THE QUADRATIC CAPACITATED VEHICLE ROUTING PROBLEM

Rafael Martinelli
École Polytechnique de Montréal and GERAD
3000 chemin de la Côte-Ste-Catherine, Montréal (Québec), Canada H3T 2A7
rafael.martinelli@gerad.ca
Claudio Contardo
ESG UQÀM and GERAD
315 rue Ste-Catherine Est, Montréal (Québec), Canada H2X 3X2
claudio.contardo@gerad.ca

RESUMO

Neste artigo é introduzido o Problema de Roteamento de Veículos Capacitados Quadrático (PRVCQ), um problem combinatório o qual tem sua origem em aplicações nas áreas de logística e transporte. O PRVCQ estende dois problemas bem conhecidos, o Problema de Roteamento de Veículos Capacitados (PRVC) e o Problema Caixeiro-Viajante Quadrático (PCVQ). A primeira extensão é feita considerando uma matriz de custos modificada, na qual os custos de travessia passam a ser associados a pares de arestas consecutivas. A segunda é obtida a partir da introdução de demandas para os clientes e de veículos com capacidade. Uma formulação de fluxo de veículos com variáveis de três índices é apresentada para o problema, na qual as variáveis representam q-arestas (pares de arestas consecutivas). Esta formulação é então reforçada com algumas classes de desigualdades válidas. As rotinas de separação das desigualdades são apresentadas, as quais são então utilizadas em um procedimento de *branch-and-cut*. Experimentos computacionais são coduzidos com o objetivo de demonstrar a eficiência da modelagem e das rotinas utilizadas, fornecendo limites inferiores justos e soluções ótimas para instâncias de pequeno e médio porte.
PALAVRAS CHAVE. Problema de Roteamento de Veículos Quadrático, Problema Caixeiro-Viajante Quadrático, Branch-and-Cut.

Áreas Principais: Logística e Transportes, Otimização Combinatória e Programação Matemática.

ABSTRACT

In this article we introduce the Quadratic Capacitated Vehicle Routing Problem (QCVRP), a combinatorial optimization problem that arises in practical applications in logistics and transportation. The QCVRP extends two other known problems, the Capacitated Vehicle Routing Problem (CVRP) and the Quadratic Symmetric Traveling Salesman Problem (QSTSP), the former by considering a modified cost matrix in which traveling costs are now associated to pairs of two consecutive edges, and the latter by introducing customer demands and vehicle capacities. We present a three-index vehicle-flow formulation for the problem in which variables represent q-edges (pairs of consecutive edges), and strengthen it with some classes of valid inequalities. We present efficient separation routines for the inequalities used in this paper and derive an exact solver based on the branch-and-cut paradigm. We conduct computational experiments to show the efficiency of the modeling and solution approaches by providing tight lower bounds and optimal solutions for small- and medium-size instances.
KEY WORDS. Quadratic Capacitated Vehicle Routing, Quadratic Symmetric Traveling Salesman, Branch-and-Cut.

Main areas: Logistics and Transportation, Combinatorial Optimization and Mathematical Programming.

1. Introduction

This paper introduces the Quadratic Capacitated Vehicle Routing Problem (QCVRP), an extension of two important problems in combinatorial optimization and logistics: the Capacitated Vehicle Routing Problem (CVRP) and the Quadratic Symmetric Traveling Salesman Problem (QSTSP). In the QCVRP, a fleet of K homogeneous vehicles must be routed from a single depot to visit exactly once all customers from a predefined set of customers $V^+ = \{1,\ldots,n\}$. For notational convenience, the depot node is labeled as the node 0, and we denote $V = V^+ \cup \{0\}$. With each vehicle is associated a capacity Q, and with each customer $i \in V^+$ is associated a demand $d_i \in \mathbb{Z}^+$. The main difference of the QCVRP with respect to the traditional CVRP is the definition of the edge costs. They are not defined on every edge but rather on every pair of two consecutive edges. In the QSTSP, the cost function follows this last structure but now a single vehicle of infinite capacity is used to visit all customer nodes, instead of a limited fleet of K capacitated vehicles as for the QCVRP. An important difference between the problem introduced in this paper with respect to the QSTSP is the cost associated to edges adjacent to the depot. While in the QSTSP the costs incurred by two edges $\{0,i\}$ and $\{0, j\}$ depends on the sequence $i \to 0 \to j$ (or equivalently $j \to 0 \to i$ since the costs are symmetric), in the QCVRP we consider these two routing costs to be independent. From a strict modeling viewpoint, we can say that the QCVRP generalizes the original CVRP but only extends the QSTSP, as the latter cannot be modeled as a QCVRP in the strict sense.

The QCVRP is motivated by a real-life application, namely a very large-scale vehicle routing problem, as follows. Imagine that a very large number of n customers has to be visited using K vehicles, with $K \ll n$, and that it is untractable to handle this large-scale problem with standard state-of-the-art algorithms due to its size. Instead, customers are grouped into a much smaller number of clusters, each of which contains an accumulated demand lower than Q. Now, a planner must decide how to route these K vehicles so as to visit each cluster exactly once, at minimum total cost. The traveling cost incurred in visiting all nodes inside a cluster depends on the previous and following clusters on a vehicle route, as one may assume that the vehicle doing this route will travel from one cluster to the next by connecting the closest pair of nodes belonging to these two clusters. The traveling cost inside the intermediate cluster corresponds then to solving an open traveling salesman problem, in which the starting and ending nodes are chosen using the previously mentioned policy, and so the quadratic costs structure.

Another relevant application of the QCVRP is the solution of vehicle routing problems with turn penalties. In such setting, a set of vehicles must visit a set of predetermined customers (in a node-routing context) or edges (in an arc-routing context), subject to some additional constraints. Each node in the graph represents a corner in the city, and therefore some turns (normally left turns), are penalized or sometimes simply forbidden, so as to mimic some usual traffic regulations in urban transportation. In [6], the authors introduced the mixed capacitated general vehicle routing problem (MCGVRP), a node-and-arc routing problem that considers turn penalties. The authors propose a graph transformation of the MCGVRP to model it as an asymmetric CVRP (ACVRP), and use a memetic algorithm especially tailored for the ACVRP to solve it. The MCGVRP extends the mixed capacitated arc-routing problem (MCARP) that includes applications in snow-plowing [20] and waste-collection [4], in which left turns are forbidden.

The main contributions of this article are mainly two. First, we formally introduce the QCVRP by providing a strict mathematical formulation of the problem based on q-edges. Second, we develop an exact algorithm for the QCVRP based on the branch-and-cut paradigm, and show that the proposed modeling approach and solution method are useful to derive strong lower and upper bounds of the problem in short computing times.

The remainder of the article is structured as follows. In Section 2 we present a literature review of related problems and methodologies. In Section 3 we formalize the QCVRP by providing a strict mathematical formulation based on q-edges. In Section 4 we present some classes of valid inequalities for the problem. In Section 5 we present the different separation algorithms used to find violated inequalities. In Section 6 we present the exact algorithm based on the branch-and-cut paradigm to derive lower and upper bounds in an iterative manner. In Section 7 we provide a computational study on several classes of instances to show the effectiveness of the proposed approach. Finally, Section 8 concludes the article and provides a discussion of several potential avenues for future research.

2. Literature review

At the best of our knowledge, the QCVRP has not yet been considered in the literature. However, it is closely related to some other classes of combinatorial optimization problems.

The Symmetric Traveling Salesman Problem (STSP) is one of the most classical problems in combinatorial optimization. In the STSP, a traveling salesman (or equivalently, a vehicle with infinite capacity) must visit a set of vertices and get back to the origin vertex in the minimum possible time. The traveling distances between each pair of vertices are supposed to be symmetric, this is $c_{ij} = c_{ji}$ for every pair of vertices i and j. The STSP was introduced in the seminal work of [10] in which the authors present a compact two-index formulation containing an exponential number of constraints. This integer program is solved using a branch-and-cut method, a new technique at that time. State-of-the-art algorithms for the STSP are based on the work of [10] by considering several new classes of valid inequalities and scalable separation algorithms [19, 16, 2].

The Quadratic Symmetric Traveling Salesman Problem (QSTSP) is a natural extension of the STSP in which the distances depend of every pair of two consecutive edges (namely q-edges). The QSTSP was formally introduced by [13]. The authors introduce a three-index vehicle-flow formulation of the problem which they strengthen using several classes of valid inequalities. They perform a polyhedral study and show that several classes of valid inequalities induce facets of the QSTSP polytope. Practical applications of the QSTSP include the Angle STSP (A-STSP), a variation of the STSP in which the traveling times on two consecutive edges depend on their angle at the middle vertex [1].

The Capacitated Vehicle Routing Problem (CVRP) is the most classical variation of the STSP, in which a fleet of K identical vehicles (instead of a single vehicle as for the STSP) is used to visit exactly once each node from a predefined set of customer nodes. Every vehicle must start and end its route at a depot node, and cannot exceed its capacity while collecting the demands of the customers visited through the route. The CVRP was introduced by [11] who formally stated the problem. In [17], the authors proposed a compact two-index vehicle-flow formulation for the CVRP strengthened with capacity cuts, and developed the first exact algorithm for the problem based on the branch-and-cut paradigm. The CVRP shares with the STSP several structural properties but they also differ from an algorithmic point of view. Indeed, the capacities on vehicles induce an additional computational complexity that makes it much harder to solve in practice than the STSP. For the reader to have an idea, while the most efficient algorithm for the STSP [2] can solve problems with several thousands of customers to optimality, the most efficient exact algorithms for the CVRP [14, 3, 8] cannot solve problems containing more than 200 customers.

3. Mathematical formulation

Before formulating the QCVRP, let us define some notation. Let $V = \{0, 1, \ldots, n\}$ be the set of nodes, and let node 0 be the depot node. Let $V^+ = V \setminus \{0\}$ be the set of customer nodes. With every customer $i \in V^+$ we associate a demand $d_i \in \mathbb{Z}^+$. We are given a fleet of exactly K homogeneous vehicles, each of which has a capacity of Q units of demand. Let E be the set of edges, namely $E = \{\{i, j\} : i, j \in V, i < j\}$, and let E^q be the set of q-edges, namely $E^q = \{(\{i,k\}, j) : i, j, k \in V, i < k, i \neq j, j \neq k, j \neq 0\}$. With every q-edge $e = (\{i,k\}, j) \in E^q$ is associated a traveling cost c_{ijk}. Note that in the definitions of sets E and E^q the symmetry of the network (and of the routing costs) is implicit. Indeed, if $\{i, j\}$ and $(\{i,k\}, j)$ encode edges and q-edges in E and E^q, respectively, then $\{j,i\}$ and $(\{k,i\},j)$ encode exactly the same two objects. For the single-customer trips represented by routes of the form $0 \to i \to 0$ with $i \in V^+$, we let c_{0j} be the routing cost associated. For every customer set $S \subseteq V^+$ we define $r(S) = \lceil \sum_{i\in S} d_i/Q \rceil$, which is a lower bound on the number of vehicles needed to visit the customers in S due to the capacity constraints.

Now, let us define the variables of the model. For every $i \in V^+$ we let w_i be a binary variable equal to 1 iff customer i is served using a single-customer route. For every edge $\{i,j\} \in E^+$, with $E^+ = \{e = \{i,j\} \in E : i,j \in V^+, i < j\}$, we let z_{ij} be a binary variable equal to 1 iff customers i and j are the only two customers visited in a route. For every $e = (\{i,k\},j) \in E^q$ we let y_{ijk} be a binary variable equal to 1 iff the q-edge e is used in a multiple-customer route (this is, a route visiting at least three customers). In addition, we let for every edge $e = \{i,j\} \in E$, x_{ij} be a binary variable equal to 1 iff edge $\{i,j\}$ is used in a multiple-customer route visiting three or more customers. Variables x are not strictly necessary as they can be derived from variables y but are still included in the model. Finally, we define binary variables ξ_{ij} for every edge $\{i,j\} \in E$ equal to 1 iff the edge $\{i,j\}$ is used by any vehicle regardless of the number of customers served in its route. Again, these edge variables are not strictly necessary as they can be derived from the previous ones, but will help in the presentation of the article as they will allow linking some of our results with previous results for the CVRP. Indeed, variables ξ correspond to the usual two-index vehicle-flow variables of the CVRP used, for instance, in [17, 18]. Note that in order to be consequent with the notation and the network symmetry, we may use variables z_{ji}, ξ_{ji}, x_{ji} and y_{kji} to encode the exact same variables as z_{ij}, ξ_{ij}, x_{ij} and y_{ijk}, respectively.

Now, let us define the following additional notation. For every edge subset $F \subseteq E$ we let $x(F) \overset{\text{def}}{=} \sum_{e\in F} x_e$ and $\xi(F) \overset{\text{def}}{=} \sum_{e\in F} \xi_e$. Analogously, for every q-edge subset $F \subseteq E^q$ we define $y(F) \overset{\text{def}}{=} \sum_{e\in F} y_e$. For every $F \subseteq E^+$ we let $z(F) \overset{\text{def}}{=} \sum_{\{i,j\}\in F} z_{ij}$. For every customer subset $U \subseteq V^+$ we let $w(U) \overset{\text{def}}{=} \sum_{i\in U} w_i$. For any two vertices sets $U,T \subset V$, we define $(U:T) \overset{\text{def}}{=} \{e = \{i,j\} \in E : (i \in U \wedge j \in T) \vee (i \in T \wedge j \in U)\}$. Also, for any three subsets $U,T,W \subset V$ we define $(U:T:W) \overset{\text{def}}{=} \{e = (\{i,k\},j) \in E^q : (i \in U \wedge k \in W) \vee (i \in W \wedge k \in U), j \in T\}$. Now, for every vertex subset $S \subseteq V$ we let $E(S) \overset{\text{def}}{=} (S:S)$ and $\delta(S) \overset{\text{def}}{=} (S:V\setminus S)$. We also define, for every vertex subset $S \subseteq V$, $E^q(S) \overset{\text{def}}{=} (S:S:S)$ in addition to the quantities $\delta^{ioo}(S) \overset{\text{def}}{=} (S:V\setminus S:V\setminus S)$, $\delta^{ioi}(S) \overset{\text{def}}{=} (S:V\setminus S:S)$, $\delta^{iio}(S) \overset{\text{def}}{=} (S:S:V\setminus S)$ and $\delta^{oio}(S) \overset{\text{def}}{=} (V\setminus S:S:V\setminus S)$.

In addition, if $S \subseteq V^+$, we let the sets $\delta^+(S)$, $\delta^{ioo+}(S)$, $\delta^{ioi+}(S)$, $\delta^{iio+}(S)$, $\delta^{oio+}(S)$ be defined as before but without including edges or q-edges linking S to the depot. Note that if any of the sets involved in $(U:T)$ or $(U:T:W)$ is a singleton $\{i\}$, we will denote instead i.

The QCVRP can be formulated as the following integer linear program:

$$\min \quad \sum_{i\in V^+} c_{0i}w_i + \sum_{\{i,j\}\in E^+} (c_{0ij}+c_{0ji})z_{ij} + \sum_{(\{i,k\},j)\in E^q} c_{ijk}y_{ijk} \tag{1}$$

subject to

$$\xi(\delta(\{i\})) = 2 \qquad i \in V^+ \tag{2}$$

$$\xi(\delta(\{0\})) = 2K \tag{3}$$

$$\xi(\delta(S)) \geq 2r(S) \qquad S \subseteq V^+, 2 \leq |S| \leq |V^+|-1 \tag{4}$$

$$y_{0ij} + y_{0ji} + z_{ij} \leq 1 \qquad i,j \in V^+, i<j \tag{5}$$

$$x_{0i} = y(0:i:V^+\setminus\{i\}) \qquad i \in V^+ \tag{6}$$

$$x_{ij} = y(i:j:V\setminus\{i,j\}) = y(V\setminus\{i,j\}:i:j) \qquad i,j \in V^+, i<j \tag{7}$$

$$\xi_{ij} = x_{ij} + z_{ij} \qquad \{i.j\} \in E^+ \tag{8}$$

$$\xi_{0i} = x_{0i} + 2w_i + z(\delta^+(\{i\})) \qquad i \in V^+ \tag{9}$$

$$w_j \in \{0,1\} \qquad j \in V^+ \tag{10}$$
$$z_{ij} \in \{0,1\} \qquad \{i,j\} \in E^+ \tag{11}$$
$$y_{ijk} \in \{0,1\} \qquad (\{i,k\},j) \in E^q \tag{12}$$
$$x_{ij} \geq 0 \qquad \{i,j\} \in E \tag{13}$$
$$\xi_{ij} \geq 0 \qquad \{i,j\} \in E. \tag{14}$$

The objective represented by the expression (1) aims to minimize the total traveling costs. The quadratic nature of the costs is expressed in terms of variables y and z. Constraints (2) are the degree constraints. They impose that every customer be visited exactly once or, equivalently, that two edges are adjacent to it. Constraints (3) are the fleet-size constraints. They impose that exactly K vehicles be routed from the depot. Constraints (4) are the subtour-elimination constraints and capacity cuts. They forbid the appearance of subtours (closed tours not linked to the depot) and of routes exceeding the capacity of the vehicles. Constraints (5) impose that two-customer routes of the form $0 \to i \to j \to 0$ cannot be associated with the y variables but rather with the corresponding z variables. Constraints (6)-(7) are the linking constraints between the edges in E and the q-edges in E^q. Constraints (8)-(9) are the linking constraints between the ξ variables and variables x,w,z. Finally, constraints (10)-(14) are the integrality and non-negativity constraints of the decision variables. Note that variables x and ξ need not be imposed as integers, as this will be a direct consequence of the integrality of variables y,w,z and the linking constraints.

This formulation is an adaptation of the three-index formulation introduced by [13] for the QSTSP, and relies on the particular quadratic structure of the costs (they depend on pairs of two consecutive edges rather than on arbitrary pairs of edges) to derive a linear programming model with a cubic number of variables. As pointed out by the authors, this trade-off between the number of variables and the nature of the objective function (linear or quadratic) is usually positive towards the use of this cubic number of variables with a linear objective, rather than using a quadratic number of variables with a quadratic objective.

4. Valid inequalities

4.1 Small routes inequalities

The small routes inequalities were introduced by [5] for the Capacitated Location-Routing Problem (CLRP) under the name of *depot degree constraints*. For the QCVRP we can derive, under certain assumptions, three families of valid inequalities. The first assumption, shared by all three families, is that the number of vehicles is tight with respect to the demands, i.e. that $\lceil d(V^+)/Q \rceil = K$.

Let $S \subseteq V^+, |S| \geq 2$ be a subset of customers such that $d_i + d_j \leq Q$ for every two different customers $i,j \in S$. Let us assume that S and the traveling distances $(c_{0i})_{i\in V^+}, (c_{ijk})_{(\{i,k\},j)\in E^q}$ satisfy the following *triangle property*: for every pair of customers $i,j \in S, i<j$, $c_{0i}+c_{0j} \geq c_{0ij}+c_{0ji}$. The following single-customer routes cut is valid for the QCVRP:

$$w(S) \leq 1. \tag{15}$$

Proposition 4.1. *The single-customer routes cuts* (15) *are valid for the QCVRP.*

Proof. Two customers $i,j \in S$ cannot be simultaneously served by single-customer routes. Indeed, it will always be cheaper to route those two customers using a two-customer route of the form $0 \to i \to j \to 0$. □

Now, let $S \subseteq V^+, |S| \geq 4$ be such that $d_i + d_j + d_k + d_l \leq Q$ for every four different customers in S. Let us assume that S and the traveling distances satisfy the following *pentagon* *property*: for every four different customers $i,j,k,l \in S$, $c_{0ji}+c_{0kl} \geq c_{ijk}+c_{jkl}$. The following two-customer routes cut is valid for the QCVRP:

$$z(E(S)) \leq 1. \tag{16}$$

Proposition 4.2. *The two-customer routes cuts* (16) *are valid for the QCVRP.*

Proof. Four customers $i,j,k,l \in S$ cannot be simultaneously served by two different two-customer routes. Indeed, it will always be cheaper to route those four customers using one route servicing them all. □

Finally, let $S \subseteq V^+, |S| \geq 4$ be such that for every four different customers $i,j,k,l \in S$, $d_i+d_j+d_k+d_l \leq Q$. Let us assume that S and the traveling distances satisfy the triangle property, the pentagon property, in addition to the following *square property*: for every three different customers $i,j,k \in S$, $c_{0i}+c_{0jk} \geq c_{0ij}+c_{ijk}$. The following mixed-customer routes cut is valid for the QCVRP:

$$w(S)+z(E(S)) \leq 1. \tag{17}$$

Proposition 4.3. *The mixed-customer routes cuts* (17) *are valid for the QCVRP.*

Proof. We already know that any two different customers in S cannot be served by single-customer routes, and that any four different customers in S cannot be served by two different two-customer routes either. In addition, the square property ensures that three different customers cannot be served one by a single-customer route, and the other two by a two-customer route. Indeed, it will always be cheaper to visit those three customers using the same vehicle. □

4.2 Lifted capacity cuts

In this section we present two families of valid inequalities that, in the same spirit of the capacity cuts (4), are used to forbid vehicle routes from visiting customers that would not fit into a vehicle due to the capacity restrictions. One of the two families is called lifted capacity cuts. As its name suggests, they represent a lifting of the original capacity cuts (4) and of the y-capacity cuts introduced by [5, 9] for the CLRP. Following the same reasoning used by [13] for the QSTSP, one can observe that the q-edges with both extremities inside of S and the middle vertex outside of it can be omitted from the cut since they represent vehicles that never left set S. The following inequality is then valid for the QCVRP for every set $S \subseteq V^+$:

$$\xi(\delta(S)) - 2y(\delta^{ioi+}(S)) \geq 2r(S). \tag{18}$$

Lemma 4.4. *Inequalities* (18) *are valid for the QCVRP.*

Proof. First, note that without loss of generality we may assume that $2w(S)+2z(E(S))+2z(\delta^+(S)) = 0$. Indeed, for each of these terms equals to one, one can remove the corresponding customers from S and prove the inequality for the remaining ones, because each time that a variable w or z is equal to one, the accumulated demand associated to each route is not greater than Q and thus $r(S)$ cannot decrease by more than one unit. Let us assume first that $r(S) = 1$. If the q-edges leaving from S were only of the form $(\{i,k\},j), i,k \in S, j \in V^+ \setminus S$, then there would be a subtour. Thus, there must be at least two edges that are not of this form which will make the left-hand side of constraint (18) become at least 2. Let us assume now that $r(S) > 1$. If there were $2k$ q-edges leaving from S, with $k < r(S)$, one could follow these q-edges (thanks to the degree constraints) that would eventually get together to the depot creating at most k routes, one of which at least would violate the capacity constraint, or some of them would eventually meet in some other customer, thus creating a subtour. □

Note that this inequality is valid for every set S and not only for sets of size strictly smaller than $|V^+|/2$ as for the QSTSP since tours must necessarily be connected to the depot.

The next lifting of this last inequality is a generalization of the following observation due to [5] for the CLRP. One can strengthen the capacity cut by not considering some routes visiting one or two customers inside S. More precisely, the following inequality is valid for every set $S \subseteq V^+$, and for every subset $S' \subset S$ such that $r(S) = r(S \setminus S')$:

$$x(\delta(S)) - 2y(\delta^{ioi+}(S)) + 2(w(S \setminus S') + z(E(S \setminus S')) + z(\delta^+(S \setminus S'))) \geq 2r(S). \tag{19}$$

Proposition 4.5. *Inequalities* (19) *are valid for the QCVRP.*

Proof. For the sake of brevity we only include a brief sketch of the proof. Let $S'' \subseteq S'$ be the set of customers $i \in S'$ that are either served by a single-customer route (in which case variable w_i would be equal to one) or by a two-customer route (in which case the corresponding z_{ij} variable would take the value 1). It can be shown that the left-hand side of inequality (19) is greater than or equal to the left-hand side of inequality (18) evaluated in the set $S \setminus S''$. The result follows from applying inequality (18) to the set $S \setminus S''$ and by noticing that $r(S) = r(S \setminus S'')$. □

Remark The lifted capacity cuts (19) dominate the y-capacity cuts introduced by [5] because of the tighter left-hand side. They also extend the lifted subtour elimination constraints of [13] valid for the QSTSP.

The second family of capacity cuts takes into account q-edges only. Given that q-edges visit two customers each, one can derive an inequality that does not dominate, nor is dominated by the previous family of valid inequalities. Let $S \subset V^+$ be a customer subset of size $|S| \geq 3$. Let us define $\rho(S) = \min\{|S| - 1, r(S)\}$. The following family of q-capacity cuts is valid for the QCVRP:

$$y(E^q(S)) \leq |S| - 1 - \rho(S). \tag{20}$$

Lemma 4.6. *The q-capacity cuts* (20) *are valid for the QCVRP.*

Proof. If $r(S) = |S|$ then all customers in S must be visited by different vehicles and thus $y(E^q(S)) = 0$. Therefore, let us assume that $r(S) \leq |S| - 1$. Because there are at least $r(S)$ vehicles servicing S, then S can be partitioned into $n \geq r(S)$ non-empty sets $S_i, i = 1, \ldots, n$ such that

$$|S_i| - y(E^q(S_i)) = \begin{cases} 1 & \text{if } |S_i| = 1 \\ 2 & \text{if } |S_i| \geq 2 \end{cases}$$

and $|S| - y(E^q(S)) = \sum_{1 \leq i \leq n} |S_i| - y(E^q(S_i))$. Let us define $l_1 = |\{i : 1 \leq i \leq n, |S_i| = 1\}|, l_2 = |\{i : 1 \leq i \leq n, |S_i| \geq 2\}|$. Thus, the following identity holds:

$$|S| - y(E^q(S)) - 1 - r(S) = l_1 + 2l_2 - 1 - r(S).$$

If $l_2 = 0$, then $l_1 = |S|$ and one has $l_1 + 2l_2 - r(S) - 1 = |S| - r(S) - 1 \geq 0$. If $l_2 \geq 1$ then $l_1 + 2l_2 - r(S) - 1 = (n - r(S)) + (l_2 - 1) \geq 0$. □

As for the lifted capacity cuts (18)-(19), one can derive a similar lifting to strengthen the left-hand side of the inequality. More precisely, let $S' \subset S$ be such that $\rho(S) = \rho(S \setminus S')$. The following lifted q-capacity cut is valid for the QCVRP:

$$y(E^q(S)) + w(S') + 2z(E(S')) + z(S' : V^+ \setminus S) + y(V \setminus S : S' : V \setminus S) \leq |S| - 1 - \rho(S). \tag{21}$$

Proposition 4.7. *The lifted q-capacity cuts* (21) *are valid for the QCVRP.*

Proof. Let $S'' \subseteq S'$ be the subset of customers of S' being served by single-customer routes, two-customer routes or by multiple-customer routes using a q-edge in $(V \setminus S : S' : V \setminus S)$. Let us define $\gamma(S,S') = w(S') + 2z(E(S')) + z(S' : V^+ \setminus S) + y(V \setminus S : S' : V \setminus S)$. One has that

$$\gamma(S,S') = |S''|.$$

Also, because of the definition of S'', one also has that

$$y(E^q(S \setminus S'')) = y(E^q(S)).$$

Using these two identities in addition to the q-capacity cut (20), one can realize that the following identity and inequality also hold:

$$y(E^q(S)) + \gamma(S,S') = y(E^q(S \setminus S'')) + |S''| \leq |S \setminus S''| + |S''| - 1 - \rho(S \setminus S'').$$

Because $\rho(S \setminus S') \leq \rho(S \setminus S'') \leq \rho(S)$ the result follows. □

5. Separation algorithms

Let $QCVRP_L$ be the linear relaxation of (1)-(14). This is, the linear program resulting from replacing the binary conditions (10)-(14) by the corresponding linear conditions. Given a solution $(\overline{\xi}, \overline{x}, \overline{y}, \overline{w}, \overline{z})$ of $QCVRP_L$, a separation algorithm for a family of inequalities $\mathscr{F}$ is a method receiving $(\overline{\xi}, \overline{x}, \overline{y}, \overline{w}, \overline{z})$ as input and returning an inequality valid for $\mathscr{F}$ violated by $(\overline{\xi}, \overline{x}, \overline{y}, \overline{w}, \overline{z})$, if one exists. Note that a separation algorithm may fail in finding a violated valid inequality, in which case we refer to it as *heuristic*. Otherwise, it is said *exact*.

In this section we present separation algorithms, exact and heuristics, for all the classes of valid inequalities introduced in this paper.

The small routes inequalities (15)-(17) are not separated dynamically but rather included from the beginning of the algorithm. In particular, we do not add them for every possible subset $S \subseteq V^+$, but we rather consider the sets $V^+_{Q/2} = \{i \in V^+ : d_i \leq Q/2\}$ and $V^+_{Q/4} = \{i \in V^+ : d_i \leq Q/4\}$. We check if $V^+_{Q/2}$ satisfies the triangle property, in which case we add the corresponding inequality (15). Then, we check if the pentagon property holds for the set $V^+_{Q/4}$, in which case we add the corresponding inequality (16). Finally, if the triangle and pentagon properties hold for $V^+_{Q/4}$, we additionally check the square property before adding the corresponding inequality (17).

Now, let us present the heuristic separation algorithm designed for finding the lifted capacity cuts (19). For any two positive integers $a, b \in \mathbb{Z}_+$, let $a\%b$ be the remainder of dividing a by b or, said oterwise, if $\lfloor a/b \rfloor = r$, $a\%b := a - rb$. The separation algorithm for the lifted capacity cuts (19) uses two sequential steps, as follows.

We first use a modified version of the implementation of [18] for the separation of capacity cuts (4). The modification makes use of a parameter $\varepsilon > 0$ to find customer subsets S that either violate a capacity cut (4) or, if not, the difference between its right-hand side and left-hand side is of at most ε. For every set S found, we consider the lifted capacity cut (18) produced by removing all q-edges leaving and re-entering set S. Let us define $\kappa = (d(S) + Q - 1)\%Q$. If $\kappa = 0$, we make $S' = \emptyset$ and check if S violates inequality (18), in which case we add this cut to (1)-(14). Otherwise (i.e., if $\kappa > 0$), we solve the following 0-1 Quadratic Knapsack Problem (0-1 QKP) to find the set S':

$$\max \quad \sum_{i \in S} (\overline{w}_i + \overline{z}(i : V^+ \setminus S))\mu_i + \sum_{i \in S} \sum_{j \in S \setminus \{i\}} \overline{z}_{ij} \mu_i \mu_j \tag{22}$$

subject to

$$\sum_{i \in S} d_i \mu_i \leq \kappa \tag{23}$$

$$\mu_i \in \{0,1\} \qquad i \in S. \tag{24}$$

The set S' corresponds to the customers $i \in S$ such that $\mu_i = 1$ in the optimal solution of (22)-(24). Unfortunately, solving the 0-1 QKP is $\mathcal{NP}$-hard in the strong sense [7]. However, any feasible solution of (23)-(24) can be used to construct a valid inequality (19). We make use of a dynamic programming heuristic proposed by [12] to find a good solution of problem (22)-(24). The heuristic uses a similar recursion to that of the classical 0-1 (Linear) Knapsack Problem (0-1 KP) but with the observation that the Bellman optimality principle does not hold anymore. The method thus yields lower bounds for the 0-1 QKP. The authors report, however, that their heuristic finds the optimal solutions in a 100% of the problems tested when combined with a simple local search heuristic due to [15].

The separation problem for the q-capacity cuts (20) and their lifted version (21) is done in an analogous fashion as for inequalities (18)-(19). For the same candidate sets S, we first check whether $r(S) \leq |S|-1$ or not. If $r(s) = |S|$ we then simply check for the violation of the q-capacity cut (20) in which case we add it to problem (1)-(14). If $r(S) \leq |S|-1$, we first detect a subset $T \subset S$ such that $r(T) \leq |T|-1$. Ideally, the set T should be as small as possible. We then solve the following 0-1 QKP involving customers in $W = S \setminus T$:

$$\max \quad \sum_{i\in W} (\overline{w}_i + \overline{z}(i : V^+ \setminus S))\mu_i + \sum_{i\in W}\sum_{j\in W\setminus\{i\}} \overline{z}_{ij}\mu_i\mu_j + \sum_{i\in W}\sum_{j,k\in V\setminus S, j\leq k} \overline{y}_{jik} \tag{25}$$

subject to

$$\sum_{i\in W} d_i\mu_i \leq \kappa \tag{26}$$

$$\mu_i \in \{0,1\} \qquad i \in W. \tag{27}$$

The set S' corresponds to the customers $i \in W$ such that $\mu_i = 1$ in the solution of the above problem. Because $r(T) \leq |T|-1$ then the set S' is guaranteed to satisfy $\rho(S \setminus S') = \rho(S)$. To select a good set T, we first sort the customers in S in non-decreasing order of $\{\overline{w}_i + \overline{z}(i : V^+ \setminus \{i\}) + \sum_{j,k\in V\setminus S, j\leq k} \overline{y}_{jik} : i \in S\}$. Then, we set $T = \emptyset$ and iteratively enlarge it by adding the next customer in the sorted list. We stop when the resulting set T satisfies the property $r(T) \leq |T|-1$.

6. The branch-and-cut algorithm

In this section we present in detail the exact solver used to solve program (1)-(14), using the branch-and-cut paradigm. Initially, we consider a relaxation of (1)-(14) in which the capacity cuts (4) and the integrality constraints (10)-(12) are relaxed. The solution of this problem yields a lower bound of the QCVRP that can be used as follows: If the associated fractional solution of this problem $\overline{X} = (\overline{\xi}, \overline{x}, \overline{y}, \overline{w}, \overline{z})$ is integer, then it represents either a feasible solution of the problem, that in the first iteration also corresponds to the optimal solution of (1)-(14), or a solution violating a capacity cut that can be detected by simple inspection by following the edges used in the candidate integer solution. If $\overline{X}$ is not integer, then we have two options: we can either look for valid inequalities that may be violated by $\overline{X}$ and strengthen the problem; or we may decide to partially break the fractionality of the solution by creating two new problems, each of which explores disjoint parts of the feasible space. Each of these subproblems is then solved using the same strategy, in a recursive manner.

The branching strategy is as follows. We first try to branch on cutsets. If we cannot detect any cutset of fractional value, we branch on variables x, z and w, in that order. To branch on cutsets, we use the following trick already used in successful implementations of branch-and-cut algorithms for the CLRP [5, 9]. At the root node, each lifted capacity cut (18)-(19) is added as an equality constraint by adding a slack variable s with coefficient of -2 to the left-hand side of the inequality. We allow the solver to branch on these slack variables, and impose a branching priority to branch in these variables first. Preliminaty tests have helped us to restrict this trick to sets S such that $r(S) \leq 2$. Indeed, branching on small cutsets has a larger impact on either the bounds or the feasibility of the resulting children nodes. To choose which variable to select among a series of candidate variables of the same family, we use the strong branching rule implemented in CPLEX.

Note that our implementations of the separation procedures are thread-safe, which allows taking advantage of the parallel implementation of the IP solver of CPLEX. During our computational analyses we assess the efficiency of using this feature when compared to a serial implementation of the proposed branch-and-cut solver.

7. Computational results

To assess the efficiency of the modeling an solution approaches proposed in this article, we have run our algorithm on a selected number of instances adapted from classic instances from the literature for the CVRP, namely the sets A, B, E and P. More precisely, we consider all instances in these sets containing 60 customers or less. The routing costs have been modified to introduce penalties on the angles incurred on each pair of consecutive edges. Namely, let $(\gamma_{ij})_{\{i,j\}\in E}$ be the routing costs of the original CVRP instances. For every vertex $i \in V$ let p_i be the position of node i in the plane. For every edge $\{i,j\} \in E$ we let $\overrightarrow{p_i p_j}$ be the vector in the plane with tail in p_i and head in p_j. We also let $\overrightarrow{p_j p_i} = -\overrightarrow{p_i p_j}$. In addition, $< \cdot, \cdot >$ and $\|\cdot\|$ denote the inner product of two vectors and the norm of a vector, respectively. For every q-edge $(\{i,k\}, j) \in E^q$, the angle α_{ijk} is defined as

$$\alpha_{ijk} = \arccos\left(\frac{< \overrightarrow{p_j p_i}, \overrightarrow{p_j p_k} >}{\|\overrightarrow{p_j p_i}\|\|\overrightarrow{p_j p_k}\|}\right). \tag{28}$$

We let $\lambda \geq 0$ be a parameter representing the penalty incurred when the angle α_{ijk} is lower than π. The routing cost of a q-edge $(\{i,k\}, j) \in E^q$ is computed as follows:

$$c_{ijk} = \begin{cases} \frac{1}{2}\lfloor (2\gamma_{ij} + \gamma_{jk})(1 + \lambda \cos(\alpha_{ijk}/2)) + 0.5 \rfloor & \text{if } i = 0, j,k \in V^+ \\ \frac{1}{2}\lfloor (\gamma_{ij} + \gamma_{jk})(1 + \lambda \cos(\alpha_{ijk}/2)) + 0.5 \rfloor & \text{if } i,j,k \in V^+ \end{cases}. \tag{29}$$

Also, for the single-customer routes, the routing costs $(c_{0i})_{i\in V^+}$ are computed as $c_{0i} = \lfloor 2\gamma_{0i}(1+\lambda) + 0.5 \rfloor$ for every $i \in V^+$. Note that when $\lambda = 0$, the routing costs correspond to those of the original CVRP.

Our algorithm has been run for a maximum time of two hours for each instance. In Table 1 we report the following aggregate results: The family of instances (Column labeled *Family*); the number of instances considered for each family (Column labeled *#Inst*); the number of instances solved to optimality (Column labeled *#opt*); the average number of branch-and-bound nodes inspected (Column labeled *#N*); and the average CPU Time, in seconds, taken by the algorithm on the instances solved to optimality (Column labeled *T*).

Family	#Inst	#opt	#N	T
A	20	13	492	649
B	17	9	867	772
E	7	6	602	431
P	19	9	684	200

Table 1: Aggregate results of the exact solver for the QCVRP

The largest instance solved by our algorithm (in terms of the number of customers) is instance A-n53-k7, which took 847.8 seconds. Instead, the smallest instance that our algorithm could not solve in the allowed time of two hours is instance B-n43-k6 for which only an upper bound was available at the end of the computation, of value 839, while the best lower bound was of 830. In total, 37 instances out of 63 were solved to optimality in less than two hours. The average optimality gap at the root node for the instances solved to optimality was of 2.4%, which shows the strength of our model by its capability for producing tight lower bounds.

8. Concluding remarks

This article introduces the Quadratic Capacitated Vehicle Routing Problem (QCVRP), a combinatorial optimization problem arising in practical applications in logistics and transportation. We propose a mathematical formulation of the QCVRP based on q-edges, and strengthen it with the inclusion of some classes of valid inequalities. We show that these inequalities are not only valid for the QCVRP, but that in some cases they also dominate some known valid inequalities for the CVRP. We have implemented an exact solver based on the branch-and-cut paradigm, and tested it on several generated instances to assess its efficiency. We show that our modeling and solution approaches are efficient to provide tight lower bounds and optimal solutions of the QCVRP in moderate computing times even for some medium-size problems. As of potential avenues of future research, we believe that embedding some of the inequalities introduced in this paper into a column generation-based exact solver would result in a much more robust algorithm. Also, we believe that some new cutting planes could still be derived for the QCVRP that could help strengthening the bounds. Finally, we believe that the development of heuristic methods capable of providing good quality solutions is critical and would also accelerate the branch-and-cut method by helping it to detect non-promising branching directions.

Acknowledgements

R. Martinelli has been partially funded by the *Fonds de recherche du Québec - Nature et technologies* (FRQNT). C. Contardo has been partially funded by the ESG UQÀM under the *PAFARC* program and the *Natural Sciences and Engineering Research Council of Canada* (NSERC). These supports are gratefully acknowledged.

References

[1] **Aggarwal, A., Coppersmith, D., Khanna, S., Motwani, R. and Schieber, B.** (2000), The angular-metric traveling salesman problem. *SIAM Journal on Computing*, v. 29, p. 697–711.

[2] **Applegate, D. L., Bixby, R. E., Chvátal, V. and Cook, W. J.** *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, Princeton, 2007.

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

[4] **Bautista, J., Fernández, E. and Pereira, J.** (2008), Solving an urban waste collection problem using ants heuristics. *Computers and Operations Research*, v. 35, p. 302–309.

[5] **Belenguer, J. M., Benavent, E., Prins, C., Prodhon, C. and Wolfler-Calvo, R.** (2011), A branch-and-cut algorithm for the capacitated location routing problem. *Computers & Operations Research*, v. 38, p. 931–941.

[6] **Bräysy, O., Martínez, E., Nagata, Y. and Soler, D.** (2011), The mixed capacitated general routing problem with turn penalties. *Expert Systems with Applications*, v. 38, p. 12954–12966.

[7] **Caprara, A., Pisinger, D. and Toth, P.** (1998), Exact solution of the quadratic knapsack problem. *INFORMS Journal on Computing*, v. 11, p. 125–137.

[8] **Contardo, C.** A new exact algorithm for the multi-depot vehicle routing problem under capacity and route length constraints. Submitted for publication at *Discrete Optimization*, 2012.

[9] **Contardo, C., Cordeau, J.-F. and Gendron, B.** A computational comparison of flow formulations for the capacitated location-routing problem. Technical Report CIRRELT-2011-47, Université de Montréal, Canada, 2011.

[10] **Dantzig, G. B., Fulkerson, D. R. and Johnson, S. M.** (1954), Solution of a large-scale traveling salesman problem. *Operations Research*, v. 2, p. 393–410.

[11] **Dantzig, G. B. and Ramser, J. H.** (1959), The truck dispatching problem. *Management Science*, v. 6, p. 80–91.

[12] **Djeumou Fomeni, F. and Letchford, A. N.** (2013), A dynamic programming heuristic for the quadratic knapsack problem. *INFORMS Journal of Computing*. Forthcoming.

[13] **Fischer, A. and Helmberg, C.** (2012), The symmetric quadratic traveling salesman problem. *Mathematical Programming*. Forthcoming.

[14] **Fukasawa, R., Longo, H., Lysgaard, J., Poggi de Aragão, M., Reis, M., Uchoa, E. and Werneck, R. F.** (2006), Robust branch-and-cut-and-price for the capacitated vehicle routing problem. *Mathematical Programming Series A*, v. 106, p. 491–511.

[15] **Gallo, G., Hammer, P. L. and Simeone, B.** (1980), Quadratic knapsack problems. *Mathematical Programming Studies*, v. 12, p. 132–149.

[16] **Jünger, M., Thienel, S. and Reinelt, G.** (1994), Provably good solutions for the traveling salesman problem. *Zeitschrift für Operations Research*, v. 40, p. 183–217.

[17] **Laporte, G., Nobert, Y. and Desrochers, M.** (1985), Optimal routing under capacity and distance restrictions. *Operations Research*, v. 33, p. 1050–1073.

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

[19] **Padberg, M. W. and Hong, S.** (1980), On the symmetric travelling salesman problem: a computational study. *Mathematical Programming Study*, v. 12, p. 78–107.

[20] **Perrier, N., Langevin, A. and Amaya, C. A.** (2008), Vehicle routing for urban snow plowing operations. *Transportation Science*, v. 42, p. 44–56.