{\varphi_{\text{ERI}}(\boldsymbol{y})}_{=0} \;<\; \underbrace{\varphi_{\text{ERI}}(\boldsymbol{x} + \boldsymbol{y})}_{=+\infty},$$

which shows that φ_{ERI} is indeed not subadditive. □

Lemma 6 (Expectiles)**.** *Fix an ambiguity set $\mathcal{P}$ of the form (6) with $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$. The worst-case expectile risk measure φ_{E} with*

$$\varphi_{\text{E}}(\boldsymbol{x}) \;=\; \operatorname*{arg\,min}_{u \in \mathbb{R}} \left\{ \alpha \cdot \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - u\right]_+^2 \right] + (1 - \alpha) \cdot \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left[u - \boldsymbol{x}^\top \tilde{\boldsymbol{q}}\right]_+^2 \right] \right\},$$

where $\alpha \in [1/2, 1)$, can be computed in time $\mathcal{O}(n^2 W + nW \log nW)$. Moreover, φ_{E} is monotone, convex and subadditive.

Proof. Both worst-case expectations in the definition of φ_{E} satisfy the conditions of Theorem 4, which implies that the expression inside the minimum defining φ_{E} simplifies to

$$\sum_{w \in \mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \left[\alpha \left(\left[\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj} - u\right]_+ \right)^2 + (1 - \alpha) \left(\left[u - \boldsymbol{x}^\top \boldsymbol{q}^\star_{wj}\right]_+ \right)^2 \right], \tag{10}$$

where $p^\star_{wj}$ and $\boldsymbol{q}^\star_{wj}$, $j = 1, \ldots, 2n+1$ and $w \in \mathcal{W}$, do not depend on u or $\boldsymbol{x}$. The expression (10) is piecewise affine and convex in u with breakpoints $\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj}$, $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$. We can calculate these breakpoints in time $\mathcal{O}(n^2 W)$, sort them in time $\mathcal{O}(nW \log nW)$ and conduct a trisection search over them to determine a value of u that minimizes (10). The trisection search requires $\mathcal{O}(\log nW)$ iterations, and the evaluation of (10) in each iteration requires time $\mathcal{O}(nW)$ since the breakpoints have been computed previously. Note that since the worst-case distribution $\mathbb{P}^\star$ is independent of u and $\boldsymbol{x}$ by Theorem 4, the parameters $p^\star_{wj}$ and $\boldsymbol{q}^\star_{wj}$ can be determined once per 2VF(d) instance.

The above discussion implies that

$$\varphi_{\text{E}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \operatorname*{arg\,min}_{u \in \mathbb{R}} \left\{ \mathbb{E}_{\mathbb{P}^\star} \left[\alpha \left(\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - u\right]_+ \right)^2 \right] + \mathbb{E}_{\mathbb{P}^\star} \left[(1 - \alpha) \left(\left[u - \boldsymbol{x}^\top \tilde{\boldsymbol{q}}\right]_+ \right)^2 \right] \right\}$$

for the worst-case distribution $\mathbb{P}^\star$ of Theorem 4. Proposition 6 of Bellini et al. (2014) then implies that φ_{E} is coherent and thus, *a fortiori*, monotone, convex and subadditive. □

Lemma 7 (Entropic Risk)**.** *Fix an ambiguity set* $\mathcal{P}$ *of the form* (6) *with* $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$*. The worst-case entropic risk* $\varphi_{\text{ent}}(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-ent}(\boldsymbol{x}^\top\tilde{\boldsymbol{q}})$ *with*

$$\mathbb{P}\text{-ent}\big(\boldsymbol{x}^\top\tilde{\boldsymbol{q}}\big) \;=\; \frac{1}{\theta}\log\mathbb{E}_{\mathbb{P}}\left[\exp\left(\theta\cdot\boldsymbol{x}^\top\tilde{\boldsymbol{q}}\right)\right],$$

where $\theta > 0$*, affords a* $W(2n{+}1)$*-point worst-case distribution that can be computed with Algorithm 1 and that is independent of* $\boldsymbol{x}$*. Moreover,* φ_{ent} *is monotone, convex and* not *subadditive.*

Proof. Since $x \mapsto \log(x)$ is monotonically increasing, we can exchange the order of the supremum and logarithm operators in the definition of φ_{ent} and conclude that

$$\varphi_{\text{ent}}(\boldsymbol{x}) \;=\; \frac{1}{\theta}\log\left(\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[\exp\left(\theta\cdot\boldsymbol{x}^\top\tilde{\boldsymbol{q}}\right)\right]\right).$$

The worst-case expectation embedded in the above expression satisfies the conditions of Theorem 4, which implies the first part of the statement.

By Definition 2.3 of Föllmer and Knispel (2011), $\mathbb{P}$-ent is monotone and convex for every $\mathbb{P}\in\mathcal{P}$, and Lemma 3 implies that both properties carry over to the worst-case entropic risk φ_{ent}. To see that φ_{ent} is not subadditive, finally, fix $\theta = 1$ and consider the ambiguity set from the proof of Lemma 4 together with $\boldsymbol{x} = (1,0)^\top$ and $\boldsymbol{y} = (0,1)^\top$. We then have

$$\underbrace{\varphi_{\text{ent}}(\boldsymbol{x})}_{=\,5.79} + \underbrace{\varphi_{\text{ent}}(\boldsymbol{y})}_{=\,5.84} \;<\; \underbrace{\varphi_{\text{ent}}(\boldsymbol{x}+\boldsymbol{y})}_{=\,17.81},$$

which shows that φ_{ent} is indeed not subadditive. □

Lemma 8 (Requirements Violation Index)**.** *Fix an ambiguity set* $\mathcal{P}$ *of the form* (6) *with* $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$*. The requirements violation index* φ_{RV} *with*

$$\varphi_{\text{RV}}(\boldsymbol{x}) \;=\; \inf\left\{\alpha \geqslant 0 \,:\, C_\alpha(\boldsymbol{x}^\top\tilde{\boldsymbol{q}}) \leqslant \overline{\rho}\right\},$$

where C_α *is the worst-case certainty equivalent under an exponential disutility,*

$$C_\alpha(\boldsymbol{x}^\top\tilde{\boldsymbol{q}}) = \begin{cases} \displaystyle\sup_{\mathbb{P}\in\mathcal{P}} \alpha\log\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{\boldsymbol{x}^\top\tilde{\boldsymbol{q}}}{\alpha}\right)\right] & \text{if } \alpha > 0 \\ \displaystyle\lim_{\gamma\to 0} C_\gamma(\boldsymbol{x}^\top\tilde{\boldsymbol{q}}) & \text{if } \alpha = 0, \end{cases}$$

and $\overline{\rho}$ *is the acceptable demand threshold, can be computed to* ϵ*-accuracy via bisection search. Moreover,* φ_{RV} *is monotone, convex and* not *subadditive.*

Proof. Following similar arguments as in the proof of Lemma 7, one can show that the worst-case certainty equivalent C_α can be computed for any fixed value of α in time $\mathcal{O}(nW)$, with an initial computation of time $\mathcal{O}(n^2W)$ to compute the worst-case distribution $\mathbb{P}^\star$ as well as the expressions $\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj}$. Note that the worst-case certainty equivalent C_α is monotonically non-increasing in α (Jaillet et al., 2016, Lemma 1). We can thus conduct a bisection search to determine the smallest value of α that satisfies $C_\alpha(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \leqslant \overline{\rho}$. The bisection search can be started with the lower bound $\underline{\alpha} = \epsilon$, where ϵ is a sufficiently small positive quantity, and any upper bound $\overline{\alpha}$ satisfying $\overline{\alpha} \geqslant \boldsymbol{x}^\top \boldsymbol{q}^\star_{wj}$ for all $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$ as well as

$$\overline{\alpha} \;\geqslant\; \frac{(\exp(1) - 2) \displaystyle\sum_{w \in \mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot (\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj})^2}{\overline{\rho} - \displaystyle\sum_{w \in \mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot \boldsymbol{x}^\top \boldsymbol{q}^\star_{wj}},$$

which guarantees that $C_{\overline{\alpha}}(\boldsymbol{x}^\top \boldsymbol{q}) \leqslant \overline{\rho}$. Details are omitted for the sake of brevity.

As for the second part of the statement, the convexity of φ_{RV} follows from Proposition 1 of Jaillet et al. (2016). One readily verifies that φ_{RV} is monotone since $\tilde{\boldsymbol{q}} \geqslant \boldsymbol{0}$ $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}$. To see that φ_{RV} is not subadditive, finally, fix $\overline{\rho} = 12$ and consider the ambiguity set from the proof of Lemma 5 together with $\boldsymbol{x} = (1, 0)^\top$ and $\boldsymbol{y} = (0, 1)^\top$. We then have

$$\underbrace{\varphi_{\text{RV}}(\boldsymbol{x})}_{=0} + \underbrace{\varphi_{\text{RV}}(\boldsymbol{y})}_{=0} \;<\; \underbrace{\varphi_{\text{RV}}(\boldsymbol{x} + \boldsymbol{y})}_{=+\infty},$$

which shows that φ_{RV} is indeed not subadditive. □

Proof of Theorem 5. The proof directly follows from the Lemmas 4–8. □

We split the proof of Theorem 6 into the following two lemmas.

Lemma 9 (CVaR)**.** *Fix an ambiguity set $\mathcal{P}$ of the form* (6) *with $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$. The worst-case conditional value-at-risk at level $1 - \epsilon$, $\varphi_{\text{CVaR}}(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_{1-\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})$ with*

$$\mathbb{P}\text{-CVaR}_{1-\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \inf_{u \in \mathbb{R}} \; u + \frac{1}{1-\epsilon} \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - u \right]_+ ,$$

where $\epsilon \in [0, 1)$, can be computed in time $\mathcal{O}(n^2 W + nW \log nW)$. Moreover, φ_{CVaR} is monotone, convex and subadditive.

Proof. Proposition 3.1 of Shapiro and Kleywegt (2002) implies that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_{1-\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \sup_{\mathbb{P}\in\mathcal{P}} \inf_{u\in\mathbb{R}} u + \frac{1}{1-\epsilon}\mathbb{E}_{\mathbb{P}}[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - u]_+ \;=\; \inf_{u\in\mathbb{R}} u + \frac{1}{1-\epsilon} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - u]_+.$$

Since the worst-case expectation in the right-most expression above satisfies the conditions of Theorem 4, the worst-case CVaR further simplifies to

$$\inf_{u\in\mathbb{R}} u + \frac{1}{1-\epsilon} \sum_{w\in\mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot \left[\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj} - u\right]_+ ,$$

where $p^\star_{wj}$ and $\boldsymbol{q}^\star_{wj}$, $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$, do not depend on u or $\boldsymbol{x}$. The function inside the above minimization is piecewise affine and convex in u with breakpoints $\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj}$, $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$. We can thus obtain a minimizer by computing the breakpoints in time $\mathcal{O}(n^2 W)$, sorting the breakpoints in time $\mathcal{O}(nW \log nW)$ and performing a trisection search over the breakpoints. The trisection search requires $\mathcal{O}(\log nW)$ iterations, and each iteration requires time $\mathcal{O}(nW)$ since the breakpoints have been computed previously.

By Corollary 12 of Rockafeller and Uryasev (2002), each constituent risk measure $\mathbb{P}$-CVaR, $\mathbb{P} \in \mathcal{P}$, is coherent and thus, *a fortiori*, monotone, convex and subadditive. Lemma 3 then implies that these properties carry over to the worst-case CVaR φ_{CVaR}. □

Lemma 10 (Service Fulfilment Risk Index)**.** *Fix an ambiguity set $\mathcal{P}$ of the form* (6) *with $\mathcal{S} = \{\hat{\boldsymbol{s}}\}$. The service fulfillment risk index φ_{SRI} with*

$$\varphi_{\text{SRI}}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \inf\left\{\alpha \geqslant 0 \,:\, \varphi_{\text{CVaR}}\left(\max\left\{\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho},\, -\alpha\right\}\right) \leqslant 0\right\},$$

where $\overline{\rho}$ is the acceptable demand threshold and the worst-case CVaR is evaluated at level $1-\gamma$ with γ being the the service level, can be computed in time $\mathcal{O}(n^2 W + nW \log nW)$. Moreover, φ_{SRI} is monotone, convex and not *subadditive.*

Proof. Theorem 1 of Zhang et al. (2021) allows us to equivalently express φ_{SRI} as

$$\varphi_{\text{SRI}}(\boldsymbol{x}) \;=\; \inf\left\{\alpha \geqslant 0 \,:\, \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho} + \alpha\right]_+ \leqslant (1-\gamma)\alpha\right\},$$

The worst-case expectation embedded in this expression satisfies the conditions of Theorem 4, and the overall expression thus simplifies to

$$\varphi_{\text{SRI}}(\boldsymbol{x}) \;=\; \inf\left\{\alpha \geqslant 0 \,:\, \sum_{w\in\mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot \left[\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj} - \overline{\rho} + \alpha\right]_+ \leqslant (1-\gamma)\alpha\right\},$$

where $p^\star_{wj}$ and $\boldsymbol{q}^\star_{wj}$, $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$, do not depend on α or $\boldsymbol{x}$. The expression

$$\sum_{w \in \mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot \left[\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj} - \overline{\rho} + \alpha\right]_+ - (1-\gamma)\alpha$$

is piecewise affine and convex in α with breakpoints $\overline{\rho} - \boldsymbol{x}^\top \boldsymbol{q}^\star_{wj}$, $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$. We can thus obtain the smallest root of this expression by computing the breakpoints in time $\mathcal{O}(n^2 W)$, sorting the breakpoints in time $\mathcal{O}(nW \log nW)$ and performing a trisection search over these breakpoints. The trisection search requires $\mathcal{O}(\log nW)$ iterations, and each iteration requires time $\mathcal{O}(nW)$ since the breakpoints have been computed previously.

Proposition 1 of Zhang et al. (2021) implies that φ_{SRI} is convex. Moreover, only readily verifies that φ_{SRI} is monotone. To see that φ_{SRI} is not subadditive, finally, fix $\gamma = 1$ and $\overline{\rho} = 12$ and consider the ambiguity set from the proof of Lemma 5 together with $\boldsymbol{x} = (1,0)^\top$ and $\boldsymbol{y} = (0,1)^\top$. We then have

$$\underbrace{\rho_{\text{SRI}}(\boldsymbol{x})}_{=0} + \underbrace{\rho_{\text{SRI}}(\boldsymbol{y})}_{=0} < \underbrace{\rho_{\text{SRI}}(\boldsymbol{x}+\boldsymbol{y})}_{=+\infty},$$

which shows that ρ_{SRI} is indeed not subadditive. □

Proof of Theorem 6. The proof directly follows from the Lemmas 9 and 10. □

Proof of Theorem 7. By assumption, each risk measure $\psi_{\mathbb{P}}$, $\mathbb{P} \in \mathcal{P}$, can be expressed as the expectation of a convex function and thus satisfies the conditions of Theorem 4. We thus obtain that

$$\varphi_{\text{URI}}(\boldsymbol{x}) = \inf \left\{ \frac{1}{\alpha} : \sum_{w \in \mathcal{W}} \hat{s}_w \sum_{j=1}^{2n+1} p^\star_{wj} \cdot f(\alpha \cdot [\boldsymbol{x}^\top \boldsymbol{q}^\star_{wj} - \overline{\rho}]) \leqslant 0, \;\; \alpha > 0 \right\},$$

where $p^\star_{wj}$ and $\boldsymbol{q}^\star_{wj}$, $w \in \mathcal{W}$ and $j = 1, \ldots, 2n+1$, do not depend on α or $\boldsymbol{x}$. Note that the summation on the left-hand side of the first inequality above inherits convexity from f, and we can thus conduct a bisection search to determine the maximum α that satisfies the inequality. The bisection search can be started with the lower bound $\underline{\alpha} = \epsilon$, where ϵ is a sufficiently small positive quantity, and any upper bound $\overline{\alpha}$ satisfying

$$\psi_{\overline{\mathbb{P}}}\left(\overline{\alpha} \cdot [\mathbf{e}^\top \tilde{\boldsymbol{q}} - \overline{\rho}]\right) \leqslant 0 \quad \text{for } \overline{\mathbb{P}} \text{ satisfying } \overline{\mathbb{P}}\left[\tilde{q}_i = \max_{w \in \mathcal{W}} \overline{q}^w_i\right] = 1,\, i \in V_C,$$

which can be determined once per 2VF(d) instance via bisection search.

By Definition 3 of Hall et al. (2015), φ_{URI} is monotone, convex and positive homogeneous, and convexity and positive homogeneity of φ_{URI} imply that φ_{URI} is subadditive as well. □

The proof of Proposition 6 relies on five auxiliary lemmas which we state and prove first.

Lemma 11. *For the* 1*-norm ambiguity set, the function* $\boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$ *in Proposition 6 can be maximized and minimized over* $\boldsymbol{s} \in \mathcal{S}$ *in time* $\mathcal{O}(W \log W)$*.*

Proof. The problem amounts to solving

$$\begin{aligned} \max / \min \quad & \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}) \\ \text{subject to} \quad & \|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_1 \leqslant \theta \\ & \mathbf{e}^\top \boldsymbol{s} = 1, \;\; \boldsymbol{s} \in \mathbb{R}^W_+, \end{aligned}$$

and this problem has been studied in the literature (Petrik and Subramanian, 2014, Theorem 3.2). In the remainder of the proof, we simplify the exposition and focus on the maximization case; the minimization problem can be solved by a straightforward adaptation of the arguments below.

The core idea behind the algorithm is to start with the initial solution $\boldsymbol{s} = \hat{\boldsymbol{s}}$ and then iteratively shift probability mass from the smallest components of $\boldsymbol{\pi}(\boldsymbol{x})$ to the largest one (respecting non-negativity of all probability weights) until the uncertainty budget θ has been exhausted. This algorithm requires the components of $\boldsymbol{\pi}(\boldsymbol{x})$ to be sorted in ascending order. This sorting, which dominates the runtime of the algorithm, can be achieved in time $\mathcal{O}(W \log W)$. □

Lemma 12. *For the* ∞*-norm ambiguity set, the function* $\boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$ *in Proposition 6 can be maximized and minimized over* $\boldsymbol{s} \in \mathcal{S}$ *in time* $\mathcal{O}(W \log W)$*.*

Proof. The problem amounts to solving

$$\begin{aligned} \max / \min \quad & \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}) \\ \text{subject to} \quad & \|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_\infty \leqslant \theta \\ & \mathbf{e}^\top \boldsymbol{s} = 1, \;\; \boldsymbol{s} \in \mathbb{R}^W_+, \end{aligned}$$

and this problem has been studied in the literature (Megiddo and Ichimori, 1985, page 3). In the remainder of the proof, we simplify the exposition and focus on the maximization case; the minimization problem can be solved by a straightforward adaptation of the arguments below.

The core idea behind the algorithm is to start with the initial solution $\boldsymbol{s} = [\hat{\boldsymbol{s}} - \theta \mathbf{e}]_+$ and then iteratively increase these probability weights, starting with the weight corresponding to the largest component of $\boldsymbol{\pi}(\boldsymbol{x})$ and moving towards the weight corresponding to the smallest component of

$\boldsymbol{\pi}(\boldsymbol{x})$, until either $\mathbf{e}^\top \boldsymbol{s} = 1$ or the uncertainty budget θ has been exhausted for the particular weight. This algorithm requires the components of $\boldsymbol{\pi}(\boldsymbol{x})$ to be sorted in ascending order. This sorting, which dominates the runtime of the algorithm, can be achieved in time $\mathcal{O}(W \log W)$. □

Lemma 13. *For the axis-parallel ellipsoidal ambiguity set, the function $\boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$ in Proposition 6 can be maximized and minimized over $\boldsymbol{s} \in \mathcal{S}$ to ϵ-accuracy in time $\mathcal{O}(W \log W \cdot \log \epsilon^{-1})$.*

Lemma 2 of Pessoa and Poss (2015) and Corollary 4 of Ghosal and Wiesemann (2020) study the related problems of maximizing a linear function over the intersection of a 2-norm ball with an ∞-norm ball as well as the intersection of an axis-parallel ellipsoid with a hyperrectangle, respectively. Both of those approaches consider the Lagrange relaxation of the ∞-norm and hyperrectangle constraints (while keeping intact the ellipsoidal constraint), which admits a closed-form solution. In our context, this approach would require the dualization of both the non-negativity and the probability simplex constraints, which would result in two sets of Lagrange multipliers that appear difficult to handle. Instead, our proof below relies on a reduction of the dual problem.

Proof of Lemma 13. The problem amounts to solving

$$\begin{aligned} \max/\min \quad & \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}) \\ \text{subject to} \quad & (\boldsymbol{s} - \hat{\boldsymbol{s}})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{s} - \hat{\boldsymbol{s}}) \leqslant \theta \\ & \mathbf{e}^\top \boldsymbol{s} = 1, \quad \boldsymbol{s} \in \mathbb{R}^W_+. \end{aligned} \tag{11}$$

In the remainder of the proof, we simplify the exposition by suppressing the dependence of $\boldsymbol{\pi}(\boldsymbol{x})$ on $\boldsymbol{x}$ and focusing on the maximization case; the minimization problem can be solved by a straightforward adaptation of the arguments below.

Strong convex duality, which is guaranteed by the existence of a Slater point due to the assumptions $\theta > 0$ and $\boldsymbol{\Sigma} \succ \mathbf{0}$, implies that

$$\begin{aligned} & \max_{\boldsymbol{s} \in \mathbb{R}^W} \min_{\substack{\alpha \in \mathbb{R}_+,\, \beta \in \mathbb{R}, \\ \boldsymbol{\gamma} \in \mathbb{R}^W_+}} \boldsymbol{s}^\top \boldsymbol{\pi} + \alpha \left[\theta - (\boldsymbol{s} - \hat{\boldsymbol{s}})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{s} - \hat{\boldsymbol{s}})\right] + \beta \left(\mathbf{e}^\top \boldsymbol{s} - 1\right) + \boldsymbol{\gamma}^\top \boldsymbol{s} \\ = & \min_{\substack{\alpha \in \mathbb{R}_+,\, \beta \in \mathbb{R}, \\ \boldsymbol{\gamma} \in \mathbb{R}^W_+}} \max_{\boldsymbol{s} \in \mathbb{R}^W} -\alpha \boldsymbol{s}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{s} + \boldsymbol{s}^\top \left(\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma} + 2\alpha \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{s}}\right) + \alpha \left(\theta - \hat{\boldsymbol{s}}^\top \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{s}}\right) - \beta, \end{aligned} \tag{12}$$

where the dual multipliers α, β and $\boldsymbol{\gamma}$ correspond to the ellipsoidal, the probability simplex as well as the non-negativity constraints in (11), respectively. We conduct a case distinction that determines the best solutions to problem (12) under the additional constraint that $\alpha = 0$ or $\alpha > 0$, respectively. The lower of the two corresponding optimal values then coincides with the optimal value of problem (12), which in turn is equal to the optimal value of problem (11).

Under the additional constraint that $\alpha = 0$, problem (12) reduces to

$$\begin{array}{ll} \text{minimize} & -\beta \\ \text{subject to} & \boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma} = \mathbf{0} \\ & \beta \in \mathbb{R}, \ \ \boldsymbol{\gamma} \in \mathbb{R}^W_+. \end{array}$$

Eliminating the slack variables $\boldsymbol{\gamma}$ from this problem, the first constraint becomes $-\beta \geqslant \max\{\pi_w : w \in \mathcal{W}\}$, and the optimal objective value is thus readily identified as $\max\{\pi_w : w \in \mathcal{W}\}$.

Under the additional constraint that $\alpha > 0$, the first-order necessary optimality conditions of the maximization problem embedded in (12) imply that

$$-2\alpha \boldsymbol{\Sigma}^{-1} \boldsymbol{s}^\star + \boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma} + 2\alpha \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{s}} = \mathbf{0} \quad \Longleftrightarrow \quad \boldsymbol{s}^\star = \frac{1}{2\alpha} \boldsymbol{\Sigma}(\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma} + 2\alpha \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{s}}).$$

Substituting this solution into the outer minimization problem in (12), we obtain

$$\begin{aligned} & \min_{\substack{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \\ \boldsymbol{\gamma} \in \mathbb{R}^W_+}} \frac{1}{4\alpha} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma} + 2\alpha \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{s}})^\top \boldsymbol{\Sigma} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma} + 2\alpha \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{s}}) + \alpha \left(\theta - \hat{\boldsymbol{s}}^\top \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{s}} \right) - \beta \\ = & \min_{\substack{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \\ \boldsymbol{\gamma} \in \mathbb{R}^W_+}} \frac{1}{4\alpha} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma})^\top \boldsymbol{\Sigma} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma}) + (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma})^\top \hat{\boldsymbol{s}} + \alpha \theta - \beta \\ = & \min_{\substack{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \\ \boldsymbol{\gamma} \in \mathbb{R}^W_+}} \frac{1}{4\alpha} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma})^\top \boldsymbol{\Sigma} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma}) + (\boldsymbol{\pi} + \boldsymbol{\gamma})^\top \hat{\boldsymbol{s}} + \alpha \theta. \end{aligned}$$

Since $\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\sigma}^2)$, the above problem reduces to

$$\begin{aligned} & \min_{\substack{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \\ \boldsymbol{\gamma} \in \mathbb{R}^W_+}} \alpha \theta + \boldsymbol{\pi}^\top \hat{\boldsymbol{s}} + \sum_{w \in \mathcal{W}} \frac{\sigma_w^2}{4\alpha} (\pi_w + \beta + \gamma_w)^2 + \gamma_w \hat{s}_w \\ = & \min_{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}} \alpha \theta + \boldsymbol{\pi}^\top \hat{\boldsymbol{s}} + \sum_{w \in \mathcal{W}} \min_{\gamma_w \in \mathbb{R}_+} \frac{\sigma_w^2}{4\alpha} (\pi_w + \beta + \gamma_w)^2 + \gamma_w \hat{s}_w \\ = & \min_{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}} \alpha \theta + \boldsymbol{\pi}^\top \hat{\boldsymbol{s}} + \sum_{w \in \mathcal{W}} \frac{\sigma_w^2}{4\alpha} \left(\pi_w + \beta + \left[-\pi_w - \beta - \frac{2\alpha}{\sigma_w^2} \hat{s}_w \right]_+ \right)^2 + \left[-\pi_w - \beta - \frac{2\alpha}{\sigma_w^2} \hat{s}_w \right]_+ \hat{s}_w, \end{aligned}$$

where the last equality holds since for any fixed $\alpha \in \mathbb{R}_+$ and $\beta \in \mathbb{R}$, the optimal solution $\boldsymbol{\gamma}^\star$ satisfies $\gamma_w^\star = -\pi_w - \beta - \frac{2\alpha}{\sigma_w^2} \hat{s}_w$ if $\pi_w + \beta + \frac{2\alpha}{\sigma_w^2} \hat{s}_w < 0$ and $\gamma_w^\star = 0$ otherwise. For any fixed $\alpha \in \mathbb{R}_+$, the above optimization problem is convex in β with at most W breakpoints, which can be solved using a trisection search. The trisection search requires $\mathcal{O}(\log W)$ iterations of complexity $\mathcal{O}(W)$ each. By applying an outer trisection over α we obtain the overall complexity $\mathcal{O}(W \log W \cdot \log \epsilon^{-1})$. $\square$

Lemma 14. *For the ellipsoidal ambiguity set, the function $\boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$ in Proposition 6 can be maximized and minimized over $\boldsymbol{s} \in \mathcal{S}$ to ϵ-accuracy in polynomial time via FISTA.*

Theorem 7 of Ghosal and Wiesemann (2020) studies the related problem of maximizing a linear function over the intersection of an ellipsoid with a hyperrectangle. Our proof of Lemma 14 follows a similar strategy as that result: We dualize the optimization problem, simplify the dual through a variable elimination and subsequently solve the simplified problem with FISTA.

Proof of Lemma 14. We focus on the maximization variant and follow the same strategy as in the proof of Lemma 13: We dualize the optimization problem and distinguish the two cases where $\alpha = 0$ and $\alpha > 0$. In the latter case, the dual problem is

$$\min_{\substack{\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \\ \boldsymbol{\gamma} \in \mathbb{R}^W_+}} \frac{1}{4\alpha} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma})^\top \boldsymbol{\Sigma} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma}) + (\boldsymbol{\pi} + \boldsymbol{\gamma})^\top \hat{\boldsymbol{s}} + \alpha \theta,$$

and the first-order necessary optimality conditions as well as the non-negativity of α imply that

$$-\frac{1}{4(\alpha^\star)^2} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma})^\top \boldsymbol{\Sigma} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma}) + \theta = 0 \quad \Longleftrightarrow \quad \alpha^\star = \sqrt{\frac{1}{4\theta} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma})^\top \boldsymbol{\Sigma} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma})}.$$

Eliminating α, the problem thus simplifies to

$$\begin{array}{ll} \text{minimize} & \sqrt{\theta} \left\| \boldsymbol{\Sigma}^{\frac{1}{2}} (\boldsymbol{\pi} + \beta \mathbf{e} + \boldsymbol{\gamma}) \right\|_2 + (\boldsymbol{\pi} + \boldsymbol{\gamma})^\top \hat{\boldsymbol{s}} \\ \text{subject to} & \beta \in \mathbb{R}, \;\; \boldsymbol{\gamma} \in \mathbb{R}^W_+. \end{array}$$

The objective function of this problem constitutes the sum of a non-smooth norm expression and a smooth function of $(\beta, \boldsymbol{\gamma})$. We can solve the problem using FISTA (Beck and Teboulle, 2009) with adaptive restarts (O'Donoghue and Candès, 2015) if we move the non-negativity constraints to the objective function through indicator functions and apply a Moreau proximal smoothing (Beck and Teboulle, 2012) to the norm term in the objective function.

$\square$

Lemma 15. *For the entropy ambiguity set, the function $\boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$ in Proposition 6 can be maximized and minimized over $\boldsymbol{s} \in \mathcal{S}$ to ϵ-accuracy in time $\mathcal{O}(W \log[\overline{\pi}/\epsilon])$, where $\overline{\pi} = \max\{\pi_w(\boldsymbol{x}) : w \in \mathcal{W}\}$.*

Proof. The problem amounts to solving

$$\begin{array}{ll} \max / \min & \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}) \\ \text{subject to} & \sum_{w \in \mathcal{W}} s_W \log\left(\frac{s_w}{\hat{s}_w}\right) \leqslant \theta \\ & \mathbf{e}^\top \boldsymbol{s} = 1, \quad \boldsymbol{s} \in \mathbb{R}^W_+, \end{array}$$

and this problem has been studied in the literature (Nilim and El Ghaoui, 2005, Section 6.2). In the remainder of the proof, we simplify the exposition and focus on the maximization case; the minimization problem can be solved by a straightforward adaptation of the arguments below.

The core idea behind the algorithm is to dualize the problem and simplify it to the form

$$\min_{\lambda > 0} \lambda \log\left[\sum_{w \in \mathcal{W}} \hat{s}_w \exp\left(\frac{\pi_w(\boldsymbol{x})}{\lambda}\right)\right] + \theta\lambda.$$

The objective function of this problem is convex, and we can thus conduct a trisection search to obtain an ϵ-optimal solution $\lambda^\star$. Section 6.3 of Nilim and El Ghaoui (2005) shows that the optimal solution $\lambda^\star$ satisfies $\lambda^\star \leqslant [\overline{\pi} - \boldsymbol{\pi}(\boldsymbol{x})^\top \hat{\boldsymbol{s}}]/\theta$, which implies the stated complexity estimate. □

Proof of Proposition 6. Since f is monotonically increasing, we can first determine the optimal value $\theta^\star$ of the problem $\max_{\boldsymbol{s} \in \mathcal{S}} / \min_{\boldsymbol{s} \in \mathcal{S}} \ \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$ and then compute $\varphi(\boldsymbol{x}) = f(\theta^\star)$. The proof of the statement then follows directly from the Lemmas 11–15. □

Proof of Corollary 1. In view of the expected disutility, we note that the rectangularity of the scenario-wise ambiguity sets $\mathcal{P}^w$, $w \in \mathcal{W}$, allows us to rewrite the risk measure as

$$\varphi_{\text{ED}}(\boldsymbol{x}) \;=\; \max_{\boldsymbol{s} \in \mathcal{S}} \sup_{\substack{\mathbb{P}_w \in \mathcal{P}^w: \\ w \in \mathcal{W}}} \sum_{w \in \mathcal{W}} s_w \cdot \mathbb{E}_{\mathbb{P}_w}\left[U(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})\right] \;=\; \max_{\boldsymbol{s} \in \mathcal{S}} \sum_{w \in \mathcal{W}} s_w \cdot \underbrace{\sup_{\mathbb{P}_w \in \mathcal{P}^w} \mathbb{E}_{\mathbb{P}_w}\left[U(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})\right]}_{= \pi_w(\boldsymbol{x})},$$

which satisfies the conditions of Proposition 6 if we set $f(x) = x$.

As for the entropic risk, similar arguments allow us to rewrite the risk measure as

$$\begin{aligned} \varphi_{\text{ED}}(\boldsymbol{x}) &= \max_{\boldsymbol{s} \in \mathcal{S}} \sup_{\substack{\mathbb{P}_w \in \mathcal{P}^w: \\ w \in \mathcal{W}}} \frac{1}{\theta} \log \sum_{w \in \mathcal{W}} s_w \cdot \mathbb{E}_{\mathbb{P}_w}\left[\exp(\theta \cdot \boldsymbol{x}^\top \tilde{\boldsymbol{q}})\right] \\ &= \max_{\boldsymbol{s} \in \mathcal{S}} \frac{1}{\theta} \log \sum_{w \in \mathcal{W}} s_w \cdot \underbrace{\sup_{\mathbb{P}_w \in \mathcal{P}^w} \mathbb{E}_{\mathbb{P}_w}\left[\exp(\theta \cdot \boldsymbol{x}^\top \tilde{\boldsymbol{q}})\right]}_{= \pi_w(\boldsymbol{x})}, \end{aligned}$$

which satisfies the conditions of Proposition 6 if we set $f(x) = \frac{1}{\theta}\log(x)$. □

Proof of Proposition 7. Under the stated assumptions, one can minimize the convex function

$$f_0(u) + \sum_{\ell=1}^{L} \max_{\boldsymbol{s} \in \mathcal{S}} f_\ell \left(\boldsymbol{s}^\top \boldsymbol{\pi}_\ell(\boldsymbol{x}, u), u \right)$$

over $u \in \mathcal{U}$ in $\mathcal{O}(\log \epsilon^{-1})$ iterations of complexity $\mathcal{O}(LT)$ each using a trisection search. This shows the statement for the first worst-case risk measure φ.

In view of the second worst-case risk measure φ, we can first employ the aforementioned trisection search to minimize the left-hand side of the inequality constraint. This establishes whether or not the minimization problem is feasible, and it provides a lower (if g is monotonically decreasing) or upper (if g is monotonically increasing) bound on the constrained minimizer of g. We can subsequently identify the constrained minimizer of g by a bisection search. Both the trisection and the bisection search require $\mathcal{O}(\log \epsilon^{-1})$ iterations of complexity $\mathcal{O}(LT)$ each. □

Proof of Corollary 2. In view of the essential riskiness index, we note that the rectangularity of the scenario-wise ambiguity sets $\mathcal{P}^w$, $w \in \mathcal{W}$, allows us to rewrite the risk measure as

$$\begin{aligned}
\varphi(\boldsymbol{x}) &= \inf \left\{ u \geqslant 0 \, : \, \max_{\boldsymbol{s} \in \mathcal{S}} \sup_{\substack{\mathbb{P}_w \in \mathcal{P}^w : \\ w \in \mathcal{W}}} \sum_{w \in \mathcal{W}} s_w \cdot \mathbb{E}_{\mathbb{P}_w} \left[\max \left\{ \boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho}, \, -u \right\} \right] \leqslant 0 \right\} \\
&= \inf \left\{ u \geqslant 0 \, : \, \max_{\boldsymbol{s} \in \mathcal{S}} \sum_{w \in \mathcal{W}} s_w \cdot \underbrace{\sup_{\mathbb{P}_w \in \mathcal{P}^w} \mathbb{E}_{\mathbb{P}_w} \left[\max \left\{ \boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho}, \, -u \right\} \right]}_{= \pi_{1w}(\boldsymbol{x}, u)} \leqslant 0 \right\},
\end{aligned}$$

which satisfies the conditions of Proposition 7 if we set $\mathcal{U} = \mathbb{R}_+$, $g(u) = u$, $f_0(u) = 0$, $L = 1$ and $f_1(x, u) = x$.

As for expectiles, similar arguments allow us to rewrite the risk measure as

$$
\begin{aligned}
\varphi(\boldsymbol{x}) &= \underset{u\in\mathbb{R}}{\arg\min}\Bigg\{\alpha\cdot\max_{\boldsymbol{s}\in\mathcal{S}}\sup_{\substack{\mathbb{P}_w\in\mathcal{P}^w:\\ w\in\mathcal{W}}}\sum_{w\in\mathcal{W}} s_w\cdot\mathbb{E}_{\mathbb{P}_w}\left[\left[\boldsymbol{x}^\top\tilde{\boldsymbol{q}}-u\right]_+^2\right]+ \\
&\qquad (1-\alpha)\cdot\max_{\boldsymbol{s}\in\mathcal{S}}\sup_{\substack{\mathbb{P}_w\in\mathcal{P}^w:\\ w\in\mathcal{W}}}\sum_{w\in\mathcal{W}} s_w\cdot\mathbb{E}_{\mathbb{P}_w}\left[\left[u-\boldsymbol{x}^\top\tilde{\boldsymbol{q}}\right]_+^2\right]\Bigg\} \\
&= \underset{u\in\mathbb{R}}{\arg\min}\Bigg\{\alpha\cdot\max_{\boldsymbol{s}\in\mathcal{S}}\sum_{w\in\mathcal{W}} s_w\cdot\underbrace{\sup_{\mathbb{P}_w\in\mathcal{P}^w}\mathbb{E}_{\mathbb{P}_w}\left[\left[\boldsymbol{x}^\top\tilde{\boldsymbol{q}}-u\right]_+^2\right]}_{=\pi_{1w}(\boldsymbol{x},u)}+ \\
&\qquad (1-\alpha)\cdot\max_{\boldsymbol{s}\in\mathcal{S}}\sum_{w\in\mathcal{W}} s_w\cdot\underbrace{\sup_{\mathbb{P}_w\in\mathcal{P}^w}\mathbb{E}_{\mathbb{P}_w}\left[\left[u-\boldsymbol{x}^\top\tilde{\boldsymbol{q}}\right]_+^2\right]}_{=\pi_{2w}(\boldsymbol{x},u)}\Bigg\},
\end{aligned}
$$

which satisfies the conditions of Proposition 7 if we set $\mathcal{U}=\mathbb{R}$, $f_0(u)=0$, $L=2$, $f_1(x,u)=\alpha\cdot x$ and $f_2(x,u)=(1-\alpha)\cdot x$.

In view of the requirements violation index, we observe that

$$
\begin{aligned}
\varphi(\boldsymbol{x}) &= \inf\left\{u>0\;:\;\max_{\boldsymbol{s}\in\mathcal{S}}\sup_{\substack{\mathbb{P}_w\in\mathcal{P}^w:\\ w\in\mathcal{W}}} u\log\left(\sum_{w\in\mathcal{W}} s_w\cdot\mathbb{E}_{\mathbb{P}_w}\left[\exp\left(\frac{\boldsymbol{x}^\top\tilde{\boldsymbol{q}}}{u}\right)\right]\right)\leqslant\overline{\rho}\right\} \\
&= \inf\left\{u>0\;:\;-\overline{\rho}+\max_{\boldsymbol{s}\in\mathcal{S}} u\log\left(\sum_{w\in\mathcal{W}} s_w\cdot\underbrace{\sup_{\mathbb{P}_w\in\mathcal{P}^w}\mathbb{E}_{\mathbb{P}_w}\left[\exp\left(\frac{\boldsymbol{x}^\top\tilde{\boldsymbol{q}}}{u}\right)\right]}_{\pi_{1w}(\boldsymbol{x},u)}\right)\leqslant 0\right\},
\end{aligned}
$$

and the convexity of the log-sum-exp function, together with the fact that convexity is preserved under affine compositions as well as perspectives, shows that the last expression satisfies the conditions of Proposition 7 if we set $\mathcal{U}=\mathbb{R}_{++}$, $g(u)=u$, $f_0(u)=-\overline{\rho}$, $L=1$ and $f_1(x,u)=u\cdot\log(x)$.

As for the CVaR, we note that

$$
\begin{aligned}
\varphi(\boldsymbol{x}) &= \max_{\boldsymbol{s}\in\mathcal{S}}\sup_{\substack{\mathbb{P}_w\in\mathcal{P}^w:\\ w\in\mathcal{W}}}\inf_{u\mathbb{R}}\left\{u+\frac{1}{1-\epsilon}\sum_{w\in\mathcal{W}} s_w\cdot\mathbb{E}_{\mathbb{P}_w}\left[\boldsymbol{x}^\top\tilde{\boldsymbol{q}}-u\right]_+\right\} \\
&= \inf_{u\mathbb{R}}\max_{\boldsymbol{s}\in\mathcal{S}}\sup_{\substack{\mathbb{P}_w\in\mathcal{P}^w:\\ w\in\mathcal{W}}}\left\{u+\frac{1}{1-\epsilon}\sum_{w\in\mathcal{W}} s_w\cdot\mathbb{E}_{\mathbb{P}_w}\left[\boldsymbol{x}^\top\tilde{\boldsymbol{q}}-u\right]_+\right\} \\
&= \inf_{u\mathbb{R}}\left\{u+\max_{\boldsymbol{s}\in\mathcal{S}}\frac{1}{1-\epsilon}\sum_{w\in\mathcal{W}} s_w\cdot\underbrace{\sup_{\mathbb{P}_w\in\mathcal{P}^w}\mathbb{E}_{\mathbb{P}_w}\left[\boldsymbol{x}^\top\tilde{\boldsymbol{q}}-u\right]_+}_{=\pi_{1w}(\boldsymbol{x},u)}\right\},
\end{aligned}
$$

which satisfies the conditions of Proposition 7 if we set $\mathcal{U}=\mathbb{R}$, $f_0(u)=u$, $L=1$ and $f_1(x,u)=x/(1-\epsilon)$.

In view of the service fulfilment index, similar arguments show that

$$
\begin{aligned}
\varphi(\boldsymbol{x}) &= \inf\left\{u \geqslant 0 \,:\, \varphi_{\text{CVaR}}\left(\max\left\{\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho},\, -u\right\}\right) \leqslant 0\right\} \\
&= \inf\left\{u \geqslant 0 \,:\, -(1-\gamma)u + \max_{\boldsymbol{s}\in\mathcal{S}} \sup_{\substack{\mathbb{P}_w \in \mathcal{P}^w: \\ w\in\mathcal{W}}} \sum_{w\in\mathcal{W}} s_w \cdot \mathbb{E}_{\mathbb{P}_w}\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho} + u\right]_+ \leqslant 0\right\} \\
&= \inf\left\{u \geqslant 0 \,:\, -(1-\gamma)u + \max_{\boldsymbol{s}\in\mathcal{S}} \sum_{w\in\mathcal{W}} s_w \cdot \underbrace{\sup_{\mathbb{P}_w\in\mathcal{P}^w} \mathbb{E}_{\mathbb{P}_w}\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho} + u\right]_+}_{=\pi_{1w}(\boldsymbol{x},u)} \leqslant 0\right\},
\end{aligned}
$$

which satisfies the conditions of Proposition 7 if we set $\mathcal{U} = \mathbb{R}_+$, $g(u) = u$, $f_0(u) = -(1-\gamma)\cdot u$, $L = 1$ and $f_1(x,u) = x$.

For the underperformance risk index, finally, we note that

$$
\begin{aligned}
\varphi(\boldsymbol{x}) &= \inf_{u\in\mathbb{R}_{++}} \left\{\frac{1}{u} \,:\, \max_{\boldsymbol{s}\in\mathcal{S}} \sup_{\substack{\mathbb{P}_w\in\mathcal{P}^w: \\ w\in\mathcal{W}}} \sum_{w\in\mathcal{W}} s_w \cdot \mathbb{E}_{\mathbb{P}_w}\left[f\left(u\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho}\right]\right)\right] \leqslant 0\right\} \\
&= \inf_{u\in\mathbb{R}_{++}} \left\{\frac{1}{u} \,:\, \max_{\boldsymbol{s}\in\mathcal{S}} \sum_{w\in\mathcal{W}} s_w \cdot \underbrace{\sup_{\mathbb{P}_w\in\mathcal{P}^w} \mathbb{E}_{\mathbb{P}_w}\left[f\left(u\left[\boldsymbol{x}^\top \tilde{\boldsymbol{q}} - \overline{\rho}\right]\right)\right]}_{=\pi_{1w}(\boldsymbol{x},u)} \leqslant 0\right\}
\end{aligned}
$$

which satisfies the conditions of Proposition 7 if we set $\mathcal{U} = \mathbb{R}_{++}$, $g(u) = 1/u$, $f_0(u) = 0$, $L = 1$ and $f_1(x,u) = x$. □

Proof of Observation 1. The proof is immediate and left out for the sake of brevity. □

Proof of Observation 2. The constituent risk measures $\mathbb{P}$-VaR, $\mathbb{P} \in \mathcal{P}$, are monotone and positive homogeneous (Föllmer and Schied, 2010, p. 3), and Lemma 3 implies that both properties carry over to φ_{VaR}. To see that φ_{VaR} is neither convex nor subadditive, note that ambiguity sets of the form (6) with $\underline{\boldsymbol{q}}^w = \overline{\boldsymbol{q}}^w$, $w \in \mathcal{W}$, and expectation as well as mean absolute deviation conditions absent constitute singleton sets that contain single distributions, and that the value-at-risk is known to violate convexity and subadditivity in that case (Föllmer and Schied, 2010, p. 3). □

Proof of Proposition 8. We show that $\boldsymbol{R} \in \mathcal{C}_{\mathrm{CC}}$ if and only if $\boldsymbol{R} \in \mathcal{C}_{\mathrm{mVaR}}$. Indeed,

$$\begin{aligned}
\boldsymbol{R} \in \mathcal{C}_{\mathrm{CC}} \quad &\iff \quad \mathbb{P}\left[\sum_{i \in \boldsymbol{R}} \tilde{q}_i \leqslant B\right] \geqslant 1-\epsilon && \forall \mathbb{P} \in \mathcal{P} \\
&\iff \quad \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}(\mathbf{1}_{\boldsymbol{R}}{}^\top \tilde{\boldsymbol{q}}) \leqslant B \\
&\iff \quad \left\lceil \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon}(\mathbf{1}_{\boldsymbol{R}}{}^\top \tilde{\boldsymbol{q}})/B \right\rceil \leqslant 1 \\
&\iff \quad \left\lceil \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\mathbf{1}_{\boldsymbol{R}}{}^\top \tilde{\boldsymbol{q}})/B \right\rceil \leqslant 1 && \text{for } k = 2 \\
&\iff \quad \left\lceil \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\mathbf{1}_{\boldsymbol{R}}{}^\top \tilde{\boldsymbol{q}})/B \right\rceil \leqslant 1 && \forall k \in K : k \geqslant 2 \\
&\iff \quad \min\left\{k, \left\lceil \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\mathbf{1}_{\boldsymbol{R}}{}^\top \tilde{\boldsymbol{q}})/B \right\rceil\right\} \leqslant 1 && \forall k \in K : k \geqslant 2 \\
&\iff \quad \max_{k=2,\ldots,m} a(\mathbf{1}_{\boldsymbol{R}}, k) \leqslant 1 \iff \max_{k \in K} a(\mathbf{1}_{\boldsymbol{R}}, k) = 1 \\
&\iff \quad \varphi_{\mathrm{mVaR}}(\mathbf{1}_{\boldsymbol{R}}) = B \iff \boldsymbol{R} \in \mathcal{C}_{\mathrm{mVaR}},
\end{aligned}$$

where the first two equivalences follow from the definition of the set $\mathcal{C}_{\mathrm{CC}}$ and Observation 1, respectively, while the fifth equivalence holds since the worst-case value-at-risk is monotonically non-decreasing in its risk threshold. The eighth equivalence holds since $a(\mathbf{1}_{\boldsymbol{R}}, 1) = 1$ by definition, the penultimate equivalence follows from the definition of φ_{mVaR}, and the last equivalence holds since any $\boldsymbol{R} \in \mathcal{C}_{\mathrm{mVaR}}$ must satisfy $\varphi_{\mathrm{mVaR}}(\mathbf{1}_{\boldsymbol{R}}) = B$ as $a(\mathbf{1}_{\boldsymbol{R}}, 1) = 1$. □

The proof of Proposition 9 relies on five auxiliary results, which we state and prove first.

Lemma 16. *Let* $k^\star = \min \arg\max\{a(\boldsymbol{x}, k) : k \in K\}$. *Then* $a(\boldsymbol{x}, k^\star) = k^\star$.

Proof. If $k^\star = 1$, then we have $a(\boldsymbol{x}, k^\star) = 1$ by definition of a. In the remainder of the proof, we thus assume that $k^\star \geqslant 2$. Define $Z_k = \lceil \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}\left(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}\right)/B \rceil$ so that $a(\boldsymbol{x}, k) = \min\{k, Z_k\}$. The statement of the lemma follows if we show that $k^\star \leqslant Z_{k^\star}$.

Assume to the contrary that $Z_{k^\star} < k^\star$, which implies that $a(\boldsymbol{x}, k^\star) = Z_{k^\star}$. Then $Z_{k^\star} \leqslant k^\star - 1$ because $Z_{k^\star} \in \mathbb{Z}$, as well as $Z_{k^\star} \leqslant Z_{k^\star-1}$ since the worst-case value-at-risk is monotonically non-decreasing in its risk threshold. We consider two possible cases, both of which will lead to a contradiction: If $k^\star - 1 \leqslant Z_{k^\star-1}$, then $a(\boldsymbol{x}, k^\star - 1) = k^\star - 1 \geqslant Z_{k^\star} = a(\boldsymbol{x}, k^\star)$. If $k^\star - 1 > Z_{k^\star-1}$, on the other hand, then $a(\boldsymbol{x}, k^\star - 1) = Z_{k^\star-1} \geqslant Z_{k^\star} = a(\boldsymbol{x}, k^\star)$. Either case, however, violates the assumption that $k^\star$ is the smallest maximizer of $a(\boldsymbol{x}, \cdot)$. □

Lemma 17. *Let* $k^\star = \min \arg\max\{a(\boldsymbol{x}, k) \,:\, k \in K\}$. *Then* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-k^\star\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \leqslant k^\star B$.

Proof. Note that

$$a(\boldsymbol{x}, k^\star + 1) \;=\; \min\left\{k^\star + 1,\, \left\lceil \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-k^\star\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})/B \right\rceil\right\} \;\leqslant\; a(\boldsymbol{x}, k^\star) \;=\; k^\star,$$

where the first identity holds by definition, the inequality holds since $k^\star$ maximizes $a(\boldsymbol{x}, \cdot)$, and the second identity follows from Lemma 16. Since $k^\star + 1 > k^\star$, the above equation implies that

$$\left\lceil \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-k^\star\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})/B \right\rceil \;\leqslant\; k^\star,$$

which immediately implies the statement of the lemma. □

Lemma 18. *For any two random variables* $\tilde{X}_1$ *and* $\tilde{X}_2$ *and risk thresholds* $\epsilon_1, \epsilon_2 \in (0,1)$, *we have* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon_1-\epsilon_2}(\tilde{X}_1 + \tilde{X}_2) \leqslant \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon_1}(\tilde{X}_1) + \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon_2}(\tilde{X}_2)$.

Proof. Define $\vartheta_i = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon_i}(\tilde{X}_i)$ for $i = 1, 2$. We need to show that $\mathbb{P}(\tilde{X}_1 + \tilde{X}_2 \leqslant \vartheta_1 + \vartheta_2) \geqslant 1 - \epsilon_1 - \epsilon_2$ for all $\mathbb{P} \in \mathcal{P}$. To this end, fix any $\mathbb{P} \in \mathcal{P}$ and observe that

$$\begin{aligned} \mathbb{P}(\tilde{X}_1 + \tilde{X}_2 \leqslant \vartheta_1 + \vartheta_2) \;&\geqslant\; \mathbb{P}(\tilde{X}_1 \leqslant \vartheta_1 \text{ and } \tilde{X}_2 \leqslant \vartheta_2) &&=\; 1 - \mathbb{P}(\tilde{X}_1 > \vartheta_1 \text{ or } \tilde{X}_2 > \vartheta_2) \\ &\geqslant\; 1 - [\mathbb{P}(\tilde{X}_1 > \vartheta_1) + \mathbb{P}(\tilde{X}_2 > \vartheta_2)] &&=\; 1 - \epsilon_1 - \epsilon_2, \end{aligned}$$

where the second inequality is due to Bonferroni's inequality. Since $\mathbb{P} \in \mathcal{P}$ was selected arbitrarily, the statement of the lemma follows. □

Lemma 19. *The worst-case risk measure* φ_{mVaR} *is monotone.*

Proof. Fix any $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ with $\boldsymbol{x} \leqslant \boldsymbol{y}$ and define $Z_k = \lceil \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})/B \rceil$ as well as $Z'_k = \lceil \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\boldsymbol{y}^\top \tilde{\boldsymbol{q}})/B \rceil$. Since $\mathbb{P}$-VaR is monotone (Föllmer and Schied, 2010, p. 3) and $\tilde{\boldsymbol{q}} \geqslant \boldsymbol{0}$ $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}$, Lemma 3 implies that

$$\begin{aligned} Z_k \;\leqslant\; Z'_k \quad \forall k \in K \;\;&\Longleftrightarrow\;\; \min\{k,\, Z_k\} \;\leqslant\; \min\{k,\, Z'_k\} && \forall k \in K \\ &\Longleftrightarrow\;\; \min\{k,\, Z_k\} \;\leqslant\; \max_{k\in K}\, \min\{k,\, Z'_k\} && \forall k \in K \\ &\Longleftrightarrow\;\; \max_{k\in K}\, \min\{k,\, Z_k\} \;\leqslant\; \max_{k\in K}\, \min\{k,\, Z'_k\}, \end{aligned}$$

and the last inequality immediately implies that $\varphi_{\text{mVaR}}(\boldsymbol{x}) \leqslant \varphi_{\text{mVaR}}(\boldsymbol{y})$. □

Lemma 20. *The worst-case risk measure* φ_{mVaR} *is subadditive.*

Proof. Assume to the contrary that φ_{mVaR} is not subadditive, that is, there are $\boldsymbol{x}_1, \boldsymbol{x}_2 \in [0,1]^n$, $\boldsymbol{x}_1 + \boldsymbol{x}_2 \leqslant \mathbf{e}$, with corresponding maximizers $k_1^\star$, $k_2^\star$ of function a (*cf.* Lemma 16) such that

$$\begin{aligned}
& \varphi_{\text{mVaR}}(\boldsymbol{x}_1 + \boldsymbol{x}_2) \;>\; \varphi_{\text{mVaR}}(\boldsymbol{x}_1) + \varphi_{\text{mVaR}}(\boldsymbol{x}_2) \\
\iff \quad & B \cdot \max_{k \in K} a(\boldsymbol{x}_1 + \boldsymbol{x}_2, k) \;>\; B \cdot (k_1^\star + k_2^\star) \\
\iff \quad & \max_{k \in K} a(\boldsymbol{x}_1 + \boldsymbol{x}_2, k) \;>\; k_1^\star + k_2^\star \\
\iff \quad & a(\boldsymbol{x}_1 + \boldsymbol{x}_2, k') \;>\; k_1^\star + k_2^\star \qquad \text{for some } k' \in \{k_1^\star + k_2^\star + 1, \ldots, m\} \\
\iff \quad & \min\left\{k', \left\lceil \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k'-1)\epsilon}\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)^\top \tilde{\boldsymbol{q}}\right)/B \right\rceil\right\} \;>\; k_1^\star + k_2^\star \\
\iff \quad & \left\lceil \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k'-1)\epsilon}\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)^\top \tilde{\boldsymbol{q}}\right)/B \right\rceil \;>\; k_1^\star + k_2^\star \\
\iff \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k'-1)\epsilon}\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)^\top \tilde{\boldsymbol{q}}\right) \;>\; B \cdot (k_1^\star + k_2^\star),
\end{aligned} \tag{13}$$

where the third equivalence follows from the fact that $a(\boldsymbol{x}_1 + \boldsymbol{x}_2, k') \leqslant k'$ for all $k' \in K$.

Define $\epsilon_1 = k_1^\star \epsilon$ and $\epsilon_2 = k_2^\star \epsilon$. Our assumption $\epsilon < 1/m$ implies that $\epsilon_1, \epsilon_2 \in (0,1)$. We have

$$\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k'-1)\epsilon}\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)^\top \tilde{\boldsymbol{q}}\right) &\leqslant \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k_1^\star + k_2^\star)\epsilon}\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)^\top \tilde{\boldsymbol{q}}\right) \\
&= \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon_1-\epsilon_2}\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)^\top \tilde{\boldsymbol{q}}\right) \\
&\leqslant \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon_1}\left(\boldsymbol{x}_1^\top \tilde{\boldsymbol{q}}\right) + \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-\epsilon_2}\left(\boldsymbol{x}_2^\top \tilde{\boldsymbol{q}}\right) \\
&\leqslant B \cdot (k_1^\star + k_2^\star),
\end{aligned}$$

where the first inequality holds since $k' \geqslant k_1^\star + k_2^\star + 1$, while the other two inequalities are due to Lemmas 18 and 17, respectively. Since this inequality chain violates (13), the statement follows. □

Proof of Proposition 9. The monotonicity and subadditivity of φ_{mVaR} follow from Lemmas 19 and 20, respectively. Moreover, φ_{mVaR} cannot be positive homogeneous since its image is restricted to integer numbers. To see that φ_{mVaR} is not convex either, consider an ambiguity set $\mathcal{P}$ of the form (6) with $W = 1$ and $n = 3$, $\underline{\boldsymbol{q}}^1 = (1,5,1)^\top$ and $\overline{\boldsymbol{q}}^1 = (30,20,30)^\top$, $\boldsymbol{\mu}^1 = (16,10,16)^\top$ as well as $\boldsymbol{\nu}^1 = (2,0.5,2)^\top$. For $\epsilon = 0.1$, $B = 20.6$ as well as $\boldsymbol{x} = (1,0,1)^\top$ and $\boldsymbol{y} = (0,1,0)^\top$, we have

$$\underbrace{\varphi_{\text{mVaR}}(0.67 \cdot \boldsymbol{x} + 0.33 \cdot \boldsymbol{y})}_{=2\cdot B} \;>\; 0.67 \cdot \underbrace{\varphi_{\text{mVaR}}(\boldsymbol{x})}_{=2\cdot B} + 0.33 \cdot \underbrace{\varphi_{\text{mVaR}}(\boldsymbol{y})}_{=B},$$

which shows that φ_{mVaR} is indeed not convex. □

Proof of Theorem 8. Define $Z_k = \lceil \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})/B \rceil$ such that $a(\boldsymbol{x}, 1) = 1$ and $a(\boldsymbol{x}, k) = \min\{k, \lceil Z_k \rceil\}$ for $k \geqslant 2$. Let $k^\star = \min \arg\max\{a(\boldsymbol{x}, k) \, : \, k \in K\}$. We claim that for all $k \in K$, *(i)* if $k > Z_k$, then $k > k^\star$; and *(ii)* if $k \leqslant Z_k$, then $k \leqslant k^\star$. This will imply that $k^\star$ can be determined via binary search as long as we can compute Z_k, $k \in K$, efficiently.

In view of *(i)*, assume to the contrary that there is $k' > Z_{k'}$ such that $1 \leqslant k' \leqslant k^\star$. We then have $Z_{k'} < k' \leqslant k^\star \leqslant Z_{k^\star}$, where the last inequality follows from Lemma 16. This, however, contradicts the fact that $Z_{k^\star} \leqslant Z_{k'}$ since $k' \leqslant k^\star$ by assumption and Z_k is monotonically non-increasing in k. As for *(ii)*, assume to the contrary that there is $k' \leqslant Z_{k'}$ such that $k' > k^\star$. We then have $k^\star < k' \leqslant Z_{k'} \leqslant Z_{k^\star}$, where the last inequality again holds since Z_k is monotonically non-increasing in k. This implies that $a(\boldsymbol{x}, k') = k' > a(\boldsymbol{x}, k^\star)$, which contradicts the fact that $k^\star \in \arg\max\, a(\boldsymbol{x}, k)$.

It remains to be shown how Z_k can be evaluated efficiently. To this end, we note that

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) &= \inf_{u\in\mathbb{R}} \left\{ u \, : \, \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}} \leqslant u) \geqslant 1-(k-1)\epsilon \right\} \\
&= \inf_{u\in\mathbb{R}} \left\{ u \, : \, \inf_{\substack{\mathbb{P}_w\in\mathcal{P}^w: \\ w\in\mathcal{W}}} \sum_{w\in\mathcal{W}} \hat{s}_w \cdot \mathbb{P}_w(\boldsymbol{x}^\top \tilde{\boldsymbol{q}} \leqslant u) \geqslant 1-(k-1)\epsilon \right\} \\
&= \inf_{u\in\mathbb{R}} \left\{ u \, : \, \sum_{w\in\mathcal{W}} \hat{s}_w \cdot \inf_{\mathbb{P}_w\in\mathcal{P}^w} \mathbb{P}_w(\boldsymbol{x}^\top \tilde{\boldsymbol{q}} \leqslant u) \geqslant 1-(k-1)\epsilon \right\},
\end{aligned}$$

where the first equality holds by definition and the other two identities are due to the law of total probability as well as the rectangularity of the ambiguity set, respectively. Verifying whether a fixed u under- or overestimates the worst-case value at risk thus reduces to computing the quantity

$$\begin{aligned}
\inf_{\mathbb{P}_w\in\mathcal{P}^w} \mathbb{P}_w(\boldsymbol{x}^\top \tilde{\boldsymbol{q}} \leqslant u) &= \sup_{\theta_w\in[0,1]} \left\{ 1-\theta_w \, : \, \sup_{\mathbb{P}_w\in\mathcal{P}^w} \mathbb{P}_w\text{-VaR}_{1-\theta_w}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \leqslant u \right\} \\
&= \sup_{\theta_w\in[0,1]} \left\{ 1-\theta_w \, : \, \sum_{i\in V_C} x_i \left(\mu_i^w + \min\left\{ \overline{q}_i^w - \mu_i^w, \, \left(\frac{1-\theta_w}{\theta_w}\right)(\mu_i^w - \underline{q}_i^w), \, \frac{\nu_i^w}{2\theta_w} \right\} \right) \leqslant u \right\},
\end{aligned}$$

for all $w \in \mathcal{W}$ and verifying whether their $\hat{\boldsymbol{s}}$-weighted sum weakly exceeds $1-(k-1)\epsilon$. Here, the first identity follows from the definition of the value-at-risk, whereas the second identity is due to Proposition 2 of Ghosal and Wiesemann (2020). Note that the sum embedded in the final maximization problem above is monotonically non-increasing and piecewise smooth in θ_w with at most $3n$ breakpoints $(\mu_i^w - \underline{q}_i^w)/(\overline{q}_i^w - \underline{q}_i^w)$, $\nu_i^w/(2[\overline{q}_i^w - \mu_i^w])$ and $(\mu_i^w - \underline{q}_i^w - \nu_i^w/2)/(\mu_i^w - \underline{q}_i^w)$, $i \in V_C$.

We can sort these breakpoints in time $\mathcal{O}(n \log n)$ and compute the maximizer via a binary search. The binary search takes $\mathcal{O}(\log n)$ iterations of time $\mathcal{O}(n)$. We can embed the binary searches over θ_w, $w \in \mathcal{W}$, in a binary search over $u \in \mathbb{R}$ to compute Z_k. The outer binary search can be initialized with the lower bound $\underline{u} = 0$ and the upper bound $\overline{u} = \max_{w \in \mathcal{W}} \mathbf{e}^\top \overline{\boldsymbol{q}}^w$, and it can be terminated once the bounds differ by less than the accuracy κ. Finally, we need to conduct a binary search over the number of vehicles $k \in K$ to compute the maximizer of $a(\boldsymbol{x}, \cdot)$. □

Proof of Proposition 10. A similar reasoning as in the proof of Theorem 8 applies; the main difference lies in the computation of $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}})$. We now have

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-VaR}_{1-(k-1)\epsilon}(\boldsymbol{x}^\top \tilde{\boldsymbol{q}}) \;=\; \inf_{u \in \mathbb{R}} \left\{ u \,:\, \min_{\boldsymbol{s} \in \mathcal{S}} \; \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}, u) \geqslant 1 - (k-1)\epsilon \right\}, \tag{14}$$

where each component $\pi_w(\boldsymbol{x}, u) = \inf_{\mathbb{P}_w \in \mathcal{P}^w} \mathbb{P}_w(\boldsymbol{x}^\top \tilde{\boldsymbol{q}} \leqslant u)$, $w \in \mathcal{W}$, can be computed in time $\mathcal{O}(n \log n)$ as detailed in the proof of Theorem 8, and Proposition 6 provides the complexity estimates for the computation of $\min_{\boldsymbol{s} \in \mathcal{S}} \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$.

A mathematical subtlety arises for the entropy and (axis-parallel as well as generic) ellipsoidal ambiguity sets, where the quantities $\min_{\boldsymbol{s} \in \mathcal{S}} \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$ are only computed to a limited accuracy $\delta > 0$. We need to choose δ small enough so that for any u outside the κ-neighborhood of a minimizer $u^\star$ in (14), the δ-neighborhood of $\min_{\boldsymbol{s} \in \mathcal{S}} \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}, u)$ is either fully contained in the interval $[0,\, 1-(k-1)\epsilon)$ or fully contained in the interval $[1-(k-1)\epsilon,\, 1]$, as this guarantees that the bisection decisions are not influenced by the inaccurate computation of $\min_{\boldsymbol{s} \in \mathcal{S}} \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x})$. Since $\boldsymbol{\pi}$ is monotonically non-decreasing in u, any u outside the κ-neighborhood of a minimizer $u^\star$ satisfies

$$\left| \min_{\boldsymbol{s} \in \mathcal{S}} \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}, u) \;-\; \min_{\boldsymbol{s} \in \mathcal{S}} \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}, u^\star) \right| \;\geqslant\; \kappa \cdot \min_{u \in \mathbb{R}} \left| \frac{\mathrm{d}}{\mathrm{d}u} \min_{\boldsymbol{s} \in \mathcal{S}} \; \boldsymbol{s}^\top \boldsymbol{\pi}(\boldsymbol{x}, u) \right| \;\geqslant\; \kappa \cdot \min_{u \in \mathbb{R},\, w \in \mathcal{W}} \left| \frac{\mathrm{d}}{\mathrm{d}u} \pi_w(\boldsymbol{x}, u) \right|.$$

Straightforward but tedious calculations show that

$$\frac{\mathrm{d}}{\mathrm{d}u} \pi_w(\boldsymbol{x}, u) \;\geqslant\; \underline{\pi}_w \;:=\; \frac{\min\{\mu_j^w - \underline{q}_j^w \,:\, j \in V_C\}}{\max\{\mathbf{e}^\top \overline{\boldsymbol{q}}^w \,:\, w \in \mathcal{W}\}^2}$$

uniformly across $u \in \mathbb{R}$, and it is thus sufficient to select $\delta \leqslant \kappa \cdot \min\{\underline{\pi}_w \,:\, w \in \mathcal{W}\}$, which implies the stated complexity estimates. □

Appendix B: Worst-Case Distribution for Theorem 4

Algorithm 1 computes the worst-case probabilities $(p^\star_{wj})_{w,j}$ with associated demand realizations $(\boldsymbol{q}^\star_{wj})_{w,j}$ for Theorem 4. The intuition behind this algorithm is outlined in Section 5.1, and the correctness of the algorithm is proven by Long et al. (2020).

Algorithm 1: Algorithm for determining the worst-case distribution (Long et al., 2020)

Input: $(\underline{\boldsymbol{q}}^w, \overline{\boldsymbol{q}}^w)$, $\boldsymbol{\mu}^w$ and $\boldsymbol{\nu}^w$, $w \in \mathcal{W}$, for the ambiguity set (6).

for $w \in \mathcal{W}$ **do**

Compute for all customers $i \in V_C$ the marginal worst-case distribution:

- $\mathbb{P}^w\big[\tilde{q}^w_i = \underline{q}^w_i\big] = \frac{\hat{\nu}^w_i}{2(\mu^w_i - \underline{q}^w_i)}$
- $\mathbb{P}^w\big[\tilde{q}^w_i = \mu^w_i\big] = 1 - \frac{\hat{\nu}^w_i(\overline{q}^w_i - \underline{q}^w_i)}{2(\overline{q}^w_i - \mu^w_i)(\mu^w_i - \underline{q}^w_i)}$
- $\mathbb{P}^w\big[\tilde{q}^w_i = \overline{q}^w_i\big] = \frac{\hat{\nu}^w_i}{2(\overline{q}^w_i - \mu^w_i)}$, where $\hat{\nu}^w_i = \min\left\{\nu^w_i, \frac{2(\overline{q}^w_i - \mu^w_i)(\mu^w_i - \underline{q}^w_i)}{\overline{q}^w_i - \underline{q}^w_i}\right\}$

Set $\boldsymbol{q}^\star_{w,1} = \underline{\boldsymbol{q}}^w$ and $\boldsymbol{m} = (m_i)_{i\in V_C} = (\mathbb{P}^w[\tilde{q}^w_i = \underline{q}^w_i])_{i\in V_C}$

for $j = 1, 2, \ldots, 2n$ **do**

Let $k = \min \arg\min \{m_i \,:\, i \in V_C\}$

Set $p^\star_{wj} = m_k$, $\boldsymbol{q}^\star_{w,j+1} = \boldsymbol{q}^\star_{w,j}$ and $\boldsymbol{m} = \boldsymbol{m} - p^\star_{wj}\mathbf{e}$

If $q^\star_{w,j+1,k} = \underline{q}^w_k$ **then** set $q^\star_{w,j+1,k} = \mu^w_k$ **else** set $q^\star_{w,j+1,k} = \overline{q}^w_k$

Set $m_k = \mathbb{P}^w[\tilde{q}^w_k = q^\star_{w,j+1,k}]$

end

end

Output: Worst-case probabilities $(p^\star_{wj})_{w,j}$ with associated demand realizations $(\boldsymbol{q}^\star_{wj})_{w,j}$.

Appendix C: Detailed Numerical Results

Table 1 reports the best feasible solution ('Opt'; accompanied by an asterisk if it is confirmed to be optimal) and the best lower bound ('LB'; value in brackets unless solved to optimality) identified by, as well as the runtime ('t'; unless not solved to optimality, in which case the runtime is 12h) incurred by our branch-and-cut scheme for the deterministic CVRP ('Deterministic'), the distributionally robust CVRP with known ('Stochastic') and unknown scenario probabilities ('Ambiguous').

	Deterministic		**Stochastic**		**Ambiguous**	
Problem	Opt	t (sec) [LB]	Opt	t (sec) [LB]	Opt	t (sec) [LB]
A-n32-k5	745.0*	0.1	745.0*	0.32	747.0*	0.26
A-n33-k5	617.0*	0.09	639.0*	5.14	639.0*	4.85
A-n33-k6	703.0*	0.4	707.0*	31.94	711.0*	27.86
A-n34-k5	701.0*	0.1	701.0*	1.26	702.0*	1.34
A-n36-k5	732.0*	0.15	743.0*	13.35	758.0*	231.02
A-n37-k5	651.0*	0.16	653.0*	3.6	655.0*	4.22
A-n37-k6	861.0*	1.8	877.0*	232.77	879.0*	96.01
A-n38-k5	648.0*	0.07	654.0*	1.08	654.0*	0.98
A-n39-k5	735.0*	0.42	758.0*	47.12	762.0*	67.48
A-n39-k6	774.0*	0.6	774.0*	11.65	774.0*	11.1
A-n44-k6	891.0*	29.07	892.0*	692.84	897.0*	970.76
A-n45-k6	869.0*	1.93	872.0*	87.95	873.0*	26.96
A-n45-k7	1034.0*	5.42	1051.0*	514.83	1064.0*	2226.95
A-n46-k7	851.0*	0.52	871.0*	43.66	874.0*	59.06
A-n48-k7	967.0*	0.91	967.0*	10.91	979.0*	57.31
A-n53-k7	954.0*	2.22	959.0*	159.08	968.0*	423.89
A-n54-k7	1051.0*	65.48	1068.0*	3821.02	1080.0*	19054.7
A-n55-k9	985.0*	1.2	992.0*	21.01	1013.0*	195.39
A-n60-k9	1202.0*	24.35	1214.0*	4693.07	1228.0*	29396.3
A-n61-k9	939.0*	5.48	942.0*	257.96	948.0*	589.91
A-n62-k8	1132.0*	9.2	1153.0*	1283.46	1161.0*	1396.45
A-n63-k9	1446.0*	1975.6	1476.0	[1444.05]	1493.0	[1444.22]
A-n63-k10	1176.0*	34.68	1178.0*	234.11	1219.0	[1206.15]
A-n64-k9	1290.0	[1277.83]	1333.0	[1267.2]	1349.0	[1265.06]
A-n65-k9	1082.0*	56.79	1085.0*	1150.0	1089.0*	2614.33
A-n69-k9	1076.0*	190.54	1082.0*	10061.2	1091.0	[1083.65]
A-n80-k10	1612.0	[1587.4]	1641.0	[1587.1]	no-feas	[1587.42]

Table 1. Runtimes and optimality gaps for the benchmark instances of Díaz (2006). Optimally solved instances are highlighted with an asterisk and accompanied by the runtime t. For all other instances, we report the upper and lower bound after 12 hours.

Problem	Deterministic		Stochastic		Ambiguous	
	Opt	t (sec) [LB]	Opt	t (sec) [LB]	Opt	t (sec) [LB]
B-n31-k5	645.0*	0.08	651.0*	0.82	651.0*	0.84
B-n34-k5	703.0*	0.18	737.0*	0.62	740.0*	1.34
B-n35-k5	866.0*	0.04	866.0*	0.11	866.0*	0.03
B-n38-k6	726.0*	0.09	730.0*	18.46	731.0*	23.29
B-n39-k5	517.0*	0.14	521.0*	0.17	521.0*	0.64
B-n41-k6	786.0*	0.08	786.0*	4.26	789.0*	1.78
B-n43-k6	655.0*	0.87	662.0*	717.52	678.0*	9.81
B-n44-k7	819.0*	4.5	835.0*	1598.4	841.0*	81.71
B-n45-k5	630.0*	0.11	666.0*	3.61	669.0*	2.81
B-n45-k6	616.0*	0.42	626.0*	6.2	626.0*	9.77
B-n50-k7	657.0*	0.13	661.0*	0.83	661.0*	0.27
B-n50-k8	1145.0*	2.62	1202.0	[1158.33]	1212.0	[1179.61]
B-n51-k7	913.0*	0.06	917.0*	0.32	921.0*	1.28
B-n52-k7	673.0*	0.14	673.0*	0.53	674.0*	0.79
B-n56-k7	621.0*	0.43	622.0*	14.44	622.0*	6.52
B-n57-k9	1511.0*	9.16	1535.0*	5829.33	1538.0*	3979.0
B-n63-k10	1347.0*	137.91	1361.0*	8751.78	1364.0*	4281.32
B-n64-k9	790.0*	1.06	796.0*	10.3	797.0*	24.36
B-n66-k9	1170.0*	573.06	1202.0*	32405.8	1206.0*	4203.8
B-n67-k10	946.0*	3.01	974.0*	1499.96	978.0*	3519.29
B-n68-k9	1114.0*	20.66	1117.0*	394.49	1124.0*	6341.28
B-n78-k10	1079.0*	28.7	1101.0	[1089.5]	1105.0*	17351.5
E-n101-k8	780.0*	159.12	787.0	[783.05]	797.0	[780.978]
E-n101-k14	1012.0	[991.132]	1048.0	[984.161]	1057.0	[984.503]
E-n22-k4	370.0*	0.01	370.0*	0.25	370.0*	0.07
E-n23-k3	564.0*	0.0	564.0*	0.0	564.0*	0.0
E-n30-k3	475.0*	0.02	475.0*	0.02	475.0*	0.02
E-n33-k4	791.0*	0.15	791.0*	0.36	791.0*	0.41
E-n51-k5	510.0*	4.77	514.0*	364.7	515.0*	614.99
E-n76-k7	656.0*	97.12	660.0*	10538.1	661.0*	16929.2
E-n76-k8	699.0*	6245.56	703.0	[694.866]	704.0	[694.913]
E-n76-k10	772.0	[769.85]	784.0	[759.652]	790.0	[761.586]
E-n76-k14	939.0	[912.383]	960.0	[900.633]	959.0	[904.464]
F-n135-k7	1069.0*	348.29	1076.0*	4619.5	1081.0*	14412.2
F-n45-k4	706.0*	0.16	710.0*	1.11	711.0*	0.52
F-n72-k4	232.0*	0.28	232.0*	0.43	232.0*	0.59
M-n101-k10	795.0*	4.71	798.0*	108.72	798.0*	255.62
M-n121-k7	962.0	[949.444]	981.0	[949.303]	975.0	[951.297]
M-n151-k12	no-feas	[935.76]	no-feas	[932.814]	no-feas	[932.771]

Table 1. (Continued from previous page.)

Problem	**Deterministic** Opt	t (sec) [LB]	**Stochastic** Opt	t (sec) [LB]	**Ambiguous** Opt	t (sec) [LB]
P-n19-k2	195.0⋆	0.0	195.0⋆	0.0	195.0⋆	0.0
P-n20-k2	208.0⋆	0.0	208.0⋆	0.01	208.0⋆	0.01
P-n21-k2	208.0⋆	0.0	208.0⋆	0.01	208.0⋆	0.0
P-n22-k2	213.0⋆	0.01	213.0⋆	0.02	213.0⋆	0.03
P-n22-k8	549.0⋆	0.01	549.0⋆	0.02	549.0⋆	0.02
P-n23-k8	486.0⋆	0.19	491.0⋆	15.09	491.0⋆	12.08
P-n40-k5	448.0⋆	0.23	449.0⋆	1.52	449.0⋆	1.86
P-n45-k5	496.0⋆	0.52	496.0⋆	15.73	500.0⋆	41.4
P-n50-k7	531.0⋆	3.56	539.0⋆	480.6	540.0⋆	483.5
P-n50-k8	580.0⋆	19.7	584.0⋆	748.81	585.0⋆	860.38
P-n50-k10	649.0⋆	25.32	652.0⋆	1221.82	657.0⋆	2681.08
P-n51-k10	686.0⋆	19.01	688.0⋆	1116.65	688.0⋆	526.73
P-n55-k10	656.0⋆	43.25	658.0⋆	1719.87	659.0⋆	3215.28
P-n55-k7	539.0⋆	0.71	543.0⋆	120.47	548.0⋆	678.55
P-n55-k8	571.0⋆	3.01	572.0⋆	276.58	573.0⋆	273.35
P-n55-k15	868.0⋆	176.34	871.0⋆	6372.02	872.0⋆	4492.75
P-n60-k10	703.0⋆	434.76	704.0⋆	14903.3	709.0⋆	35829.7
P-n60-k15	904.0⋆	275.71	911.0⋆	34400.8	921.0	[908.389]
P-n65-k10	750.0⋆	449.65	757.0	[748.853]	760.0	[748.066]
P-n70-k10	773.0	[769.264]	781.0	[760.688]	776.0	[760.304]
P-n76-k4	588.0⋆	1.61	588.0⋆	21.87	588.0⋆	20.13
P-n76-k5	608.0⋆	8.49	612.0⋆	4477.6	613.0⋆	5235.28
P-n101-k4	673.0⋆	0.99	673.0⋆	25.56	673.0⋆	21.65
att-n48-k4	38634.0⋆	0.59	38637.0⋆	7.58	38637.0⋆	16.07

Table 1. (Continued from previous page.)