The complexity of branch-and-price algorithms for the capacitated vehicle routing problem with stochastic demands

Ricardo Fukasawa
rfukasawa@uwaterloo.ca

Joshua Gunter
jdgunter@uwaterloo.ca

September 29, 2022

Abstract

The capacitated vehicle routing problem with stochastic demands (CVRPSD) is a variant of the deterministic capacitated vehicle routing problem where customer demands are random variables. While the most successful formulations for several deterministic vehicle-routing problem variants are based on a set-partitioning formulation, adapting such formulations for the CVRPSD under mild assumptions on the demands remains challenging. In this work we provide an explanation to such challenge, by proving that when demands are given as a finite set of scenarios, solving the LP relaxation of such formulation is strongly NP-Hard. We also prove a hardness result for the case of independent normal demands.

1 Introduction

The capacitated vehicle routing problem with stochastic demands (CVRPSD) is a variant of the deterministic capacitated vehicle routing problem (CVRP). The CVRP, which has been widely studied [Laporte, 1992, Toth and Vigo, 2001], consists of planning routes for vehicles with a given capacity to deliver goods to a set of customers with known demands, with the goal of finding the cheapest such set of routes. In the CVRPSD, rather than being deterministic, customer demands are random variables from a given probability distribution. This creates the possibility of route failures occurring, where the realized demand on a route is greater than the vehicle capacity.

Both chance-constrained and 2-stage stochastic programming models for the CVRPSD have been explored. In chance-constrained models, the additional costs incurred by route failures is not computed explicitly, and instead routes are required to have a low probability of failing in the first place. Stewart and Golden [1983] solved a chance-constrained model via a reduction to a deterministic VRP under the assumption of independent demands. They additionally provided heuristics for solving a chance-constrained model with correlated demands. Laporte et al. [1989] were the first to exactly solve a chance-constrained model using a branch-and-cut algorithm, but required demands to be both independent and normal. More recently, Dinh et al. [2018] exactly solved a chance-constrained model for both joint normal demands and demands given by scenarios with a branch-cut-and-price algorithm.

More work has been devoted to formulations of the CVRPSD as a 2-stage stochastic program with recourse. Bertsimas [1992] derived upper and lower bounds on the optimal value of the CVRPSD, while Gendreau et al. [1995] were the first to successfully solve the CVRPSD as a 2-stage stochastic program exactly. They achieved this by way of the Integer L-shaped algorithm, a branch-and-cut algorithm first developed by Laporte and Louveaux [1993] as a general framework for solving stochastic programs with complete recourse. Hjorring and Holt [1999] found new optimality cuts for this method for the case of a single vehicle and independent demands.

For other vehicle routing problems, such as the deterministic CVRP and the vehicle routing problem with time windows (VRPTW), the current state-of-the-art solving techniques are branch-and-price algorithms [Pessoa et al., 2020, Fukasawa et al., 2006]. The column generation subproblems for these are NP-hard, but due to pseudo-polynomial time algorithms for column generation these branch-and-price algorithms can still be solved efficiently. These algorithms are especially effective in the case of restrictive capacity constraints, a class of problems with which the L-shaped algorithm struggles.

The efficacy of branch-and-price algorithms in those cases motivated study into whether they could similarly be applied to the CVRPSD, with Christiansen and Lysgaard [2007] being the first to develop a branch-and-price algorithm for the CVRPSD using a set partitioning formulation. While this method was not able to solve problems containing as many vertices as the L-shaped method, their results showed it was more effective when problem instances had a higher number of vehicles. This method was further improved by Gauvin et al. [2014]. However, these algorithms rely on the assumption that customer demands are independent.

A core component of branch-and-price algorithms is the efficient solution of the column generation subproblem. We will refer to this subproblem as the minimum cost 2-stage q-route problem (2SQ). This consists of finding the q-route, or walk starting and ending at the depot, with minimum expected cost that satisfies capacity constraints. In this paper, we will examine results on this problem and the implications they have on solving this formulation for the CVRPSD.

Our main contribution is two proofs on the hardness of the minimum cost 2-stage q-route problem. First we prove that solving this problem in the case of distributions specified by a finite set of scenarios is strongly NP-hard, which indicates that this problem is very hard in general and can not be simply extended to instances with dependent customers demands. Secondly we prove that under the assumption of $O(1)$ elementary operations, there is a reduction from the Hamiltonian cycle problem to the minimum cost 2-stage q-route problem with independent normal demands, even when the means and variances are polynomially bounded in n. This suggests that exactly pricing the 2-stage cost of q-routes may not be feasible in pseudopolynomial time. However this does not constitute a proof of strong NP-hardness, due to some of the required assumptions and potentially large edge costs in the construction.

There are a few points to make to clarify the implications of these results. First of all, the results talk about the hardness of calculating the *exact* two-stage cost. For instance this still allows an approach like the one from Christiansen and Lysgaard [2007] that relaxes the fact that visiting a customer twice means that the demands of the repeated customer in the q-route is identical every time it happens in the q-route.

The second point is that state-of-the-art methods for deterministic routing problems use more complex structures like ng-routes [Baldacci et al., 2011]. Our hardness results actually are stronger due to the fact that we consider q-routes instead of ng-routes, since q-routes are a special case of ng-routes and, therefore, the existence of a (pseudo-)polynomial time algorithm for ng-routes would imply the same algorithm could be applied for q-routes.

Finally, our work deals with a simple recourse policy. On the other hand, there are several recent works that deal with a more complex way of calculating the failure cost [Florio et al., 2020b,a, 2021, 2022, Salavati-Khoshghalb et al., 2019a,b, Louveaux and Salazar-González, 2018]. Once again, considering a simpler recourse policy and showing a hardness result seems to point to the core issue that makes the problem hard is the uncertainty itself, rather than the more complicated recourse policy. Although, contrary to the ng-route case, we do not believe that our result directly implies hardness of the more complex recourse policy. It is worth pointing out that all such works assume the demands are independent, which we do not assume in the scenario case.

This paper is organized as follows: in Section 2, we cover a problem definition and formulation for the CVRPSD. In Section 3, we examine the differences between the pricing problem for the CVRP and CVRPSD and take a closer look at the difficulties introduced by stochastic demands. In Section 4, we present our results on the hardness of the pricing problem.

2 Problem Definition & Model Formulation

The CVRPSD is defined on an undirected graph $G = (V, E)$, where $V = \{0, ..., n\}$. Vertex 0 represents the depot, and $V_+ = \{1, ..., n\}$ represents the set of customers. For each edge $\{i, j\} \in E$ there is an associated travel cost d_{ij}.

Each customer $i \in V_+$ has an associated random variable ξ_i for their demand, with expected value $\mathbb{E}[\xi_i]$ and variance $V[\xi_i]$. There are t identical vehicles, each with a capacity denoted by Q. A q-route is defined as a walk $(0, v_1, v_2, ..., v_k, 0)$ where $v_i \in V_+$ for $i = 1, ..., k$ and $v_i \neq v_{i+1}$ for $i = 1, ..., k-1$. We will say the *length* of such a q-route is k. A q-route is an *elementary route* if $v_i \neq v_j$ for all $i \neq j$, otherwise it is *non-elementary*. Let μ denote the cumulative expected demand on a q-route R; R is then *feasible* if $\mu \leq Q$.

We consider the following set partitioning formulation (SP):

$$\min \sum_{r\in\mathfrak{R}_q} c_r x_r \tag{1}$$

$$\text{s.t.} \sum_{r\in\mathfrak{R}_q} \alpha_{ir} x_r = 1, \quad \forall i \in V_+, \tag{2}$$

$$\sum_{r\in\mathfrak{R}_q} x_r = t, \tag{3}$$

$$x_r \in \{0,1\} \quad \forall r \in \mathfrak{R}_q. \tag{4}$$

Here $\mathfrak{R}_q$ denotes the set of all feasible q-routes; c_r denotes the expected cost of r; and α_{ir} is a parameter which counts the number of times a route r visits vertex i. The objective function (1) minimizes the expected cost of the routes. Constraints (2) ensure that each customer is visited exactly once by some route. We use a simple recourse strategy: if a customer's demand exceeds the vehicle capacity they return to the depot to replenish, then return to the failed customer and continue the route as planned.

To compute the expected cost c_r for a route, we divide it into two quantities: C_D, the deterministic cost of travelling along the route, and C_F, the expected cost incurred by failures along the route, with $c_r = C_D + C_F$. C_D is equal to the sum of the edge costs along the route, while C_F is computed by summing the *expected failure cost* at each customer i along the route. Given a q-route $r = (0, 1, ..., k, 0)$, let $\Psi_{r,i}$ denote the probability distribution of the cumulative demand of r at vertex i, and let $\psi_{r,i}$ denote a $\Psi_{r,i}$-distributed random variable. The expected number of failures at customer i, $\mathrm{FAIL}_i(r)$, is then computed by the following:

$$\mathrm{FAIL}_i(r) = \sum_{u=1}^{\infty} [\mathbb{P}\{\psi_{r,i-1} \leq uQ\} - \mathbb{P}\{\psi_{r,i} \leq uQ\}] \tag{5}$$

where $\mathbb{P}\{E\}$ denotes the probability of event E occurring. Thus $\mathbb{P}\{\psi_{r,i-1} \leq uQ\}$ is the probability that the uth failure has not occurred before visiting customer i, and $\mathbb{P}\{\psi_{r,i} \leq uQ\}$ is the probability that the uth failure has still not occurred after visiting customer i. Therefore the difference:

$$\mathbb{P}\{\psi_{r,i-1} \leq uQ\} - \mathbb{P}\{\psi_{r,i} \leq uQ\} \tag{6}$$

is the probability of the uth failure occurring exactly at customer i. The expected failure cost at customer i $\mathrm{EFC}_i(r)$ can then by computed by multiplying this by the distance travelled when a failure occurs:

$$\mathrm{EFC}_i(r) = 2d_{0i}\,\mathrm{FAIL}_i(r). \tag{7}$$

Thus $C_F = \sum_{i=1}^{k} \mathrm{EFC}_i(r)$.

Finally, we refer to the column generation subproblem of finding the q-route with minimum reduced cost as the *minimum 2-stage cost q-route problem (2SQ)* to distinguish it from the deterministic version of the problem. Formally, we wish to solve:

$$\min_{r \in \mathfrak{R}_q} c_r \tag{8}$$

Where $\mathfrak{R}_q$ is the set of all feasible q-routes and $c_r = C_D + C_F$ as defined above. We note that an instance of (2SQ) is defined by the graph G, the edge costs d, the capacity Q and the probability distribution of the demands $\mathbb{P}$.

3 Complexities in pricing introduced by stochastic demands

For the deterministic CVRP, there exists a dynamic programming algorithm to price q-routes in $O(n^2Q)$ time. The pricing problem is expressed as a shortest path problem with resource constraints (SPPRC), an overview of which can be found in Righini and Salani [2006]. For the CVRPSD the problem becomes harder, and one of the main reasons is that in the case of q-routes the demands of vertices at different points along the route will be correlated if the route contains repeated vertices, even if the demands between different customers are independent. Suppose $r = (0, v_1, ..., v_k, 0)$ is a q-route such that $v_i = v_j$ for $i \neq j$, $1 \leq i, j, \leq k$. Then v_i and v_j have perfectly correlated demand.

One problem stemming from this core issue is that the cumulative variance of a q-route is not necessarily additive. If the route is elementary and the random variables are independent, then the cumulative variance is additive, but if the route is not elementary then the cumulative variance is also not additive. Consider having k copies of a random variable ξ with expected value $\mathbb{E}[\xi]$ and variance $V[\xi]$. By linearity of expectation, the expected value of $k\xi$ is $k\mathbb{E}[\xi]$. However variance is not linear. Indeed, it is well-known that $V[k\xi] = k^2V[\xi]$. Thus if a customer i is repeated k times in a route, it should contribute $k^2V[\xi_i]$ to the total variance of the route, rather than $kV[\xi_i]$. This means that given a path $P = (u_1, ..., u_h)$ and a vertex v, to calculate the cumulative variance of the augmented path $P' = (u_1, ..., u_h, v)$ we need to know how many times v has already been visited by P, not simply the cumulative variance of P and the variance of v. This is the core issue that differentiates pricing q-routes for the deterministic CVRP and the CVRPSD.

To illustrate this, we present the following example. Consider a CVRPSD instance consisting of three customer nodes $1, 2, 3$, each represented by independent normally distributed random variables with expected demand $\mu_i = 100$ and variance $\sigma_i^2 = 100$ for $i = 1, 2, 3$. Let the vehicle capacity be $Q = 305$. Consider the two routes $r^1 = (0, 1, 2, 3, 0)$ and $r^2 = (0, 1, 2, 1, 0)$.

The expected number of failures on the elementary route r^1 is $\mathrm{FAIL}_1(100, 100) + \mathrm{FAIL}_2(200, 200) + \mathrm{FAIL}_1(300, 300) = 0.386$, while the expected number of failures on the non-elementary route r^2 is $\mathrm{FAIL}_1(100, 100) + \mathrm{FAIL}_2(200, 200) + \mathrm{FAIL}_1(300, 600) = 0.419$. Thus the non-elementary route r^2 has a failure rate higher than that of r^1, even though each route visits a total of 3 vertices with each vertex having identical mean and variance.

This fact that we need to know information about the number of times a vertex has occurred so far on the route suggests that this problem may be similarly hard as the elementary route pricing problem, which is known to be strongly NP-Hard. It is worth noting that previous approaches to this problem such as by Christiansen and Lysgaard [2007] and Gauvin et al. [2014] considered two visits to the same customer as if they were two completely independent visits to different customers.

4 Complexity of solving the minimum 2-stage cost q-route problem

We now present our results on the complexity of solving the minimum cost 2-stage q-route problem. We first show that in the finite scenario case, pricing q-routes when including the 2-stage costs is strongly NP-hard. Secondly, we show that while it is not necessarily strongly NP-hard in the case of independent normal demands, under the assumption that elementary operations are $O(1)$ and that the expected number of route failures on a given route is computable in polynomial time, there is a reduction from the Hamiltonian cycle problem to the minimum cost 2-stage q-route problem. This holds even when demands and variances are polynomially bounded in n. Since both results rely on a reduction from the Hamiltonian cycle problem (a strongly NP-hard problem), we formally define it here:

Definition 1. *Given a graph $G' = (V', E')$, a* Hamiltonian cycle *is a cycle that visits every vertex exactly once, or equivalently a connected subgraph H of G' such that $V(H) = V'$ and $deg_H(v) = 2, \forall v \in V'$. The strongly NP-hard Hamiltonian cycle problem is to decide, given the graph G', if G' has a Hamiltonian cycle.*

Our strategy to prove hardness is to provide a polynomial-time reduction from the Hamiltonian cycle problem to the desired pricing problem.

4.1 Finite Scenarios

In the finite scenario case, we are solving (8) with the distribution of customer demands being given as a finite set of sample scenarios $s^j, 1 \leq j \leq S$. Each scenario s^j has a probability $p_j, 1 \leq j \leq S$ of being realized, with $\sum_{j=1}^{S} p_j = 1$. The customer demand vector ξ takes on the value s^j with probability p_j.

Theorem 1. *Given an instance $G' = (V', E')$ of the Hamiltonian cycle problem, with $n = |V'| - 1 \geq 3$. Then one can construct in time polynomial in n an instance of (2SQ) with the property that its optimal solution has value $2n^3 - n$ if and only if G' has a Hamiltonian cycle.*

Proof. Let $\{0\}$ be an arbitrary vertex in V' and $V_+ = V' \setminus \{0\}$. Consider the following instance of (2SQ):

Let $G = (V_+ \cup \{0\}, E)$ be the complete graph on the same vertex set as G', and assign edge costs as follows:

$$\begin{array}{ll} d_{0i} = n^3 & \forall 0i \in E' \\ d_{0i} = n^3 + 1 & \forall 0i \notin E' \\ d_{ij} = -1 & \forall ij \in E' \\ d_{ij} = 0 & \forall ij \notin E' \end{array}$$

Let the capacity $Q = 2n - 1$. Construct n scenarios s^v, one for each vertex $v \in V_+$ with vertex v having demand n and all other vertices having demand 1. We note that $\mathbb{E}[\xi_i] = \frac{2n-1}{n}, \forall i \in V_+$ and therefore, any q-route has length at most n.

Let $\tilde{c}_r$ denote the expected failure cost of a q-route, with c_r denoting the total cost of a q-route, i.e. the sum of the deterministic travel costs and $\tilde{c}_r$. We first show that elementary and non-elementary q-routes have a significant difference in expected failure cost.

Claim 1: Let r be a q-route of (2SQ). If r is elementary then $\tilde{c}_r = 0$. If r is non-elementary, then $\tilde{c}_r \geq n^2$.

Proof: If r is elementary, then the total demand on r in any scenario is at most $2n - 1$. Therefore r will never fail, and so $\tilde{c}_r = 0$. If r is non-elementary, then there exists some vertex $v \in V_+$ which is visited twice. Therefore r fails in scenario s^v, implying that $\tilde{c}_r \geq \frac{1}{n}n^3 = n^2$. ■

We now show that this implies that the optimum solution to (2SQ) is an elementary q-route.

Claim 2: The minimum cost q-route of (2SQ) is elementary.

Proof: For any vertex $v \in V_+$, the elementary route $(0, v, 0)$ has cost $\leq 2n^3+2$. Therefore the minimum cost q-route must have cost less than or equal to $2n^3+2$.

If r is a non-elementary route, then by the claim 1, we have $\tilde{c}_r \geq n^2$ and so we have $c_r \geq 2n^3 + \tilde{c}_r \geq 2n^3 + n^2$. As noted before, any q-route has length at most n. Thus any non-elementary route can have its cost reduced by at most n by travelling along negative cost edges, and so $c_r \geq 2n^3+n^2-n > 2n^3+2$. Note that the last inequality holds for all $n \geq 3$. Thus any non-elementary route r cannot be a minimum cost q-route. ■

We are now ready to prove the theorem. Let r be a minimum cost q-route of (2SQ). From Claim 2, it is an elementary route, and from Claim 1, we have $\tilde{c}_r = 0$. From the observation that r visits at most n vertices in V_+, it is easy to see that $c_r \geq 2n^3 - n$ with equality being achieved if and only if r is a Hamiltonian cycle contained in G' □

The hardness result for the finite scenario case now follows as a corollary.

Corollary 1. *When demands are given as a finite set of scenarios, solving the minimum cost 2-stage q-route problem is strongly NP-Hard.*

4.2 Independent Normal Demands

For the independent normal case we are again solving (8) with each customer demand $\xi_i = N(\mu_i, \sigma_i^2)$, i.e. the demand at each customer is independent and normally distributed with known means μ_i and variances σ_i^2, $1 \leq i \leq n$. Throughout, let $\Phi(x)$ be the CDF of the standard $N(0,1)$ normal distribution, and $\Phi^c(x) = 1 - \Phi(x)$.

Before proving a reduction from the Hamiltonian cycle problem to the minimum cost 2-stage q-route problem under the assumption of $O(1)$ elementary operations, we first need to define a new function RF and prove several necessary properties. This function counts the total expected number of failures on a route, as opposed to FAIL_i, which counts the number of failures at the vertex i on a route.

Definition 2 (Route Fail)**.** $\mathrm{RF}(\mu, \sigma^2)$ *denotes the expected number of failures on a route with a cumulative demand of μ and cumulative variance of σ^2, that is,* $\mathrm{RF}(\mu, \sigma^2) :=$

$$\sum_{u=1}^{\infty} u[\mathbb{P}\{\Psi(\mu,\sigma^2) > uQ\} - \mathbb{P}\{\Psi(\mu,\sigma^2) > (u+1)Q\}]$$

$$= \sum_{u=1}^{\infty} \mathbb{P}\{\Psi(\mu,\sigma^2) > uQ\} = \sum_{u=1}^{\infty} \Phi^c\left(\frac{uQ - \mu}{\sigma}\right)$$

By using this definition and some properties that will be proven later on in this section, we can show the following result:

Theorem 2. *Suppose one can solve (2SQ) problem under the following assumptions:*

1. *Demands are independent and identically distributed normal with mean μ and variance σ^2.*
2. *μ and σ are constant integers which do not grow in n.*
3. *The capacity is polynomially bounded in n.*
4. *All elementary operations can be performed in $O(1)$ time.*
5. $\mathrm{RF}(\mu, \sigma^2)$ *is computable in polynomial time.*

Then there exists an algorithm using polynomially many operations that solves the Hamiltonian cycle problem with polynomially many calls to this algorithm.

Proof. Let $G' = (V', E')$ be an instance of the Hamiltonian cycle problem, with $n = |V'| - 1 \geq 3$ (if $n < 3$ one can solve the Hamiltonian cycle problem in constant time without any auxiliary problems).

Construct the following instance of the (2SQ). Let $\{0\}$ be an arbitrary vertex in V' and $V_+ = V' \setminus \{0\}$. Let $G = (V_+ \cup \{0\}, E)$ be a complete graph. Let

$\mu_i = 1$ and $\sigma_i^2 = 1$ for all i, and let $Q = n + \frac{1}{2}$. The first thing that needs to be highlighted is that any q-route must have length at most n, as it would otherwise violate the capacity constraint.

The idea of the proof is to carefully pick our edge costs in such a way that the minimum cost route must only contain edges in E'. We will then derive an upper bound UB_H on the cost of a route which is a Hamiltonian cycle in G, along with lower bounds LB_S and LB_Q on the costs of routes which have length at most $n-1$ and routes which are non-elementary of length n. We will then show that $UB_H < LB_S$ and $UB_H < LB_Q$. If we have all these ingredients, then we can just call (2SQ) for the constructed instance and know that G' has a Hamiltonian cycle if and only if the optimal value is at most UB_H.

The choice of costs will be done in a manner similar to how they were chosen for the scenario case. The costs are defined as:

$$\begin{aligned}
d_{0i} &= M_n - 1 && \forall 0i \in E' \\
d_{0i} &= M_n && \forall 0i \notin E' \\
d_{ij} &= -W_n && \forall ij \in E' \\
d_{ij} &= nW_n && \forall ij \notin E'
\end{aligned}$$

where M_n and W_n are some appropriately chosen positive constants that will be defined later.

Notice that the function RF can be used to provide bounds on EFC in the following manner. The expected failure cost of a route $r = (0, 1, ..., k, 0)$ is the sum of the expected failure costs at each vertex, i.e. $\mathrm{EFC}(r) = \sum_{i=1}^{k} \mathrm{EFC}_i(\mu_i, \sigma_k^i)$, where μ_i and σ_i^2 denote the cumulative demand and variance at vertex i along the route R. Furthermore $\mathrm{EFC}_i(\mu_i, \sigma_i^2) = 2d_{0i}\,\mathrm{FAIL}_i(\mu_i, \sigma_i^2)$, where $\mathrm{FAIL}_i(\mu_i, \sigma_i^2)$ counts the expected number of failures at vertex i. Since RF counts the total expected number of failures on the route, we have $\mathrm{RF}(\mu_k, \sigma_k^2) = \sum_{i=1}^{k} \mathrm{FAIL}_i(\mu_i, \sigma_i^2)$. Letting $d^+ = \max\{d_{0i} : 1 \le i \le k\}$ and $d^- = \min\{d_{0i} : 1 \le i \le k\}$, we obtain the following bounds for $\mathrm{EFC}(r)$ using RF:

$$\begin{aligned}
\sum_{i=1}^{k} 2d_{0i}\,\mathrm{FAIL}_i(\mu_i, \sigma_i^2) &\le \sum_{i=1}^{k} 2d^+\,\mathrm{FAIL}_i(\mu_i, \sigma_i^2) \\
&= 2d^+\mathrm{RF}(\mu_k, \sigma_k^2)
\end{aligned}$$

By a similar argument, one can argue that $2d^-\mathrm{RF}(\mu_k, \sigma_k^2) \le \mathrm{EFC}(r) \le 2d^+\mathrm{RF}(\mu_k, \sigma_k^2)$.

Minimum cost q-route only contains edges in E':

Let us now argue why a minimum cost q-route will only contain edges in E'. Let v be a vertex such that $0v \in E'$, and consider the q-route $r_v = (0, v, 0)$. This route has cost:

$$c_{r_v} = 2(M_n - 1) + 2(M_n - 1)\mathrm{RF}(1, 1).$$

Now consider a q-route $r_u = (0, u_1, u_2, ..., u_k, 0)$ such that $u_i u_{i+1} \notin E'$ for some $1 \le i \le k-1$. Since the length of r_u is at most n, the cost of r_u is:

$$c_{r_u} = d_{0u_1} + d_{0u_k} + \sum_{i=1}^{k-1} d_{u_i u_{i+1}} + \mathrm{EFC}(r_u)$$

$$\geq 2(M_n - 1) + (n - (k-2))W_n + 2(M_n - 1)\mathrm{RF}(k, k)$$
$$> 2(M_n - 1) + 2(M_n - 1)\mathrm{RF}(1, 1) = c_{r_v}.$$

We note that the last inequality follows from the fact that $\mathrm{RF}(\mu, \sigma^2)$ is strictly increasing in both μ and σ, as will be shown later in Lemma 2.

Upper bounds on Hamiltonian cycles of G'

Next we derive an upper bound on the cost of a q-route which is a Hamiltonian cycle in G'. The cost of a q-route r_H which is a Hamiltonian cycle in G' is:

$$c(r_H) = \underbrace{2(M_n - 1) - (n-1)W_n}_{\text{deterministic travel cost}} + \underbrace{\mathrm{EFC}(R_H)}_{\text{recourse cost}}$$

As the maximum cost of any edge $0i$, $1 \leq i \leq n$, is M_n, we have $2M_n\mathrm{RF}(n, n) \geq \mathrm{EFC}(R_H)$. Thus:

$$UB_H = 2(M_n - 1) - (n-1)W_n + 2M_n\mathrm{RF}(n, n)$$

is an upper bound for $c(r_H)$.

Lower bounds on q-routes of length at most $n - 1$

Next we will obtain a lower bound LB_S on the cost of a q-route with length at most $n - 1$.

Let $R_{E_k} = (0, v_1, v_2, ..., v_k, 0)$, $1 \leq k < n$, be a q-route. The cost of R_{E_k} is:

$$c(R_{E_k}) = d_{0v_1} + d_{0v_k} + \sum_{i=1}^{k-1} d_{v_i v_{i+1}} + \mathrm{EFC}(R_{E_1})$$

$$\geq 2(M_n - 1) - (k-1)W_n + 2(M_n - 1)\mathrm{RF}(k, k) =: LB_{E_k}$$

We will use $LB_S := LB_{E_{n-1}}$. To be able to bound this cost, we will make the following assumption on W_n:

(AW) $W_n > 2M_n[\mathrm{RF}(k+1, k+1) - \mathrm{RF}(k, k)] + 2 \cdot \mathrm{RF}(k, k), \forall k = 1, \ldots, n-1$

Claim 1: If $k < n - 1$, then $LB_{E_k} \geq LB_{E_{n-1}} > UB_H$
Proof:

Note that $LB_{E_{k+1}} - LB_{E_k} = 2(M_n - 1)[\mathrm{RF}(k+1, k+1) - \mathrm{RF}(k, k)] - W_n \leq 2M_n[\mathrm{RF}(k+1, k+1) - \mathrm{RF}(k, k)] + 2 \cdot \mathrm{RF}(k, k) - W_n < 0$ where the last inequality follows from (AW).

Therefore, $LB_{E_k} \geq LB_{E_{n-1}}$.

Finally note that it also follows from (AW) that

$$UB_H - LB_{E_{n-1}} =$$
$$= 2M_n[\mathrm{RF}(n, n) - \mathrm{RF}(n-1, n-1)] +$$
$$+ 2 \cdot \mathrm{RF}(n-1, n-1) - W_n < 0$$

Therefore $UB_H < LB_{E_{n-1}}$. ■

Lower bounds on non-elementary q-routes of length n

Next we consider a non-elementary route of length n, $R_Q = (0, v_1, v_2, ..., v_n, 0)$. Also, notice that the expected demand of R_Q is n and, since at least one vertex repeats in R_Q, its variance is strictly more than n. It will suffice for us to use the lower bound $n+1$ on the variance (though that lower bound can be improved).

Like in the previous case,

$$c(R_Q) = d_{0v_1} + d_{0v_n} + \sum_{i=1}^{n-1} d_{v_i v_{i+1}} + \mathrm{EFC}(R_Q)$$

$$\geq 2(M_n - 1) - (n-1)W_n + 2(M_n - 1)\mathrm{RF}(n, n+1) =: LB_Q$$

To bound this cost, we will make the following assumption on M_n:

(AM) $M_n > \frac{\mathrm{RF}(n,n+1)}{\mathrm{RF}(n,n+1)-\mathrm{RF}(n,n)}$

With this assumption, it is easy to see that

$$UB_H - LB_Q$$

$$= 2 \cdot \mathrm{RF}(n, n+1) - 2M_n[\mathrm{RF}(n, n+1) - \mathrm{RF}(n, n)]$$
$$< 2 \cdot \mathrm{RF}(n, n+1) - 2\mathrm{RF}(n, n+1) = 0.$$

Therefore $UB_H < LB_Q$.

We can now conclude the proof of the Theorem, since if $\mathrm{RF}(\mu, \sigma^2)$ can be computed in polynomial time for all integers (μ, σ^2), then one can easily first pick M_n satisfying (AM), and then pick W_n satisfying (AW) also in polynomial time. Thus one can solve the Hamiltonian cycle problem in G' by solving the minimum (2SQ) in G and checking whether the output solution has cost $\leq UB_H$. $\square$

There were some missing details that needed to be proven before using Theorem 2. First of all, it is actually not entirely clear that RF is actually a finite quantity, given that it consists of an infinite sum. It is reasonable to expect that it is, but a formal proof is required. Thus, we will now prove that RF is convergent.

Lemma 1. $\mathrm{RF}(\mu, \sigma^2)$ *converges.*

Proof. We first note that $\Phi^c(t) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt < \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{t}{x} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \frac{1}{t} e^{-\frac{t^2}{2}}$. Therefore

$$\mathrm{RF}(\mu, \sigma^2) \sum_{u=1}^{\infty} \Phi^c \left(\frac{uQ - \mu}{\sigma} \right) <$$

$$\sum_{u=1}^{\infty} \frac{\sigma}{\sqrt{2\pi}(uQ - \mu)} e^{-\frac{(uQ-\mu)^2}{2\sigma^2}}$$

which converges by applying the ratio test:

$$\lim_{u\to\infty}\left|\frac{\frac{\sigma}{\sqrt{2\pi}((u+1)Q-\mu)}e^{-\frac{((u+1)Q-\mu)^2}{2\sigma^2}}}{\frac{\sigma}{\sqrt{2\pi}(uQ-\mu)}e^{-\frac{(uQ-\mu)^2}{2\sigma^2}}}\right|$$

$$=\lim_{u\to\infty}\left|\frac{uQ-\mu}{uQ+Q-\mu}e^{-\frac{2uQ^2+Q^2-2Q\mu}{2\sigma^2}}\right|=0$$

Therefore $\mathrm{RF}(\mu,\sigma^2)$ converges. $\square$

Next we will prove another fact that is used throughout the proof of Theorem 2, which is that RF is strictly increasing in μ, and strictly increasing in σ^2 when $\mu \leq Q$.

Lemma 2. $\mathrm{RF}(\mu,\sigma^2)$ *is strictly increasing in* μ*, and strictly increasing in* σ^2 *when* $\mu \leq Q$*.*

Proof. First we will show RF is strictly increasing in μ. Let $\epsilon > 0$. Then $\mathrm{RF}(\mu+\epsilon,\sigma^2)-\mathrm{RF}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\sum_{u=1}^{\infty}\int_{\frac{uQ-\mu-\epsilon}{\sigma}}^{\frac{uQ-\mu}{\sigma}} e^{-\frac{t^2}{2}}dt$. Moreover, for all $u \in \mathbb{Z}, u \geq 1$ we have $\frac{uQ-\mu}{\sigma} > \frac{uQ-\mu-\epsilon}{\sigma}$. Thus $\mathrm{RF}(\mu+\epsilon,\sigma^2) > \mathrm{RF}(\mu,\sigma^2)$.

Now suppose that $\mu \leq Q$, and again let $\epsilon > 0$. Then $\mathrm{RF}(\mu,\sigma^2+\epsilon) - \mathrm{RF}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\sum_{u=1}^{\infty}\int_{\frac{uQ-\mu}{\sqrt{\sigma^2+\epsilon}}}^{\frac{uQ-\mu}{\sigma}} e^{-\frac{t^2}{2}}dt$. Since $\mu \leq Q$, for $u=1$ we have $\frac{uQ-\mu}{\sigma} \geq \frac{uQ-\mu}{\sqrt{\sigma^2+\epsilon}}$ and for $u \geq 2, u \in \mathbb{Z}$, we have $\frac{uQ-\mu}{\sigma} > \frac{uQ-\mu}{\sqrt{\sigma^2+\epsilon}}$. Thus $\mathrm{RF}(\mu,\sigma^2+\epsilon) > \mathrm{RF}(\mu,\sigma^2)$. $\square$

We conclude this section by commenting that Theorem 2 does not show strong NP-hardness of the pricing problem, since some of the data used in the reduction may be very large and/or hard to compute. However, we believe this is an indication that the pricing problem is hard. Notice, for instance, that the pricing problem proposed in Christiansen and Lysgaard [2007] would be polynomial-time when μ, σ, Q are polynomially bounded. Also, putting in perspective, weakly NP-hard problems like the knapsack problem (which is closely related to the q-route problem) can be solved in polynomial-time if either the weights of the items or their profits are polynomially bounded. This does not happen in our problem.

5 Conclusion

In this paper we identified some difficulties that stochastic demands introduce to the pricing problem in set partitioning formulations for the CVRPSD. Specifically, we show that even when customer demands are independent, vertices along q-routes may still have correlated demands. One issue that this introduces is the cumulative variance of a q-route may not be the sum of the variances of all the customers on the q-route. This in turn means that we cannot as easily reuse the dynamic programming algorithm for pricing q-routes in the CVRP if we wish to *exactly* price q-routes in the CVRPSD.

In addition, we proved that when the distribution is specified as a finite set of scenarios, solving the minimum 2-stage cost q-route pricing problem is strongly NP-hard. We have also given a hardness result in the case of independent normal demands, even when demands and variances are bounded by a constant. Therefore if any pseudo-polynomial time dynamic programming algorithm exists, it would need to exploit bounded edge costs, rather than bounded demands and variances.

Further questions on this topic would first be whether the minimum cost 2-stage q-route pricing problem is strongly NP-hard in the case of independent normal demands. While our results suggest this may be case, further work needs to be done for a complete proof. Secondly, given that pricing is strongly NP-hard in the scenarios case, other avenues must be explored to solve this class of problems. One possibility is the use of Monte Carlo methods to verify route feasibility, an idea very recently applied to the vehicle routing with stochastic demands and probabilistic duration constraints by Florio et al. [2021]. Another possibility worth exploring is combining the set-partitioning formulation for the CVRPSD with another formulation such as in Dinh et al. [2018], or to compute approximate costs instead of exact ones.

Acknowledgments

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

References

R. Baldacci, A. Mingozzi, and R. Roberti. New route relaxation and pricing strategies for the vehicle routing problem. *Operations Research*, 59(5):1269–1283, 2011.

D. J. Bertsimas. A vehicle routing problem with stochastic demand. *Operations Research*, 40:574–585, 1992.

C. H. Christiansen and J. Lysgaard. A branch-cut-and-price algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research Letters*, 37:773–781, 2007.

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

A. M. Florio, R. F. Hartl, and S. Minne. Optimal a priori tour and restocking policy for the single-vehicle routing problem with stochastic demands. *European Journal on Operations Research*, 285:172–182, 2020a.

A. M. Florio, R. F. Hartl, and S. Minner. New exact algorithm for the vehicle routing problem with stochastic demands. *Transportation Science*, 54(4): 1073–1090, 2020b.

A. M. Florio, R. F. Hartl, S. Minner, and J.-J. Salazar-González. A branch-and-price algorithm for the vehicle routing problem with stochastic demands and probabilistic duration constraints. *Transportation Science*, 55(1):122–138, 2021.

A. M. Florio, D. Feillet, M. Poggi, and T. Vidal. Vehicle routing with stochastic demands and partial reoptimization. *Transportation Science*, Article in advance, 2022.

R. Fukasawa, H. Longo, J. Lysgaard, M. P. de Aragão, M. Reis, E. Uchoa, and R. F. Werneck. Robust branch-and-cut-and-price for the capacitated vehicle routing problem. *Mathematical Programming*, 108:491–511, 2006.

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

M. Gendreau, G. Laport, and R. Séguin. An exact algorithm for the vehicle routing problem with stochastic demands and customers. *Transportation Science*, 29:143–155, 1995.

C. Hjorring and J. Holt. New optimality cuts for a single-vehicle stochastic routing problem. *Annals of Operations Research*, 86:569–584, 1999.

G. Laporte. The vehicle routing problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, 59:345–358, 1992.

G. Laporte and F. Louveaux. The Integer L-shaped method for stochastic integer programs with complete recourse. *Operation Research Letters*, 13: 133–142, 1993.

G. Laporte, F. Louveaux, and H. Mercure. Models and exact solutions for a class of stochastic location-routing problems. *European Journal of Operational Research*, 39:71–78, 1989.

F. V. Louveaux and J.-J. Salazar-González. Exact approach for the vehicle routing problem with stochasticdemands and preventive returns. *Transportation Science*, 52(6):1463—-1478, 2018.

A. Pessoa, R. Sadykov, E. Uchoa, and F. Vanderbeck. A generic exact solver for vehicle routing and related problems. *Mathematical Programming B*, 183: 483–523, 2020.

G. Righini and M. Salani. Symmetry helps: Bounded bi-directional dynamic programming for the elementary shortest path problem with resource constraints. *Discrete Optimization*, 3:255–273, 2006.

M. Salavati-Khoshghalb, M. Gendreau, O. Jabali, and W. Rei. An exact algorithm to solve the vehicle routing problem with stochastic demands under an optimal restocking policy. *European Journal on Operations Research*, 273: 175–189, 2019a.

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

W. R. Stewart and B. L. Golden. Stochastic vehicle routing: a comprehensive approach. *European Journal of Operational Research*, 14:371–385, 1983.

P. Toth and D. Vigo, editors. *The Vehicle Routing Problem.* Society for Industrial and Applied Mathematics, 2001.