Hardness of pricing routes for two-stage stochastic vehicle routing problems with scenarios

Matheus J. Ota[a,*], Ricardo Fukasawa[a]

[a] *University of Waterloo, Waterloo, Ontario, Canada*

Abstract

The vehicle routing problem with stochastic demands (VRPSD) generalizes the classic vehicle routing problem by considering customer demands as random variables. Similarly to other vehicle routing variants, state-of-the-art algorithms for the VRPSD are often based on set-partitioning formulations, which require efficient routines for the associated pricing problems. However, all these set-partitioning-based approaches have strong assumptions on the correlation between the demands random variables (e.g. no correlation), a simplification that diverges from real-world settings where correlations frequently exist. In contrast, there is a significant effort in the stochastic programming community to solve problems where the uncertainty is modeled with a finite set of scenarios. This approach can approximate more diverse distributions via sampling and is particularly appealing in data-driven contexts, where historical data is readily available. To fill this gap, we focus on the VRPSD with demands given by scenarios. We show that, for any route relaxation (where repeated visits are allowed in a route) and any approximation of the recourse cost that satisfy some mild assumptions, the VRPSD pricing problem is still strongly NP-hard. This provides a very strong argument for the difficulty of developing efficient column-generation based algorithms for the VRPSD with demands following an empirical probability distribution of scenarios.

Keywords: Vehicle routing, Computational complexity, Stochastic programming

1. Introduction

Vehicle routing problems (VRPs) are ubiquitous in operations research, and many different variants have been proposed (Toth & Vigo, 2014). Among these variants, the vehicle routing problem with stochastic demands (VRPSD) addresses the case where, rather than deterministic, the customer demands are stochastic and follow a given probability distribution. As noted by Gendreau et al. (2016), the VRPSD has a rich history of over 50 years (Tillman, 1969), with an increasing interest in the last decade (Dinh et al., 2018; Florio et al., 2020; Ghosal & Wiesemann, 2020; Pessoa et al., 2021; Ledvina et al., 2022; Florio et al., 2022; Hoogendoorn & Spliet, 2023). To handle the possibility of *route failures* — that is, when the realized accumulated demand of a route exceeds the vehicle capacity — different approaches have been proposed, such as chance-constraint and dynamic reoptimization (Gendreau et al., 2016). However, the most extensively explored approach, which is the focus of our study,

*Corresponding author

Email addresses: mjota@uwaterloo.ca (Matheus J. Ota), rfukasawa@uwaterloo.ca (Ricardo Fukasawa)

models the VRPSD as a *two-stage stochastic program with recourse*. In this paradigm, the *first-stage* determines the vehicle routes; then, in the *second-stage*, the customer demands random variables are revealed upon vehicle arrival, and in the event of a route failure, a recourse policy describes the recourse actions that the vehicle should execute. The optimization model for the VRPSD then seeks to minimize the sum of the lengths of the selected routes and their expected second-stage costs.

Following a trend in exact algorithms for different vehicle routing problems (Poggi & Uchoa, 2014), several state-of-the-art exact algorithms for the VRPSD are based on set partitioning (SP) formulations (Christiansen & Lysgaard, 2007; Gauvin et al., 2014; Florio et al., 2020, 2022). In this context, a *q-route* is said to be a sequence of customers whose sum of expected demands does not exceed the vehicle capacity. SP formulations for the VRPSD typically have a variable for each q-route (or ng-route, see Gauvin et al. (2014)), and thus, they may have exponential size. The linear programming relaxation of SP formulations are then solved with a column-generation method, and consequently, the corresponding branch(-cut)-and-price algorithms rely on computationally efficient routines for solving the pricing problems. In fact, the exact algorithms of Christiansen & Lysgaard (2007); Gauvin et al. (2014); Florio et al. (2020, 2022) all make use of efficient dynamic programming / labeling algorithms to solve the pricing problems.

With the exception of Florio et al. (2022), all of the previously mentioned works assume that the customer demands follow independent probability distributions. Furthermore, Florio et al. (2022) allows only a very limited type of correlation, where all random variables are affected by a single *external factor*. Motivated by results on the sample average approximation method and following several works in the stochastic optimization literature (Birge & Louveaux, 2011; Swamy & Shmoys, 2012; Luedtke & Ahmed, 2008; Chen & Luedtke, 2022; Verweij et al., 2003), we instead consider the VRPSD with an empirical probability distribution of scenarios — henceforth simply called *VRPSD with scenarios*. The main goal of this paper is to show that, under some mild assumptions on the second-stage costs, it is strongly $\mathcal{NP}$-hard to solve the pricing problem of the VRPSD with scenarios.

To the best of our knowledge, the only work that explicitly addresses the complexity of solving the VRPSD pricing problem is a paper by Fukasawa & Gunter (2023). Although their work also considers the VRPSD with scenarios, our results extend and generalize theirs in at least two ways. To explain one of the main novelties of our work, we must make a few comments about the paper of Fukasawa & Gunter (2023). First, the authors only consider a particular recourse policy (the classical or detour-to-depot recourse policy). In addition, while they give an indication of strong $\mathcal{NP}$-hardness for pricing q-routes with demands following independent normals, they argue that such a result still allows the existence of a pseudo-polynomial pricing algorithm such as the one of Christiansen & Lysgaard (2007), since the second-stage cost considered by Christiansen & Lysgaard (2007) is exact at the elementary q-routes, but only approximate at the non-elementary ones. (A q-route is *elementary* if each customer appears in it at most once.) One may then attempt to make the pricing problem of the VRPSD with scenarios easier either by considering different recourse policies and/or allowing approximate second-stage costs for non-elementary q-routes. In this paper, we address these possibilities by replacing the second-stage cost of the VRPSD pricing problem with a generic *recourse cost function*. We show that, for any choice of a recourse cost function that satisfies a set of mild assumptions, the corresponding pricing problem for the VRPSD with scenarios is strongly $\mathcal{NP}$-hard. In doing so, we not only extend the results of Fukasawa & Gunter (2023) for alternative recourse policies, but we also prove that the VRPSD with scenarios is strongly $\mathcal{NP}$-hard regardless of the use of approximate second-stage costs for the non-elementary q-routes.

The second contribution of our work is as follows. Typical hardness results for pricing problems model the pricing problem as a generic optimization problem (with unconstrained costs), and then a reduction from a known hard problem to that generic optimization problem is shown. For example, one usually states that pricing elementary q-routes is strongly $\mathcal{NP}$-hard because the elementary shortest path problem with resource constraints (ESPPRC) is strongly $\mathcal{NP}$-hard (Irnich & Desaulniers, 2005). However, if we are interested in designing a branch(-cut)-and-price algorithm that uses only elementary q-routes, we can restrict ourselves to the case where the cost vector of the ESPPRC comes from the reduced costs of a restricted master problem. Very recently, Spliet (2023) has shown that, even if we assume that the cost vector has the form of reduced costs, pricing elementary q-routes for several VRP variants is still strongly $\mathcal{NP}$-hard. Similarly, we show that even when edge costs are restricted to the form of reduced costs, the pricing problem of the VRPSD with scenarios is still strongly $\mathcal{NP}$ hard.

Lastly, we mention that, contrary to most hardness results for VRP's pricing problems (Fukasawa & Gunter, 2023; Irnich & Desaulniers, 2005; Spliet, 2023), our reduction is not from the Hamiltonian cycle problem. Instead, we explore the fact that scenarios can be used to model the edges of a graph, and we derive our results via a reduction from the independent set problem. Loosely speaking, this is the main reason why our proof works for any recourse cost function that satisfies the previously mentioned assumptions.

The remainder of this paper is organized as follows. Section 2 defines the VRPSD and discusses set partitioning formulations. Section 3 presents our definitions of the recourse cost functions and the VRPSD pricing problem(s). The proofs of the complexity results are shown in Section 4. Section 5 concludes the paper and discusses further research directions.

2. Preliminaries

2.1. The vehicle routing problem with stochastic demands

We begin by defining the parameters in an instance of the vehicle routing problem with stochastic demands (VRPSD). Let $G = (V = \{0\} \cup V_+, E)$ be a complete undirected graph. The vertex 0 represents the depot and the set V_+ indicates the customers. With each edge $e \in E$, we associate a cost $c(e) \in \mathbb{Q}_{++}$. We may denote an edge (resp. arc) by either $\{u, v\}$ (resp. (u, v)) or uv. We also adopt the usual notation $\delta(v) := \{uw \in E : u = v\}$, for every $v \in V$. The letters $k \in \mathbb{Z}_{++}$ and $B \in \mathbb{Q}_{++}$ are used to refer to the desired number of routes and the vehicle capacity, respectively. Let d be a random vector with entries in V_+ indicating the random demand of each customer. The vector d is governed by a probability distribution $\mathbb{P}$, and we define $\bar{d} := \mathbb{E}[d] \in \mathbb{Q}_{++}^{V_+}$, where each entry of $\bar{d}$ is at most B. We remark that we enforce G to be complete, as well as c and $\bar{d}$ to be strictly positive because this conforms to the majority of the VRPSD benchmark instances (Laporte et al., 2002; Jabali et al., 2014; Florio et al., 2022).

As mentioned in the Introduction, the *VRPSD with scenarios* assumes that $\mathbb{P}$ is an empirical probability distribution of scenarios. Let S be a finite set of scenarios and for each $s \in S$, let d^s be a vector with entries $d^s_v \in \mathbb{Q}_+$, for every $v \in V_+$ (note that here we allow zero valued demands). Naturally, d^s_v indicates the demand of customer v in scenario s. We may regard d^s as a function, meaning that we write $d^s(v)$ instead of d^s_v. Each scenario $s \in S$ has an associated realization probability p_s, and these probabilities sum up to one. We say that $\mathbb{P}$ is *given by scenarios* if $\mathbb{P}(d = d^s) = p_s$, for all $s \in S$. An instance of the VRPSD is denoted by the tuple $\mathcal{I}_{\text{VRPSD}} = (G, \mathbb{P}, c, B, k)$ and we use Ω to refer to the set of all instances of the VRPSD with scenarios. (To avoid some technicalities, we consider two isomorphic graphs to be the same graph.) We always use G to denote graphs, so if we write $G \in \mathcal{I}_{\text{VRPSD}}$, we mean that G is the graph in the VRPSD instance $\mathcal{I}_{\text{VRPSD}}$.

Next, we discuss the concepts of q-routes and elementary q-routes. We define a *q-route* R of G as a tuple $(v_1, v_2, \ldots, v_\ell)$ such that

(i) $\{v_1, \ldots, v_\ell\} \subseteq V_+$;

(ii) $\sum_{j=1}^{\ell} \bar{d}(v_j) \leq B$; and

(iii) $\{\{0, v_1\}, \{v_1, v_2\} \ldots, \{v_{\ell-1}, v_\ell\}, \{v_\ell, 0\}\} \subseteq E$.

(Since we assume that G is complete, item (iii) always holds, but we still include it here for clarity.) We may regard R as a directed multigraph, meaning that $V(R)$ refers to the set $\{0, v_1, v_2, \ldots, v_\ell\}$ and $A(R)$ refers to the multiset $\{(0, v_1), (v_1, v_2), \ldots, (v_\ell, 0)\}$. We naturally extend the costs from the edges of G to the corresponding arcs, that is, if $a = (u, v)$ is an arc and $e = uv \in E$, then $c(a) = c(e)$. The first-stage cost of a q-route R is given by the sum of the cost of its arcs, i.e., $c(R) = \sum_{a \in A(R)} c(a)$. The *total expected demand* of q-route R, denoted by $\bar{d}(R)$, is the sum $\sum_{j=1}^{\ell} \bar{d}(v_j)$. A similar notation is also used for a scenario $s \in S$, so $d^s(R) = \sum_{j=1}^{\ell} d^s(v_j)$. The value of $\bar{d}(R)$ may be different from the sum $\sum_{v \in V(R) \cap V_+} \bar{d}(v)$ since q-routes allow the repetition of customers. We say that a q-route $R = (v_1, v_2, \ldots, v_\ell)$ is *elementary* if $v_1, \ldots, v_\ell$ are all pairwise distinct.

Let $R = (v_1, \ldots, v_\ell)$ be a q-route of G, we say that R *observes a failure* if $\mathbb{P}(d(R) > B) > 0$, that is, R has a non-zero probability of exceeding the vehicle capacity. *Recourse policies* describe *recourse actions* that aim to recover a vehicle from a failure. As we will see, our results are not tied to a specific recourse policy; still, for illustration purposes, we briefly mention the *classical recourse policy*. In this recourse policy, whenever collecting a customer demand causes the vehicle load to exceed its capacity, the vehicle returns to the depot to replenish its capacity and resume the route execution at the failed customer (see Dror (1990) for more details). When the customer's probability distributions are independent and have a convolution property (Gendreau et al., 2016), other more sophisticated recourse policies have been proposed, such as the *optimal restocking recourse policy* (Yee & Golden, 1980) and the *rule-based recourse policies* (Salavati-Khoshghalb et al., 2019b,a). However, to the best of our knowledge, no previous work addresses alternate recourse policies for the VRPSD with scenarios. Nevertheless, as we mentioned earlier, our results are fairly general and do not depend directly on the choice of the recourse policy.

For the moment, let us fix some recourse policy for the VRPSD with scenarios, and let $Q(R, d; \mathcal{I}_{\text{VRPSD}})$ be a random variable that indicates the (random) recourse cost of a q-route R under this fixed recourse policy. Formally speaking, given the parameters in $\mathcal{I}_{\text{VRPSD}}$, Q is a function of the tuple R and the random variable d; however, to simplify notation, we write simply $Q(R)$. A feasible solution for the VRPSD is a set of k elementary q-routes $\{R_1, \ldots, R_k\}$

that visits every customer exactly once, i.e., $\{V(R_i) \setminus \{0\}\}_{i\in[k]}$ is a partition of V_+. The objective of the problem is to minimize the sum $\sum_{i\in[k]}(c(R_i) + \mathbb{E}[Q(R_i)])$.

2.2. Set partitioning formulations

We now describe in more detail the set partitioning (SP) formulations used in state-of-the-art exact algorithms for the VRPSD (Christiansen & Lysgaard, 2007; Gauvin et al., 2014; Florio et al., 2020). For any q-route R of G and customer $v \in V_+$, we define $\textsc{count}(v, R)$ as being the number of times that vertex v appears in q-route R. Let $\mathcal{R}_q(\mathcal{I}_{\text{VRPSD}})$ be the set of all q-routes of $G \in \mathcal{I}_{\text{VRPSD}}$, when $\mathcal{I}_{\text{VRPSD}}$ is clear from the context, we write only $\mathcal{R}_q$. For each $\mathcal{R} \subseteq \mathcal{R}_q$ we consider the following associated SP formulation

$$\begin{aligned}
\mathrm{P}(\mathcal{R}) \qquad \min \quad & \sum_{R\in\mathcal{R}} (c(R) + \mathbb{E}[Q(R)]) \cdot \lambda_R \\
\text{s.t.} \quad & \sum_{R\in\mathcal{R}} \textsc{count}(v, R) \cdot \lambda_R = 1, \qquad \forall v \in V_+, && (1)\\
& \sum_{R\in\mathcal{R}} \lambda_R = k, && (2)\\
& \lambda \geq 0. && (3)
\end{aligned}$$

We make three observations regarding these formulations. First, we comment on the choice of the notation. We defined P so that it explicitly depends on the set $\mathcal{R}$ (and implicitly depends on the remaining parameters of $\mathcal{I}_{\text{VRPSD}}$), this is convenient for us because we can talk about different SP formulations depending on the choice of $\mathcal{R}$. For instance, suppose that $\mathcal{R}_+ \subseteq \mathcal{R}_q$ is any set containing $\mathcal{R}_e$, where $\mathcal{R}_e$ denotes the set of all elementary q-routes of G. In this case, we say that $\mathcal{R}_+$ is a *route-relaxation*, and the notation $\mathrm{P}(\mathcal{R}_+)$ refers to a valid formulation for the VRPSD (after enforcing integrality on the variables). Additionally, for most of this paper, we consider the column-generation procedure, and in this setting, our notation allows us to explicitly indicate the set of variables present in the restricted master problem.

The second observation is about different choices of route-relaxations $\mathcal{R}_+$ and the corresponding SP formulations. For example, if choose $\mathcal{R}_+ = \mathcal{R}_e$, then the LP relaxation of $\mathrm{P}(\mathcal{R}_+)$ is fairly strong, but at the expense of making the pricing problem strongly $\mathcal{NP}$-hard (even when there is no demand uncertainty, see (Irnich & Desaulniers, 2005; Spliet, 2023)). Hence, state-of-the-art exact algorithms for VRPs typically use route-relaxations $\mathcal{R}_+ \supsetneq \mathcal{R}_e$ that seeks a balance between the strength of the set partitioning formulation and the tractability of the pricing problem. For the case of the VRPSD, Christiansen & Lysgaard (2007) uses $\mathcal{R}_+$ as the set of q-routes without subcycles of length 2; while Gauvin et al. (2014) set $\mathcal{R}_+$ to be a route-relaxation based on ng-routes (Baldacci et al., 2011).

Finally, we mention that branch-cut-and-price algorithms for the VRPSD (Gauvin et al., 2014; Florio et al., 2020, 2022) also impose that λ belongs to a set Λ defined by additional valid linear inequalities, such as the rounded capacity inequalities (Laporte & Nobert, 1983) and the subset row cuts (Jepsen et al., 2008). As we argue later in Section 4.1, for the purpose of examining the complexity of the pricing problem, we may ignore such constraints without loss of generality.

Now let π and γ be the dual variables associated with constraints (1) and (2), respectively. Our definition of the VRPSD pricing problem depends on the following dual formulation of P.

$$
\begin{aligned}
\mathrm{D}(\mathcal{R}) \qquad \max \quad & \sum_{v \in V_+} \pi_v + \gamma k \\
\text{s.t.} \quad & \gamma + \sum_{v \in V_+} \textsc{count}(v, R) \cdot \pi_v \leq c(R) + \mathbb{E}[Q(R)], \qquad \forall R \in \mathcal{R}. \qquad (4)
\end{aligned}
$$

Fix $\mathcal{R}_+$ to be a route-relaxation and observe that formulation $\mathrm{P}(\mathcal{R}_+)$ may have an exponential number of variables (columns). We now briefly describe the column-generation (CG) procedure that solves $\mathrm{P}(\mathcal{R}_+)$. Suppose that we have $\mathcal{R} \subsetneq \mathcal{R}_+$, such that $|\mathcal{R}| \ll |\mathcal{R}_+|$ and $\mathrm{P}(\mathcal{R})$ is feasible. We call $\mathrm{P}(\mathcal{R})$ the *restricted master problem* (with respect to $\mathcal{R}_+$). Solving problem $\mathrm{P}(\mathcal{R})$ to optimality gives us dual multipliers $\bar{\pi}$ and $\bar{\alpha}$ which are optimal for $\mathrm{D}(\mathcal{R})$. Each iteration of the CG procedure solves the so called *pricing problem*, which can be seen as a separation problem in $\mathrm{D}(\mathcal{R}_+)$. In our context, the pricing problem checks if $(\bar{\pi}, \bar{\alpha})$ satisfy all constraints (4), or if there is a q-route $R \in \mathcal{R}_+ \setminus \mathcal{R}$ such that $c(R) + \mathbb{E}[Q(R)] - \sum_{v \in V_+} \textsc{count}(v, R) \cdot \bar{\pi}_v < \bar{\gamma}$. If this last case holds, the process repeats for $\mathrm{P}(\mathcal{R} \cup \{R\})$ and $\mathrm{D}(\mathcal{R} \cup \{R\})$; otherwise, $(\bar{\pi}, \bar{\alpha})$ is optimal for $\mathrm{D}(\mathcal{R}_+)$ and we are done by linear programming duality.

Let $\textsc{count}(uv, R)$ denotes how many times arc (u, v) appears in q-route R. Whenever we have a dual vector $\bar{\pi} \in \mathbb{Q}^{V_+}$, we assume that $\bar{\pi}_0 = 0$. Thus, we define

$$
c^{\bar{\pi}}(R) := \sum_{uv \in A(R)} \textsc{count}(uv, R) \cdot (c(uv) - \bar{\pi}_v) = \sum_{uv \in A(R)} \textsc{count}(uv, R) \cdot \left(c(uv) - \frac{\bar{\pi}_u + \bar{\pi}_v}{2} \right).
$$

The *reduced cost of edge uv* is given by $c^{\bar{\pi}}(uv) := c(uv) - \frac{1}{2}(\bar{\pi}_u + \bar{\pi}_v)$, while the *reduced cost of q-route* R is the sum $c^{\bar{\pi}}(R) + \mathbb{E}[Q(R)] - \bar{\gamma}$. Rewriting a violated inequality of type (4) yields

$$
c^{\bar{\pi}}(R) + \mathbb{E}[Q(R)] - \bar{\gamma} < 0.
$$

In other words, a q-route with negative reduced cost corresponds to a violated inequality in $\mathrm{D}(\mathcal{R}_+)$.

3. Recourse cost functions and the VRPSD pricing problems

Before we define the VRPSD pricing problem, we need to point out a detail that we have glossed over in Section 2. Recall that we have fixed the recourse policy *a priori*. Formally speaking, the VRPSD is, in fact, a *class of problems*, as its definition depends on the choice of the function $\mathbb{E}[Q(.)]$. Since our results are valid for the pricing problems of several different problems in this VRPSD class, we now define a class of VRPSD pricing problems, denoted $(\mathrm{Prc}(f))$. In these problems, we replace the term $\mathbb{E}[Q(R)]$ by a generic *recourse cost function* $f(R; \mathcal{I}_{\text{VRPSD}})$, where R is a q-route of G. Once again, to ease notation, we may omit $\mathcal{I}_{\text{VRPSD}}$ from the function parameters. Henceforth, whenever we write $\mathrm{P}(.)$ or $\mathrm{D}(.)$, we refer to the resulting set-partitioning based formulations with $\mathbb{E}[Q(R)]$ replaced by $f(R)$.

Problem. VRPSD pricing problem ($\mathrm{Prc}(f)$)
INSTANCE:

(a) an instance $\mathcal{I}_{\text{VRPSD}} = (G, \mathbb{P}, c, B, k)$ of the VRPSD with scenarios;

(b) a set $\mathcal{R} \subseteq \mathcal{R}_q$ such that $\mathrm{P}(\mathcal{R})$ is feasible; and

(c) a vector $\bar{\pi} \in \mathbb{Q}^{V_+}$ and a scalar $\bar{\gamma} \in \mathbb{Q}$ that are optimal for $\mathrm{D}(\mathcal{R})$.

GOAL: If there is a q-route $R \in \mathcal{R}_q \setminus \mathcal{R}$ such that $c^{\bar{\pi}}(R) + f(R) - \bar{\gamma} < 0$, returns YES and R; otherwise, returns NO.

Notice that we essentially defined ($\mathrm{Prc}(f)$) as a decision problem, while typical hardness results for pricing problems consider optimization problems. Moreover, the input for such optimization problems usually does not assume that the objective function coefficients are coupled with optimal dual multipliers. In fact, to the best of our knowledge, the only work that uses conditions similar to (b) and (c) is the recent one by Spliet (2023).

We now discuss how permitting alternative recourse cost functions f in the definition of the VRPSD pricing problem allows us to address another gap in the literature. For the moment, let us fix $\mathbb{E}[Q(.)]$ to be the expected recourse cost of the classical recourse policy. The complexity result of Fukasawa & Gunter (2023) is based on the observation that if a vertex is repeated in a q-route, then all of its occurrences are perfectly correlated. This insight is then used to prove that when the demands random variables follow independent normals (and under some additional technical conditions), problem (2SQ) can be used to solve the Hamiltonian cycle problem. Nevertheless, Christiansen & Lysgaard (2007) had previously designed a branch-and-price algorithm for the VRPSD that uses a pseudo-polynomial pricing routine. As demonstrated in Fukasawa & Gunter (2023), this seeming contradiction is explained by the fact that the pricing algorithm of Christiansen & Lysgaard (2007) may incorrectly compute the expected recourse cost of non-elementary q-routes. More precisely, Christiansen & Lysgaard (2007) solves formulation $\mathrm{P}(\mathcal{R}_q)$ with the term $\mathbb{E}[Q(R)]$ replaced by $g(R)$, where $0 \leq g(R) \leq \mathbb{E}[Q(R)]$ and $g(R)$ is guaranteed to be equal to $\mathbb{E}[Q(R)]$ only when R is elementary. Even though g is only an approximation of $\mathbb{E}[Q(.)]$, the condition that g coincides with $\mathbb{E}[Q(.)]$ at the elementary q-routes guarantees the correctness of Christiansen & Lysgaard (2007) branch-and-price algorithm. This same observation also holds for more recent branch-cut-and-price algorithms based on set partitioning formulations (Gauvin et al., 2014; Florio et al., 2020). As pointed out by Fukasawa & Gunter (2023), one may thus wonder if a similar situation holds for the case when $\mathbb{P}$ is given by scenarios. In other words, for the VRPSD with scenarios, if we only require the expected recourse costs to be computed exactly for elementary q-routes, it is not clear if it is possible to design a pseudo-polynomial time algorithm for solving the pricing problem. This question is our main motivation for studying the computational complexity of solving ($\mathrm{Prc}(f)$), for every choice of recourse cost function f that satisfies the following assumptions.

Assumptions. For every $\mathcal{I}_{\text{VRPSD}} \in \Omega$, with $G \in \mathcal{I}_{\text{VRPSD}}$, and every q-route R of G,

(A1) $f(R) \in \mathbb{Q}_+ \cup \{+\infty\}$;

(A2) the value $f(R)$ can be computed in polynomial-time with respect to $\mathcal{I}_{\text{VRPSD}}$;

(A3) if R is elementary and $\mathbb{P}(d(R) \leq B) = 1$, then $f(R) = 0$;

(A4) if R is elementary, then $f(R) \geq \mathbb{P}(d(R) > B) \cdot \min_{e \in \delta(0)}\{c(e)\}$.

Notice that, whenever R is non-elementary, only assumption (A1) constrains the value of $f(R)$. Therefore, the previously mentioned recourse cost function g used by Christiansen & Lysgaard (2007) satisfies all of the above assumptions. We prove Theorem 1 below in Section 4, and as a result, we show that, when $\mathbb{P}$ is given by scenarios, no choice of a recourse cost function g' — with $0 \leq g'(R) \leq \mathbb{E}[Q(R)]$ and $g'(R) = \mathbb{E}[Q(R)]$ whenever R is elementary — leads to a polynomial-time algorithm for the VRPSD pricing problem (unless $\mathcal{P} = \mathcal{NP}$).

Theorem 1. *Let f be a recourse cost function satisfying assumptions (A1)-(A4), then (Prc(f)) is strongly $\mathcal{NP}$-hard.*

Besides addressing the mentioned technical question on the hardness of VRPSD pricing problems that allows "approximate" expected recourse cost at non-elementary q-routes; assumptions (A1)-(A4) also let us handle many alternative recourse policies. Indeed, recourse actions should not decrease the cost of a solution (assumption (A1)) and, if possible, should be avoided (assumption (A3)). Moreover, if we seek to solve an optimization problem, it is reasonable that we should be able to evaluate its objective function quickly (assumption (A2)). Assumption (A4) was inspired by the fact that all of the recourse policies mentioned in Section 2.1 (classical, optimal restocking and rule-based recourse policies) prescribe recourse actions that unload the vehicle at the depot. Note that some lower bound assumption like (A4) is necessary, as otherwise, we may choose $f \equiv 0$ and (Prc(f)) becomes a deterministic VRP pricing problem, which is pseudo-polynomial solvable, but also provides no information about the effects of uncertainty.

It turns out that even when the recourse policy prescribes recourse actions with costs lower than a trip to the depot, the corresponding VRPSD pricing problem might still be strongly $\mathcal{NP}$-hard. Consider the following less constrained assumption.

(A4') There exists a constant $\alpha \in (0,1) \cap \mathbb{Q}$ such that, for all $\mathcal{I}_{\text{VRPSD}} \in \Omega$, with $G \in \mathcal{I}_{\text{VRPSD}}$, and every q-route R of G, if R is elementary, then

$$f(R) \geq \alpha \cdot \mathbb{P}(d(R) > B) \cdot \min_{e \in \delta(0)} \{c(e)\}.$$

We show in Section 4 that replacing (A4) by (A4') does not make the pricing problem easier.

Corollary 2. *Let f be a recourse cost function satisfying assumptions (A1)-(A3) and assumption (A4'), then (Prc(f)) is strongly $\mathcal{NP}$-hard.*

Another interesting implication of Theorem 1 is as follows. For each $\mathcal{I}_{\text{VRPSD}} \in \Omega$, let $\mathcal{R}_+(\mathcal{I}_{\text{VRPSD}})$ be a route-relaxation of $\mathcal{I}_{\text{VRPSD}}$. Consider the pricing problem of formulation $\text{P}(\mathcal{R}_+(\mathcal{I}_{\text{VRPSD}}))$ with $\mathbb{E}[Q(.)]$ replaced by f. In other words, consider the decision problem (Prc($\mathcal{R}_+, f$)) defined similarly to (Prc(f)) except that in item (b) we have $\mathcal{R} \subseteq \mathcal{R}_+(\mathcal{I}_{\text{VRPSD}})$ and our goal is to check if there is negative reduced-cost q-route in $\mathcal{R}_+(\mathcal{I}_{\text{VRPSD}}) \setminus \mathcal{R}$. Corollary 3 below shows that, if f satisfies assumptions (A1)-(A4), problem (Prc($\mathcal{R}_+, f$)) is also strongly $\mathcal{NP}$-hard.

Corollary 3. *For each $\mathcal{I}_{\text{VRPSD}} \in \Omega$, let $\mathcal{R}_+(\mathcal{I}_{\text{VRPSD}})$ be a route-relaxation of $\mathcal{I}_{\text{VRPSD}}$. Let f be a recourse cost function satisfying assumptions (A1)-(A4), then (Prc($\mathcal{R}_+, f$)) is strongly $\mathcal{NP}$-hard.*

Proof. Choose f' such that, for all $\mathcal{I}_{\text{VRPSD}} \in \Omega$, $f'(R; \mathcal{I}_{\text{VRPSD}}) = f(R; \mathcal{I}_{\text{VRPSD}})$ if R belongs to $\mathcal{R}_+(\mathcal{I}_{\text{VRPSD}})$; and $f'(R; \mathcal{I}_{\text{VRPSD}}) = +\infty$ otherwise. Problem (Prc($\mathcal{R}_+, f$)) is equivalent to problem (Prc(f')). $\square$

4. Complexity of VRPSD pricing problems

Let f be a recourse cost function satisfying (A1)-(A4), we first show a construction that, given a graph and an independent set U, uses (Prc(f)) to check if the graph contains an independent set U' larger than U. Theorem 1 then follows by solving the independent set problem with a polynomial number of calls to an algorithm that solves (Prc(f)).

We start by formally defining the independent set problem (which is well known to be strongly $\mathcal{NP}$-hard).

Problem. Independent set problem
INSTANCE: an undirected graph H and an integer $t \in \mathbb{Z}_{++}$.
GOAL: returns YES if H has an independent set of size at least t, and returns NO otherwise.

4.1. Reducing the independent set problem to (Prc(f))

Let $\mathcal{I}_{\text{IS}} = (H, t)$ be an instance of the independent set problem and let U be an independent set of H with size $t-1$. Define $m := |E(H)|$ and label the vertices of H as $\{1, \ldots, n\}$. We assume that $n \geq 2$, $m \geq n+1$ and H is connected. Without loss of generality, we also assume that $n \notin U$. The purpose of this section is to describe the construction of an instance $\mathcal{I}_{\text{PRC}} = (\mathcal{I}_{\text{VRPSD}}, \mathcal{R}, \bar{\pi}, \bar{\gamma})$ of (Prc(f)) that corresponds to the instance $\mathcal{I}_{\text{IS}}$ and the independent set U. We also prove some simple results that (hopefully) motivate our chosen objects. In Section 4.2 we then demonstrate that solving instance $\mathcal{I}_{\text{PRC}}$ of (Prc(f)) is equivalent to solving the independent set instance $\mathcal{I}_{\text{IS}}$.

Consider the graph $G = (V = \{0\} \cup V_+, E = E_1 \cup E_2 \cup E_3)$ where

$$
\begin{aligned}
V_+ &= V(H) \cup \{r^1, r^2, a\} \cup \{r_i^1, r_i^2, w_i, a_i, b_i\}_{i \in [n]}, \\
E_1 &= \{\{0, r^1\}, \{r^1, a\}, \{a, r^2\}, \{w_n, 0\}\} \cup \{\{w_{i-1}, i\}, \{i, w_i\}, \{w_{i-1}, w_i\}\}_{i \in [n]}, \\
E_2 &= \{\{0, r_i^1\}, \{r_i^1, a_i\}, \{a_i, r_i^2\}, \{r_i^2, i\}, \{r_i^2, b_i\}, \{i, b_i\}, \{b_i, 0\}\}_{i \in [n]} \\
E_3 &= \{uv : u, v \in V, u \neq v, uv \notin E_1 \cup E_2\}.
\end{aligned}
$$

To ease exposition, we sometimes regard w_0 as r^2. An illustration of the graph G is shown in Figure 1.

Let ε be a positive number smaller than $\frac{1}{n}$. The vehicle capacity is set to $B_2 = 9B_1$, where $B_1 = (n+1)m + \varepsilon$. The number of vehicles is set to $k = n+1$. Let M be a rational number larger than nB_2. The cost of each edge $uv \in E$ is set as follows.

$$
c(uv) = \begin{cases} \varepsilon, & \text{if } uv \in (E_1 \cup E_2) \setminus (\delta(0) \cup \{\{r_i^2, i\}\}_{i \in [n]}) \\ 1 + \varepsilon, & \text{if } uv \in \{\{r_i^2, i\}\}_{i \in [n]}, \\ M, & \text{if } uv \in \delta(0) \cap (E_1 \cup E_2), \\ 6B_2 M, & \text{if } uv \in E_3. \end{cases}
$$

As we prove later (see Claim 2), edges in E_3 have a "very large cost", in the sense that if G has a negative reduced cost q-route, then the edges of this q-route belongs to $E_1 \cup E_2$.

The set S of scenarios of the input probability distribution $\mathbb{P}$ have cardinality $m+n+3$, and the scenarios in S are labeled as $E(H) \dot\cup \{r^1, r^2, s_b\} \dot\cup \{s_i\}_{i \in [n]}$. With each scenario $s \in S$, we associate a vertex set $V(s) \subseteq V_+$ and a realization probability p_s as follows.

$$(V(s), p_s) = \begin{cases} \left(\{i,j\}, \frac{1}{B_2}\right), & \text{if } s = ij \in E(H), \\ \left(\{r^1, r^1_1, \ldots, r^1_n\}, \frac{5B_1}{B_2}\right), & \text{if } s = r^1, \\ \left(\{r^2, r^2_1, \ldots, r^2_n\}, \frac{3B_1}{B_2}\right), & \text{if } s = r^2, \\ \left(\{a, b_1, \ldots, b_n\}, \frac{\varepsilon}{B_2}\right), & \text{if } s = s_b, \\ \left(\{w_i, a_1, \ldots, a_n\}, \frac{m}{B_2}\right), & \text{if } s = s_i, \text{ for some } i \in [n]. \end{cases}$$

It is easy to check that $\sum_{s \in S} p_s = 1$, so indeed the scenarios in S give a valid input probability distribution $\mathbb{P}$.

Let $\mathbb{I}(.)$ be the indicator function. We set the demand of customer $v \in V_+$ in scenario $s \in S$ as $d^s(v) = B_2 \cdot \mathbb{I}(v \in V(s))$. Hence, the expected demand of each customer $v \in V_+$ is

$$\bar{d}(v) = \begin{cases} |\delta_H(v)|, & \text{if } v \in V(H), \\ 5B_1, & \text{if } v \in \{r^1, r^1_1, \ldots, r^1_n\}, \\ 3B_1, & \text{if } v \in \{r^2, r^2_1, \ldots, r^2_n\}, \\ \varepsilon, & \text{if } v \in \{a, b_1, \ldots, b_n\}, \\ nm, & \text{if } v \in \{a_1, \ldots, a_n\}, \\ m, & \text{if } v \in \{w_1, \ldots, w_n\}. \end{cases}$$

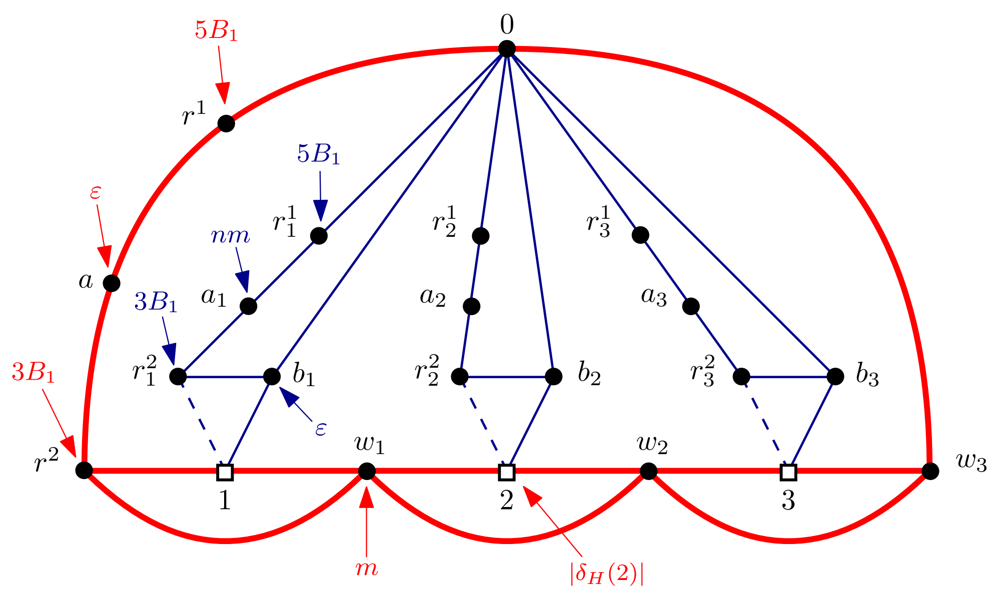

Figure 1: Illustration of the constructed graph G when H is a graph with vertex set $\{1, 2, 3\}$. We only show the edges in $E_1 \cup E_2$. The edges in E_1 (resp. E_2) are shown in red and bold lines (resp. blue and thin lines). The dashed lines indicates edges with costs $1 + \varepsilon$. The numbers next to the arrows refer to the expected demands.

The choice of the scenario demands implies some rather simple but relevant facts. Henceforth, for any subset of vertices $V' \subseteq V_+$ and q-route R, we use $\textsc{count}(V', R)$ as a short-hand for $\sum_{v \in V'} \textsc{count}(v, R)$.

Fact 1. *Let R be a q-route of G that observes no failures, then $V(R) \cap V(H)$ is an independent set of H.*

Proof. Since R observes no failures, we know that $|V(R) \cap V(s)| \leq 1$, for all $s \in S$. In particular, for each $s = ij \in E(H)$, we have that $|V(R) \cap V(H) \cap \{i, j\}| = |V(R) \cap \{i, j\}| \leq 1$. Hence, $V(R) \cap V(H)$ is an independent set of H. □

Fact 2. *For every q-route R of G it holds that* $\textsc{count}(\{r^1\} \cup \{r_i^1\}_{i\in[n]}, R) \leq 1$. *Moreover, if $E(R) \subseteq E_1 \cup E_2$, it also holds that* $\textsc{count}(\{r^1, r^2\} \cup \{r_i^1, r_i^2\}_{i\in[n]}, R) \leq 2$.

Proof. For every vertex $v \in \{r^1\} \cup \{r_i^1\}_{i\in[n]}$, we have that $2 \cdot \bar{d}(v) = 10B_1 > 9B_1 = B_2$, proving the first part of the statement. Now suppose by contradiction that the edges of R belongs to $E_1 \cup E_2$ and $\textsc{count}(\{r^1, r^2\} \cup \{r_i^1, r_i^2\}_{i\in[n]}, R) \geq 3$. Then q-route R also contains a customer u that does not belong to $\{r^1, r^2\} \cup \{r_i^1, r_i^2\}_{i\in[n]}$. This implies that $\bar{d}(R) \geq 3 \cdot \bar{d}(r^2) + \bar{d}(u) > B_2$. □

We now have a VRPSD instance $\mathcal{I}_{\text{VRPSD}} = (G, \mathbb{P}, c, B, k)$. Since our objective is to prove that (Prc(f)) is strongly $\mathcal{NP}$-hard, we also mention that M and ε can be chosen so that their unary encoding is polynomial on n. In order to have an instance of the VRPSD pricing problem, it remains to construct $\mathcal{R}$ and $(\bar{\pi}, \bar{\gamma})$. However, before we show the construction, we introduce a definition and prove a related fact.

Definition 1. *Let $T \subseteq V(H)$, we write R_T to denote the elementary q-route of G with arc set* $A(R_T) = \{(0, r^1), (r^1, a), (a, r^2), (w_n, 0)\} \cup \{(w_{i-1}, i), (i, w_i)\}_{i\in T} \cup \{(w_{i-1}, w_i)\}_{i\in V(H)\setminus T}$.

Fact 3. *Let $\vec{R} = (v_1, \ldots, v_\ell)$ and $\overleftarrow{R} = (v_\ell, \ldots, v_1)$ be elementary q-routes of G with $\ell \geq 2$ and $E(\vec{R}) = E(\overleftarrow{R}) \subseteq E_1$, then there exists $T \subseteq V(H)$ such that $A(R_T)$ is equal to either $A(\vec{R})$ or $A(\overleftarrow{R})$.*

Proof. Since $E(\vec{R}) = E(\overleftarrow{R}) \subseteq E_1$, either $v_1 = r^1$ or $v_\ell = r^1$. Without loss of generality, assume $v_1 = r^1$. Since $\ell \geq 2$, we have $v_2 = a$. Moreover, vertex r^1 does not belong to $\{v_3, \ldots, v_\ell\}$, since $5B_1 = \bar{d}(r^1) > B_2 - \bar{d}(v_1) - \bar{d}(v_2) = 4B_1 - \varepsilon$. This implies that $v_3 = r^2$ and $v_\ell = w_n$ (since the depot is only adjacent to r^1 and w_n in (V, E_1)). Since $\vec{R}$ is elementary, $\vec{R}$ visits each vertex in $\{r^2, w_1, \ldots, w_n\}$ exactly once and $\vec{R} = R_T$ with $T = V(\vec{R}) \cap V(H)$. □

For each $i \in [n]$, let us define the q-routes $R_i \coloneqq (r_i^1, a_i, r_i^2, b_i)$ and $R_i' \coloneqq (r_i^1, a_i, r_i^2, i, b_i)$. We set $\mathcal{R} = \{R_U\} \cup \{R_i\}_{i\in U} \cup \{R_i'\}_{i\in V(H)\setminus U}$ (see Figure 2). For all $i \in [n]$, it is clear that both $\bar{d}(R_i)$ and $\bar{d}(R_i')$ are at most B_2. To see that $\bar{d}(R_U) \leq B_2$, observe that $\bar{d}(V(R_U) \cap \{w_1, \ldots, w_n\}) = nm$; furthermore, since U is an independent set, $\bar{d}(V(R_U) \cap V(H)) = \sum_{i\in U} |\delta_H(i)| \leq m$.

Observe that $\{V(R) \setminus \{0\}\}_{R\in\mathcal{R}}$ forms a partition of V_+. Therefore, as $|\mathcal{R}| = n + 1 = k$, the only feasible solution to P($\mathcal{R}$) is $\bar{\lambda} \in \{0, 1\}^{\mathcal{R}_q}$ with $\bar{\lambda}_R = 1$ if and only if $R \in \mathcal{R}$. Every q-route $R \in \mathcal{R}$ observes no failures, so by assumption (A3), $f(R) = 0$. The optimal value for P($\mathcal{R}$) is thus

$$c(R_U) + \sum_{i\in U} c(R_i) + \sum_{i\in V(H)\setminus U} c(R_i') = 2M(n + 1) + |V(H) \setminus U| + \varepsilon(|V_+| - (n + 1)).$$

Let us now construct a dual feasible solution $(\bar{\pi}, \bar{\gamma})$ with the same objective function value. Set $\bar{\gamma} = 0$, and for each $v \in V_+$, set

$$\bar{\pi}_v = \begin{cases} 4M + \varepsilon, & \text{if } v = a, \\ 5M + \varepsilon, & \text{if } v \in \{a_1, \ldots, a_n\}, \\ -2M, & \text{if } v = w_n, \\ -3M, & \text{if } v \in \{b_1, \ldots, b_n\}, \\ 1 + \varepsilon, & \text{if } v \in V(H), \\ -1 + \varepsilon, & \text{if } v = w_i \text{ and } i \in U, \\ \varepsilon, & \text{otherwise.} \end{cases}$$

Summing the dual values yields $\bar{\gamma} + \sum_{v \in V_+} \bar{\pi}_v = 2M(n+1) + |V(H) \setminus U| + \varepsilon(|V_+| - (n+1))$ (recall that $n \notin U$). One can also check that $c^{\bar{\pi}}(R) + f(R) = c^{\bar{\pi}}(R) = 0$, for all $R \in \mathcal{R}$; so $(\bar{\pi}, \bar{\gamma})$ is optimal for D$(\mathcal{R})$ (see Figure 2).

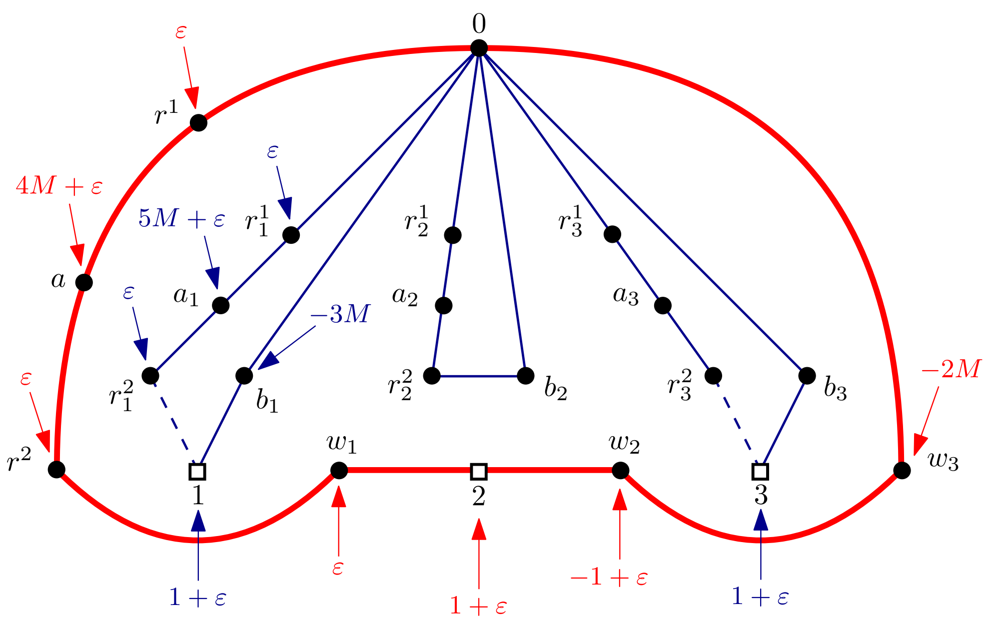

Figure 2: Illustration of the set $\mathcal{R}$ and the dual vector $\bar{\pi}$ with respect to the graph in Figure 1 and $U = \{2\}$. The q-route R_U is shown in red and bold lines, while the q-routes R_1', R_2 and R_3' are shown in blue and thin lines. Again, the dashed lines indicate edges with cost $1 + \varepsilon$. The numbers next to an arrow pointing to a vertex $v \in V_+$ refer to the dual value $\bar{\pi}_v$. The reader can check that every q-route in $\mathcal{R}$ has zero reduced cost.

In fact, the following claim gives an easy expression to calculate the reduced cost of q-route R_T, for any $T \subseteq V(H)$.

Lemma 4. *For all $T \subseteq V(H)$, $c^{\bar{\pi}}(R_T) = |U| - |T|$.*

Proof. We show that $c^{\bar{\pi}}(R_T) = \sum_{uv \in A(R_T)} \text{COUNT}(uv, R_T) \cdot (c(uv) - \bar{\pi}_v) = |U| - |T|$. Define $A_1 := \{(w_{i-1}, i), (i, w_i)\}_{i \in T}$ and $A_2 := \{(w_{i-1}, w_i)\}_{i \in V(H) \setminus T}$. Additionally, define a vector $\pi' \in \mathbb{R}^{V_+}$ obtained from $\bar{\pi}$ as follows: $\pi'_{w_n} = \varepsilon$ and $\pi'_v = \bar{\pi}_v$, for all $v \in V_+ \setminus \{w_n\}$. Notice that (check Figure 2)

$$\sum \left(c(a) : a \in \{(0, r^1), (r^1, a), (a, r^2), (w_n, 0)\}\right) - \sum \left(\bar{\pi}_v : v \in \{r^1, r^2, a, w_n\}\right) = -\varepsilon,$$

so we can rewrite the reduced cost of R_T as $c^{\bar{\pi}}(R_T) = \sum_{uv \in A_1 \cup A_2} \text{COUNT}(uv, R) \cdot (c(uv) - \pi'_v)$.

To evaluate $c^{\bar{\pi}}(R_T)$, take a vertex i in $V(H)$. Then $c((w_{i-1}, i)) + c((i, w_i)) - \pi'_i - \pi'_{w_i} = -1 + \mathbb{I}(i \in U)$ and $c((w_{i-1}, w_i)) - \pi'_{w_i} = \mathbb{I}(i \in U)$. Therefore, $\sum_{uv \in A_1} \text{COUNT}(uv, R) \cdot (c(uv) - \pi'_v) = -|T \setminus U|$ and $\sum_{uv \in A_2} \text{COUNT}(uv, R) \cdot (c(uv) - \pi'_v) = |U \setminus T|$. We conclude that

$$\sum_{uv \in A(R_T)} \text{COUNT}(uv, R_T) \cdot (c(uv) - \bar{\pi}_v) = |U \setminus T| - |T \setminus U| = |U| - |T|.$$

□

Remark 1. *We can interpret $\bar{\pi}$ with a discharging argument (Cranston & West, 2017): for each q-route $R = (v_1, \ldots, v_\ell) \in \mathcal{R}$, each edge in $e = \{v_j, v_{j+1}\} \in E(R)$ starts with a charge of $c(e)$ that is "distributed" among the customers in $V(R)$; by the end of the process, the total charge at node $v \in V_+$ is exactly $\bar{\pi}_v$. For instance, if $R = R_i$ the charges of edges $\{0, r_i^1\}$ and $\{0, b_i\}$ are distributed by sending $5M$ to vertex a_i and $-3M$ to vertex b_i. The remaining edges $\{v_j, v_{j+1}\} \in E(R_i)$ each send its charges to vertex v_{j+1}, with the exception of edge $\{i, b_i\}$ that send its charges to r_i^1. A similar procedure is applied when $R = R'_i$. The charges of the edges in R_U are distributed as follows. The two edges in $E(R_U) \cap \delta(0)$ send $4M$ of charge to vertex a and $-2M$ of charge to vertex w_n. For each of the remaining edges $e = \{v_j, v_{j+1}\} \in E(R_U)$, we have three cases: if $v_{j+1} = w_n$, then e send its charges to r^1; if $\{v_j, v_{j+1}\} = \{i, w_i\}$, then e send 1 of charge to i and $-1 + \varepsilon$ of charge to w_i; if none of the previous cases hold, e send its charges to v_{j+1}. Since the discharging procedure do not alter the net total charge, it follows that the primal and dual objectives are the same.*

Remark 2. *Recall that we mentioned in Section 2.2 that branch-cut-and-price algorithms for the VRPSD (Gauvin et al., 2014; Florio et al., 2020, 2022) also impose that λ belongs to a set Λ defined by additional valid linear inequalities. The solution $\bar{\lambda}$ that we built here is integral and feasible for $\mathrm{P}(\mathcal{R})$, so $\bar{\lambda}$ satisfies the constraints in Λ. Suppose now that we want to consider the constraints in Λ in the VRPSD pricing problem. Our construction shows that, for the purposes of studying the hardness of the pricing problem, we can assume that the dual variables associated with the constraints in Λ are all set to zero.*

4.2. Proving that (Prc(f)) solves the independent set problem

Consider the instances $\mathcal{I}_{\text{IS}} = (H, t)$ and $\mathcal{I}_{\text{PRC}} = (\mathcal{I}_{\text{VRPSD}}, \mathcal{R}, \bar{\pi}, \bar{\gamma})$ defined in Section 4.1. The main step now is to show that $\mathcal{I}_{\text{IS}}$ is a YES instance if and only if $\mathcal{I}_{\text{PRC}}$ is also a YES instance. As Claim 1 shows, one direction is an immediate implication of Lemma 4.

Claim 1. *Suppose that T is an independent set of H with size greater than U, then R_T has negative reduced cost.*

Proof. For all $s \in S$, $|V(R_T) \cap V(s)| \leq 1$, so R_T observes no failures. By assumption (A3), $f(R_T) = 0$. Using Lemma 4 we find that the reduced cost of R_T is $|U| - |T| < 0$. □

The converse direction requires more work; so we briefly comment on the intuition for the proof. Let $R^* = (v_1, \ldots, v_\ell)$ be a q-route of G with negative reduced cost. Vertex a contributes with a "large negative value" of $-5M - \varepsilon$ to the reduced cost of a q-route. Since R^* avoids edges in E_3 (see Claim 2 below), we know that if $v_j = a$, for some $j \in [\ell] \setminus \{1\}$, then $\{v_{j-1}, v_{j+1}\} = \{r^1, r^2\}$. Therefore, by Fact 2, R^* cannot contain a more than once. A similar argument also holds for a_i and r_i^1, r_i^2, for every $i \in [n]$. We can leverage this reasoning to prove that $E(R^*) \subseteq E_1$; additionally, since every vertex in $\{w_i\}_{i\in[n]}$ have expected demand m, we can also show that R^* is elementary. Applying Fact 3, Lemma 4 and assumption (A4) we then learn that R^* observes no failures. Finally, by Fact 1 and Lemma 4, it follows that $V(R^*) \cap V(H)$ is an independent set of H larger than U.

Before we proceed, we need some extra notation. For any q-route R of G, we use $\bar{\pi}(R)$ as a short-hand for $\sum_{v\in V(R)} \text{COUNT}(v, R)\cdot\bar{\pi}_v$, so $c^{\bar{\pi}}(R) = c(R) - \bar{\pi}(R)$. We will also need to refer to the contributions of vertices in $\mathcal{A} := \{a, a_1, \dots, a_n\}$ and $\mathcal{B} := \{w_n, b_1, \dots, b_n\}$ to $\bar{\pi}(R^*)$. Hence, we define

$$
\begin{aligned}
p_a &:= \sum_{v\in\mathcal{A}} \text{COUNT}(v, R^*) \cdot \bar{\pi}_v, \\
p_b &:= \sum_{v\in\mathcal{B}} \text{COUNT}(v, R^*) \cdot \bar{\pi}_v, \\
p &:= \bar{\pi}(R^*) - p_a - p_b,
\end{aligned}
$$

so $c^{\bar{\pi}}(R^*) = c(R^*) - (p + p_a + p_b)$.

Claim 2. $E(R^*) \subseteq E_1 \cup E_2$

Proof. Notice that for every q-route R of G we have that $\bar{\pi}(R) < 6B_2M$. Therefore, if R^* had an edge in $E(R^*) \cap E_3$, then R^* would have positive reduced cost. □

Claim 3. $\text{COUNT}(\mathcal{A}, R^*) = \text{COUNT}(\mathcal{B}, R^*) = 1$.

Proof. By Claim 2 and Fact 2, we know that $\text{COUNT}(\mathcal{A}, R^*) \le 1$. To see that $\text{COUNT}(\mathcal{A}, R^*) = 1$, note that

$$c^{\bar{\pi}}(R^*) + f(R^*) \ge c(R^*) - \bar{\pi}(R^*) = c(R^*) - p - p_a - p_b > M + \varepsilon - p_b - p_a,$$

where the first inequality follows from (A1) and the second inequality follows from $c(R^*) \ge 2M + \varepsilon$ and $p < (1+\varepsilon)B_2 < 2B_2 < nB_2 < M$. Since R^* has negative reduced cost, we get

$$M + \varepsilon - p_b - p_a < c^{\bar{\pi}}(R^*) + f(R^*) < 0. \tag{5}$$

This implies that $p_a > M + \varepsilon - p_b \ge 0$, since $p_b \le 0$. Therefore, $\text{COUNT}(\mathcal{A}, R^*) = 1$ and $p_a \le 5M + \varepsilon$.

Inequality (5) also implies that $-p_b < p_a - M - \varepsilon \le 4M$, and thus $\text{COUNT}(\mathcal{B}, R^*) \le 1$, since otherwise $-p_b \ge 4M$. To finish the proof, notice that, by the way we set the demands, $\text{COUNT}(\{r^1, r_1^1, \dots, r_n^1\}, R^*) \le 1$. Therefore, since the depot is only adjacent to vertices in $\{r^1, r_1^1, \dots, r_n^1\} \cup \mathcal{B}$ in the graph $(V, E_1 \cup E_2)$, it follows from Claim 2 that $\text{COUNT}(\mathcal{B}, R^*) \ge 1$. □

Claim 4. *$(v_1, v_2, v_3, v_\ell) = (r^1, a, r^2, w_n)$ or symmetrically, $(v_\ell, v_{\ell-1}, v_{\ell-2}, v_1) = (r^1, a, r^2, w_n)$.*

Proof. Since R^* has negative reduced cost, R^* has at least 2 customers. Thus, by Claim 2, either (1) $(v_1, v_\ell) \in \{r^1, r_1^1, \dots, r_n^1\} \times \mathcal{B}$ or (2) $(v_\ell, v_1) \in \mathcal{B} \times \{r^1, r_1^1, \dots, r_n^1\}$ holds. By symmetry, we assume (1), in which case $v_2 \in \mathcal{A}$ (since $\ell \ge 2$) and $v_3 \in \{r^2, r_1^2, \dots, r_n^2\}$

(since $\bar{d}(v_1) = 5B_1$). We show that v_1 cannot be equal to r_i^1, for any $i \in [n]$. Suppose not, then $(v_2, v_3) = (a_i, r_i^2)$. We have two cases. If $v_4 = b_i$, then by Claims 2 and 3 we know that $R^* = R_i \in \mathcal{R}$, a contradiction. Otherwise, by Claims 2 and 3, $v_4 = i$ and $v_5 \notin \{w_{i-1}, w_i, r_i^2\}$ since $\bar{d}(r_i^2) > \bar{d}(w_{i-1}) = \bar{d}(w_i) = m$ and $\bar{d}(i) > \varepsilon$. So $v_5 = b_i$ and $R^* = R_i'$, a contradiction. We conclude that $(v_1, v_2, v_3) = (r^1, a, r^2)$.

To see that $v_\ell = w_n$, suppose by contradiction that $v_\ell = b_i$, for some $i \in [n]$. By Claim 3, we then have that $p_a = 4M + \varepsilon$, $p_b = -3M$, $c(R^*) \geq 2M + \varepsilon$ and $p < (1+\varepsilon)B_2 < 2B_2 < mB_2 = M$. By assumption (A1), we then get the following contradiction,

$$c^{\bar{\pi}}(R^*) + f(R^*) \geq c(R^*) - p - p_a - p_b > 0.$$

□

Claim 5. *R^* is elementary and observes no failures.*

Proof. By Claim 4, assume that $(v_1, v_2, v_3, v_\ell) = (r^1, a, r^2, w_n)$. Then $B_2 - \bar{d}(r^1) - \bar{d}(a) - \bar{d}(r^2) = B_1 - \varepsilon = (n+1)m$, so the sum of the average demands of $v_3, \ldots, v_\ell$ is at most B_1, and therefore, R^* avoids vertices in $\{r_i^2\}_{i \in [n]}$. Since $\text{COUNT}(\mathcal{B}, R^*) = 1$ and $v_\ell = w_n$, R^* also avoids vertices in $\{b_i\}_{i \in [n]}$. Therefore, $E(R^*) \subseteq E_1$ and R^* contains all vertices in $\{w_1, \ldots, w_n\}$.

Now suppose that R^* is not elementary, then R^* visits a vertex $w \in \{w_0, \ldots, w_n\}$ at least twice, and thus, $\bar{d}(v_3) + \ldots + \bar{d}(v_\ell) > m(n+1) = B_1 - \varepsilon$, a contradiction. Therefore, R^* is elementary, and by Fact 3, we know that $R^* = R_T$, for some $T \subseteq V(H)$. Using Lemma 4, we have that $c^{\bar{\pi}}(R^*) = |U| - |T| \geq -n$. We conclude that R^* observes no failures, since otherwise, by (A1) and (A4), we have that

$$c^{\pi}(R^*) + f(R^*) \geq -n + \frac{M}{B_2} > 0.$$

□

We close this section by combining all of the results seen so far. (To facilitate the reading, we restate Theorem 1.)

Lemma 5. *Let $\mathcal{I}_{\text{IS}} = (H, t)$ be an instance of the independent set problem where H is a connected graph with $n \geq 2$ vertices and $m \geq n+1$ edges. Let U be an independent set in H with size $t-1$. Then, for every recourse cost function f satisfying assumptions (A1)-(A4), there exists an instance $\mathcal{I}_{\text{PRC}}$ of (Prc(f)) satisfying the following.*

(1) The instance $\mathcal{I}_{\text{PRC}}$ can be constructed in polynomial time with respect to $\mathcal{I}_{\text{IS}}$ and the size of a unary encoding of $\mathcal{I}_{\text{PRC}}$ is polynomial on the size of $\mathcal{I}_{\text{IS}}$; and

(2) $\mathcal{I}_{\text{IS}}$ is a YES *instance if and only if $\mathcal{I}_{\text{PRC}}$ is also a* YES *instance.*

Proof. Let $\mathcal{I}_{\text{PRC}} = (\mathcal{I}_{\text{VRPSD}}, \mathcal{R}, \bar{\pi}, \bar{\gamma})$ be the instance of (Prc(f)) shown in Section 4.1. It is clear that $\mathcal{I}_{\text{PRC}}$ satisfies the conditions of item (1). By Claim 1, we only need to show that if R^* is a q-route of $G \in \mathcal{I}_{\text{VRPSD}}$ with negative reduced cost, then there is an independent set in H with larger size than U. Define $U' := V(R^*) \cap V(H)$. It follows from Claim 5 and (A3) that $f(R^*) = 0$. Additionally, by Fact 1, U' is an independent set in H. Hence, by Fact 3, $c^{\bar{\pi}}(R^*) = |U| - |U'| < 0$. □

Theorem 1. *Let f be a recourse cost function satisfying assumptions (A1)-(A4), then (Prc(f)) is strongly $\mathcal{NP}$-hard.*

Proof. Let $\mathcal{I}_{\text{IS}} = (H, t)$ be an instance of the independent set problem with $n \coloneqq |V(H)| \geq 2$, $m \coloneqq |E(H)| \geq n + 1$ and $t \leq n - 1$. Suppose that we have an algorithm $\mathcal{A}$ that solves (Prc(f)). We now describe an algorithm that maintains an independent set U and uses $\mathcal{A}$ to increase U at each iteration (if possible). Clearly, this algorithm finishes with at most t calls to $\mathcal{A}$.

Start with $U \leftarrow \emptyset$. If $|U| \geq t$, we return YES. Otherwise, we construct an instance $\mathcal{I}_{\text{PRC}} = (\mathcal{I}_{\text{VRPSD}}, \mathcal{R}, (\bar{\pi}, \bar{\gamma}))$ using Lemma 5 and set U. If $\mathcal{A}(\mathcal{I}_{\text{PRC}})$ returns NO, it follows from Lemma 5 that we can stop the algorithm and return NO. Otherwise, $\mathcal{A}(\mathcal{I}_{\text{PRC}})$ returns a q-route R^* with negative reduced cost. Hence, by Claim 5, R^* is elementary and observes no failures. Fact 1 then implies that $U' = V(R^*) \cap V(H)$ is an independent set in H larger than U. Therefore, we update U to U' and repeat the procedure. □

Remark 3. *Suppose that instead of having f satisfying (A1)-(A4), we have the setting of Corollary 2, where f satisfies (A1)-(A3) and (A4') with constant α. Then we can simply choose $M \geq \frac{1}{\alpha} n B_2$ when constructing instance $\mathcal{I}_{\text{PRC}}$ in Section 4.1. The reduction is still polynomial since α was assumed to be constant and the inequality in Claim 5 still holds. As Claim 5 was the only step where we used assumption (A4), this proves Corollary 2.*

5. Conclusion

In this paper, we proved novel hardness results for pricing problems that arise when solving set partitioning formulations for the two-stage vehicle routing problem with stochastic demands. More precisely, we have shown that, when the input probability distribution is given by scenarios, for any choice of a recourse cost function that satisfies assumptions (A1)-(A4), the corresponding VRPSD pricing problem is strongly $\mathcal{NP}$-hard. Consequently, we show that, for any choice of a route relaxation, there is no hope of designing a pseudo-polynomial time algorithm for solving a VRPSD pricing problem, even if we allow the recourse cost to be inexact at non-elementary q-routes. Furthermore, based on the work of Spliet (2023), we have also shown that assuming an additional cost structure in the VRPSD pricing problem — where the edge costs have the form of reduced costs whose dual variables come from a restricted master problem — does not make the problem easier (at least from a computational complexity perspective).

We believe that several research directions can be further explored. For example, although many recourse policies have been proposed for the case where the customers have independent probability distributions, it is still not clear how one should design recourse policies for the case where the probability distribution is given by scenarios. It might be that in this case, assumption (A3) is reasonable but restrictive. Another research direction is to study the problem when additional structures are imposed on the input graph. For example, inspired by the VRPSD instances used for computational studies (Laporte et al., 2002; Jabali et al., 2014; Florio et al., 2022), we constructed our proof so that the input graph for the VRPSD is complete, and the edge costs and demand values are strictly positive. However, some VRPSD instances have even further structures in the input graph. For example, several VRPSD instances use Euclidean graphs (Florio et al., 2020; Christiansen & Lysgaard, 2007).

It could be interesting to analyze if, under such additional structures in the input graph, the VRPSD pricing problem (with scenarios) is still strongly $\mathcal{NP}$-hard. Finally, we mention that it is not clear whether it is possible to reduce "typical pricing problems" (where the edge costs are unrestricted) to "reduced cost pricing problems" (where the edge costs have the form of reduced costs). In other words, given a pricing problem with an unrestricted cost vector w, can we set up a corresponding feasible restricted master problem $\max\{c^T\lambda : A\lambda \leq b\}$, with dual vector π and $c \geq 0$, in a way that the reduced cost vector $c - A^T\pi$ matches w? If such a statement holds, then our proof can be simplified by just setting the edge costs appropriately. (Quite possibly, the proof of Spliet (2023) would also be simplified by such a result.)

Overall, our results indicate that if one wants to design an exact column-generation based algorithm for the VRPSD with scenarios, then one either has to cope with the strong $\mathcal{NP}$-hardness of the pricing problem (for example, like in Florio et al. (2022)); or has to resort to other methods to precisely compute the recourse cost. For instance, one could ignore the recourse cost in the objective function coefficient of a q-route and use the integer L-shaped method (Laporte et al., 2002; Jabali et al., 2014; Hoogendoorn & Spliet, 2023) to capture the recourse cost of a solution.

Acknowledgments

The authors acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) (funding reference number RGPIN-2020- 04030).

References

Baldacci, R., Mingozzi, A., & Roberti, R. (2011). New route relaxation and pricing strategies for the vehicle routing problem. *Operations research*, *59*, 1269–1283.

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Chen, R., & Luedtke, J. (2022). On sample average approximation for two-stage stochastic programs without relatively complete recourse. *Mathematical Programming*, *196*, 719–754.

Christiansen, C. H., & Lysgaard, J. (2007). A branch-and-price algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research Letters*, *35*, 773–781.

Cranston, D. W., & West, D. B. (2017). An introduction to the discharging method via graph coloring. *Discrete Mathematics*, *340*, 766–793.

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

Dror, M. (1990). Cost allocation: The traveling salesman, bin packing, and the knapsack. *Appl. Math. Comput.*, *35*, 191–207.

Florio, A. M., Gendreau, M., Hartl, R. F., Minner, S., & Vidal, T. (2022). Recent advances in vehicle routing with stochastic demands: Bayesian learning for correlated demands and elementary branch-price-and-cut. *European Journal of Operational Research*, .

Florio, A. M., Hartl, R. F., & Minner, S. (2020). New exact algorithm for the vehicle routing problem with stochastic demands. *Transportation Science*, *54*, 1073–1090.

Fukasawa, R., & Gunter, J. (2023). The complexity of branch-and-price algorithms for the capacitated vehicle routing problem with stochastic demands. *Operations Research Letters*, *51*, 11–16. URL: `https://www.sciencedirect.com/science/article/pii/S0167637722001468`. doi:`https://doi.org/10.1016/j.orl.2022.11.005`.

Gauvin, C., Desaulniers, G., & Gendreau, M. (2014). A branch-cut-and-price algorithm for the vehicle routing problem with stochastic demands. *Computers & Operations Research*, *50*, 141–153.

Gendreau, M., Jabali, O., & Rei, W. (2016). 50th anniversary invited article—future research directions in stochastic vehicle routing. *Transportation Science*, *50*, 1163–1173.

Ghosal, S., & Wiesemann, W. (2020). The distributionally robust chance-constrained vehicle routing problem. *Operations Research*, *68*, 716–732. URL: `https://doi.org/10.1287/opre.2019.1924`. doi:`10.1287/opre.2019.1924`. `arXiv:https://doi.org/10.1287/opre.2019.1924`.

Hoogendoorn, Y., & Spliet, R. (2023). An improved integer l-shaped method for the vehicle routing problem with stochastic demands. *INFORMS Journal on Computing*, *35*, 423–439.

Irnich, S., & Desaulniers, G. (2005). *Shortest path problems with resource constraints*. Springer.

Jabali, O., Rei, W., Gendreau, M., & Laporte, G. (2014). Partial-route inequalities for the multi-vehicle routing problem with stochastic demands. *Discrete Applied Mathematics*, *177*, 121–136.

Jepsen, M., Petersen, B., Spoorendonk, S., & Pisinger, D. (2008). Subset-row inequalities applied to the vehicle-routing problem with time windows. *Operations Research*, *56*, 497–511.

Laporte, G., Louveaux, F. V., & van Hamme, L. (2002). An integer l-shaped algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research*, *50*, 415–423.

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

Ledvina, K., Qin, H., Simchi-Levi, D., & Wei, Y. (2022). A new approach for vehicle routing with stochastic demand: Combining route assignment with process flexibility. *Operations Research*, *70*, 2655–2673. URL: `https://doi.org/10.1287/opre.2022.2304`. doi:`10.1287/opre.2022.2304`. `arXiv:https://doi.org/10.1287/opre.2022.2304`.

Luedtke, J., & Ahmed, S. (2008). A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, *19*, 674–699.

Pessoa, A. A., Poss, M., Sadykov, R., & Vanderbeck, F. (2021). Branch-cut-and-price for the robust capacitated vehicle routing problem with knapsack uncertainty. *Operations Research*, *69*, 739–754. URL: https://doi.org/10.1287/opre.2020.2035. doi:10.1287/opre.2020.2035. arXiv:https://doi.org/10.1287/opre.2020.2035.

Poggi, M., & Uchoa, E. (2014). New exact algorithms for the capacitated vehicle routing problem. In *Vehicle Routing* chapter 3. (pp. 59–86). doi:10.1137/1.9781611973594.ch3.

Salavati-Khoshghalb, M., Gendreau, M., Jabali, O., & Rei, W. (2019a). A hybrid recourse policy for the vehicle routing problem with stochastic demands. *EURO Journal on Transportation and Logistics*, *8*, 269–298.

Salavati-Khoshghalb, M., Gendreau, M., Jabali, O., & Rei, W. (2019b). A rule-based recourse for the vehicle routing problem with stochastic demands. *Transportation Science*, *53*, 1334–1353.

Spliet, R. (2023). Technical note—the complexity of the pricing problem of the set partitioning formulation of vehicle routing problems. *Operations Research*, . doi:10.1287/opre.2023.2481.

Swamy, C., & Shmoys, D. B. (2012). Sampling-based approximation algorithms for multistage stochastic optimization. *SIAM Journal on Computing*, *41*, 975–1004.

Tillman, F. A. (1969). The multiple terminal delivery problem with probabilistic demands. *Transportation Science*, *3*, 192–204.

Toth, P., & Vigo, D. (2014). *Vehicle routing: problems, methods, and applications*. SIAM.

Verweij, B., Ahmed, S., Kleywegt, A. J., Nemhauser, G., & Shapiro, A. (2003). The sample average approximation method applied to stochastic routing problems: a computational study. *Computational optimization and applications*, *24*, 289–333.

Yee, J. R., & Golden, B. L. (1980). A note on determining operating strategies for probabilistic vehicle routing. *Naval Research Logistics Quarterly*, *27*, 159–163.