Appendix A: Proof of Proposition 1

We repeat Proposition 1 and provide a proof.

PROPOSITION 1 *The p-step formulation* (1)-(5) *is valid for the CVRP.*

Proof Consider first any solution to the CVRP. Such a solution consists of several routes. We can decompose a route of length l into one or more p-steps r_i, with $i = 0, \ldots, \lceil l/p \rceil - 1$, as follows. The partial paths P_{r_i} are obtained by recursively taking the last p arcs of the route, until at most p arcs are left. We then define $d_{r_0} = 0$, and, for $i \geq 1$, we define $d_{r_i} = \sum_{j=0}^{i-1}(q(r_j) - q^-_{r_j})$, where $q^-_{r_j}$ is the demand of the last location on p-step r_j. Setting $x_r = 1$ for all p-steps that are constructed this way and $x_r = 0$ for all other p-steps provides a feasible solution to (1)-(5).

Next, consider any solution to (1)-(5). By (3), if a node appears as the last node on a p-step that is selected, it appears as the first node on a selected p-step, as well. By (2), each customer appears on a selected p-step either once as an intermediate node, or twice: once as the first node and once as the last node on a selected p-step. In both cases, the customer has one incoming arc and one outgoing arc in the selected p-steps. Any solution to (1)-(5) thus induces an integer arc-flow in G. Such an arc-flow can be decomposed into paths from 0 to $n+1$ and cycles in G (Ahuja et al. 1993, Theorem 3.5). What remains to be shown is that cycles cannot exist and that the paths respect the vehicle capacity.

We show by contradiction that cycles cannot exist. Consider any cycle and let $C \subseteq N'$ be the nodes visited by the cycle. We define R^p_C as the set of p-steps of which the start or end location is on the cycle C. Because the arc-flow is binary, it follows that any p-step $r \in R^p_C$ that is selected in the current solution actually has all locations on the cycle. Using (4) and denoting the demand of the first customer on a p-step r by q^+_r, this allows us to derive

$$\begin{aligned} 0 &\leq \sum_{i \in C} \sum_{r \in R^p} q^i_r x_r \\ &= \sum_{i \in C} \sum_{r \in R^p_C} q^i_r x_r \\ &= \sum_{r \in R^p_C} \sum_{i \in C} q^i_r x_r \\ &= \sum_{r \in R^p_C} \left(d_r + q^+_r - d_r - q(r)\right) x_r . \end{aligned}$$

Because $q^+_r - q(r) < 0$ for all $r \in R^p_C$, this shows that such a cycle cannot exist.

Next, consider a path from the starting depot to the ending depot. This path is a concatenation of p-steps $r_0, \ldots, r_k$, for some integer $k \geq 0$. All these p-steps are selected with value 1. If $k = 0$, then the path contains at most p arcs. By definition, it then holds that $d_{r_0} \leq Q - q(r_0)$. Hence, $q(r_0) \leq Q - d_{r_0} \leq Q$. Therefore, the path respects the vehicle capacity. If $k > 0$, then we can rewrite (4) for the final location on p-step r_i with $i = 0, \ldots, k-1$, as

$$d_{r_i} + q(r_i) \leq d_{r_{i+1}} + q^+_{r_{i+1}} .$$

Summing the above over $i = 0$ to $i = k-1$, we obtain

$$\sum_{i=0}^{k-1} \left(d_{r_i} + q(r_i)\right) \leq \sum_{i=1}^{k} \left(d_{r_i} + q^+_{r_i}\right) .$$

Dropping common terms and using that $q^+_{r_0} = 0$, this can be rewritten to

$$d_{r_0} + \sum_{i=0}^{k-1} \left(q(r_i) - q^+_{r_i}\right) - q^+_{r_k} \leq d_{r_k}.$$

By definition, it holds that $d_{r_k} \leq Q - q(r_k)$. Thus,

$$d_{r_0} + \sum_{i=0}^{k} (q(r_i) - q^+_{r_i}) \leq Q.$$

Because $d_{r_0} \geq 0$ and $\sum_{i=0}^{k}(q(r_i) - q^+_{r_i})$ is equal to the demand of the customers on the path, we see that the path respects the vehicle capacity. Hence, any solution to (1)-(5) represents a solution to the CVRP. □

Appendix B: Proof of Proposition 2

Before proving Proposition 2, we first provide SCF. This mixed integer programming formulation makes use of the continuous variables f_{ij} which can be interpreted as the total load carried by the vehicle traversing arc $(i, j) \in A$ and binary variables $\vec{\theta}_{ij}$ that represent whether arc $(i, j) \in A$ is selected or not. The formulation is the following.

$$\text{(SCF)} \quad \min \sum_{(i,j)\in A} c_{ij}\vec{\theta}_{ij} \tag{EC.1}$$

$$\sum_{j\in N} \vec{\theta}_{ji} = 1 \qquad \forall i \in N' \tag{EC.2}$$

$$\sum_{j\in N} \vec{\theta}_{ij} = 1 \qquad \forall i \in N' \tag{EC.3}$$

$$\sum_{j\in N} f_{ji} - \sum_{j\in N} f_{ij} = q_i \qquad \forall i \in N' \tag{EC.4}$$

$$q_j\vec{\theta}_{ij} \leq f_{ij} \qquad \forall (i, j) \in A \tag{EC.5}$$

$$f_{ij} \leq (Q - q_i)\,\vec{\theta}_{ij} \qquad \forall (i, j) \in A \tag{EC.6}$$

$$\vec{\theta}_{ij} \in \{0, 1\} \qquad \forall (i, j) \in A. \tag{EC.7}$$

The objective (EC.1) is to minimize the total costs. Constraints (EC.2) and (EC.3) are the indegree and outdegree constraints. Constraints (EC.4) ensure that after visiting a customer the vehicle load is reduced by the demand. Constraints (EC.5) specify that when visiting a customer, a vehicle should carry at least the demand. Furthermore, constraints (EC.6) ensure that the vehicle load does not exceed the capacity. Finally, (EC.7) are the binarity conditions on the arc-flow variables.

Before presenting a proof of Proposition 2, stating that the 1-step formulation has the same LP bound as SCF, we first show a stronger result. We show that actually for any choice of p, the p-step formulation is at least as strong as SCF.

LEMMA EC.1. *For any $p \in \mathbb{N}_{>0}$, the p-step formulation is at least as strong as SCF.*

Proof Consider any feasible solution x to the LP relaxation of the p-step formulation (1)-(5). We show that there exists a corresponding solution to the LP relaxation of SCF with the same objective value, proving the p-step formulation is at least as strong as SCF.

Define the binary parameters $\vec{b}^r_{ij}$ indicating whether arc $(i,j) \in A$ is included on p-step $r \in R^p$, and define for each arc $(i,j) \in A$,

$$\vec{\theta}_{ij} = \sum_{r \in R^p} \vec{b}^r_{ij} x_r.$$

We first establish that the definition of $\vec{\theta}$ yields the same objective value, and satisfaction of (EC.2), (EC.3), and $0 \leq \vec{\theta}_{ij} \leq 1$ for all $(i,j) \in A$. Then we establish existence of a corresponding $\bar{f}$ such that $\left(\vec{\theta}, \bar{f}\right)$ satisfies (EC.4)-(EC.6).

First consider the objective value. By definition of $\vec{\theta}$, the objective value of $(\vec{\theta}, \bar{f})$ is given by

$$\begin{aligned}
\sum_{(i,j)\in A} c_{ij}\vec{\theta}_{ij} &= \sum_{(i,j)\in A} c_{ij} \sum_{r \in R^p} \vec{b}^r_{ij} x_r \\
&= \sum_{r \in R^p} \left(\sum_{(i,j)\in A} c_{ij} \vec{b}^r_{ij} \right) x_r \\
&= \sum_{r \in R^p} c_r x_r.
\end{aligned}$$

Next, denote by R_i^+ the collection of p-steps that start at i, and R_i^- for the collection of p-steps that end at i. Furthermore, let $R_i^=$ be the collection of p-steps that include i but do not start nor end at i. It follows from (3) that $\sum_{r\in R_i^+} x_r = \sum_{r \in R_i^-} x_r$ for any $i \in N'$. Using this and (2), we derive for all $i \in N'$ that

$$\begin{aligned}
1 &= \frac{1}{2} \sum_{r \in R^p} a^i_r x_r \\
&= \frac{1}{2} \sum_{r \in R_i^+} x_r + \sum_{r \in R_i^=} x_r + \frac{1}{2} \sum_{r \in R_i^-} x_r \\
&= \sum_{r \in R_i^+} x_r + \sum_{r \in R_i^=} x_r \\
&= \sum_{r \in R_i^=} x_r + \sum_{r \in R_i^-} x_r.
\end{aligned}$$

We now obtain for all $i \in N'$,

$$\begin{aligned}
\sum_{j \in N} \vec{\theta}_{ji} &= \sum_{j \in N} \sum_{r \in R^p} \vec{b}^r_{ji} x_r \\
&= \sum_{r \in R^p} \left(\sum_{j \in N} \vec{b}^r_{ji} \right) x_r \\
&= \sum_{r \in R_i^= \cup R_i^-} x_r \\
&= 1,
\end{aligned}$$

showing that (EC.2) are satisfied. Similarly,

$$\begin{aligned}\sum_{j\in N}\vec{\theta}_{ij} &= \sum_{j\in N}\sum_{r\in R^p}\vec{b}^r_{ij}x_r\\ &= \sum_{r\in R^p}\left(\sum_{j\in N}\vec{b}^r_{ij}\right)x_r\\ &= \sum_{r\in R^+_i\cup R^=_i}x_r\\ &= 1,\end{aligned}$$

showing that (EC.3) are satisfied. Observe that trivially $\vec{\theta}_{ij}\geq 0$ for all $(i,j)\in A$, and combined with (EC.2) and (EC.3) this implies that also $\vec{\theta}_{ij}\leq 1$ for all $(i,j)\in A$.

Next, we prove existence of an $\bar{f}$ such that $\left(\vec{\theta},\bar{f}\right)$ satisfies (EC.4)-(EC.6). We do this in two steps, by first defining a flow f which may not satisfy (EC.4), and then proving that it can be modified such that the modified flow $\bar{f}$ does. In order to do so, define $q(r,i)$ as the sum of the demands of the customers that are visited on p-step $r\in R^p$ up to and including location $i\in N$, if i is visited by r, and define for each arc $(i,j)\in A$,

$$f_{ij} = \sum_{r\in R^p}\vec{b}^r_{ij}(Q-d_r-q(r,i))x_r.$$

Observe that for any p-step $r\in R^p$, it holds that $d_r+q(r)\leq Q$. Furthermore, if p-step r includes arc $(i,j)\in A$, then $d_r+q(r,i)+q_j\leq d_r+q(r)\leq Q$. Hence,

$$\begin{aligned}q_j\vec{\theta}_{ij} &= \sum_{r\in R^p}\vec{b}^r_{ij}x_r q_j\\ &\leq \sum_{r\in R^p}\vec{b}^r_{ij}x_r(Q-d_r-q(r,i))\\ &= f_{ij},\end{aligned}$$

showing that $\left(\vec{\theta},f\right)$ satisfies (EC.5). Similarly,

$$\begin{aligned}(Q-q_i)\vec{\theta}_{ij} &= \sum_{r\in R^p}\vec{b}^r_{ij}x_r(Q-q_i)\\ &\geq \sum_{r\in R^p}\vec{b}^r_{ij}x_r(Q-d_r-q(r,i))\\ &= f_{ij},\end{aligned}$$

showing that $\left(\vec{\theta},f\right)$ satisfies (EC.6).

We define $Q'_r = Q - d_r$ and derive for $i \in N'$ that

$$
\begin{aligned}
&\sum_{j\in N} f_{ji} - \sum_{j\in N} f_{ij}\\
&\quad= \sum_{j\in N}\sum_{r\in R^p} \left(\vec{b}^r_{ji}(Q'_r - q(r,j)) - \vec{b}^r_{ij}(Q'_r - q(r,i))\right) x_r\\
&\quad= \sum_{r\in R_i^+} x_r \sum_{j\in N} \left(-\vec{b}^r_{ij}(Q'_r - q_i)\right)\\
&\qquad+ \sum_{r\in R_i^=} q_i x_r + \sum_{r\in R_i^-} x_r \sum_{j\in N} \left(\vec{b}^r_{ji}(Q'_r - q(r,j))\right)\\
&\quad= Q\left(-\sum_{r\in R_i^+} x_r + \sum_{r\in R_i^-} x_r\right)\\
&\qquad+ \sum_{r\in R_i^+} q^i_r x_r + \sum_{r\in R_i^=} q_i x_r + \sum_{r\in R_i^-} x_r \left(q^i_r + q_i\right)\\
&\quad= q_i\left(\sum_{r\in R_i^= \cup R_i^-} x_r\right) + \sum_{r\in R_i^+\cup R_i^-} q^i_r x_r\\
&\quad\geq q_i,
\end{aligned}
$$

where we use (4) to derive the final inequality by noting that $q^i_r = 0$ for all $r \in R_i^=$. Next, we establish the existence of a flow f' such that $\left(\vec{\theta}, f - f'\right)$ satisfies (EC.4)-(EC.6).

First, define $G' = (V', A')$ as G with additional node t and additional arcs (i,t) for $i \in N' \cup \{n+1\}$. For $i \in N'$, define $f_{it} = \sum_{j\in N} f_{ji} - \sum_{j\in N} f_{ij}$ as flow from i to t. We showed that $f_{it} \geq q_i$ for all $i \in N'$. Moreover, define $f_{(n+1)t} = \sum_{j\in N} f_{j(n+1)}$. By doing so, we have flow conservation at every node of G' except 0 and t. Define $F = \sum_{j\in N} f_{0j}$ as the total flow from 0 to t. It follows that the flow over each $(0,t)$-cut is at least F.

Define for each arc $(i,j) \in A$ a capacity $u'_{ij} = f_{ij} - q_j\vec{\theta}_{ij}$, as well as $u'_{it} = f_{it} - q_i$ for $i \in N'$ and $u'_{(n+1)t} = f_{(n+1)t}$. By construction, it holds for all flows f' in G' that respect the capacities that $f - f'$ satisfies (EC.5) and (EC.6). We aim to find the maximum flow f' in G' satisfying the capacities u'. Consider any $(0,t)$-cut (S,T). It holds that

$$
\begin{aligned}
\sum_{i\in S}\sum_{j\in T} u'_{ij} &= \sum_{i\in S}(f_{it} - q_i) + \sum_{i\in S}\sum_{j'\in T\setminus\{t\}} (f_{ij} - q_j\vec{\theta}_{ij})\\
&= \sum_{i\in S}\sum_{j\in T} f_{ij} - \sum_{i\in S} q_i - \sum_{i\in S}\sum_{j'\in T\setminus\{t\}} q_j\vec{\theta}_{ij}\\
&\geq F - \sum_{i\in S} q_i - \sum_{i\in N}\sum_{j'\in T\setminus\{t\}} q_j\vec{\theta}_{ij}\\
&= F - \sum_{i\in N} q_i.
\end{aligned}
$$

In particular, for $S = N$ and $T = \{t\}$, we have

$$\sum_{i\in S}\sum_{j\in T} u'_{ij} = \sum_{i\in N} f_{it} - \sum_{i\in N} q_i = F - \sum_{i\in N} q_i.$$

This shows that there exists a flow f' in G' of value $F - \sum_{i\in N} q_i$. Defining $\bar{f} = f - f'$, $\bar{f}$ satisfies (EC.5)-(EC.6) by construction. It holds that the total flow of $\bar{f}$ equals

$$F - \left(F - \sum_{i\in N} q_i\right) = \sum_{i\in N} q_i.$$

This implies that $\bar{f}_{it} = q_i$ for all $\in N'$. It follows that $\bar{f}$ satisfies (EC.4) for all $i \in N'$ as well.

We conclude that $\left(\vec{\theta}, \bar{f}\right)$ is a feasible solution to the LP relaxation of SCF with the same objective value as the solution x to the p-step formulation. Thus, the p-step formulation is at least as strong as SCF. □

We repeat Proposition 2 and provide a proof.

PROPSITION 2 *The* 1*-step formulation is equally strong as SCF.*

Proof By Lemma EC.1 it holds that the 1-step formulation is at least as strong as SCF. What remains to be shown is that SCF is at least as strong as the 1-step formulation, by demonstrating that any feasible solution to the LP relaxation of SCF corresponds to a feasible solution to the LP relaxation of the 1-step formulation, with the same objective value. Consider a solution to the LP relaxation of SCF. For all $(i, j) \in A$ such that $\vec{\theta}_{ij} > 0$, define the 1-step

$$r(i,j) = \left((i,j), Q - \frac{f_{ij}}{\vec{\theta}_{ij}} - q_i\right)$$

and let $x_{r(i,j)} = \vec{\theta}_{ij}$. From (EC.5), it follows that

$$d_r = Q - \frac{f_{ij}}{\vec{\theta}_{ij}} - q_i \leq Q - \frac{q_j\vec{\theta}_{ij}}{\vec{\theta}_{ij}} - q_i = Q - q_i - q_j.$$

From (EC.6), it follows that

$$d_r = Q - \frac{f_{ij}}{\vec{\theta}_{ij}} - q_i \geq Q - \frac{(Q - q_i)\vec{\theta}_{ij}}{\vec{\theta}_{ij}} - q_i = 0.$$

This shows that the 1-step $((i, j), d_r)$ is feasible.

It is a standard result that (2) and (3) coincide with (EC.2) and (EC.3). It remains to be shown that the solution satisfies (4). In order to do so, note that $q^i_r \neq 0$ only if i is visited by p-step r, that is, if $P_r = (i, j)$ or if $P_r = (j,i)$ for some $j \in N$. Also note that

$$q^i_{r(i,j)} = Q - \frac{f_{ij}}{\vec{\theta}_{ij}}$$

and

$$q^i_{r(j,i)} = -\left(Q - \frac{f_{ji}}{\vec{\theta}_{ji}} - q_j\right) - q_i - q_j = -\left(Q - \frac{f_{ji}}{\vec{\theta}_{ji}} + q_i\right).$$

This allows us to derive

$$
\begin{aligned}
\sum_{r\in R^p} q_r^i x_r &= \sum_{j\in N} q^i_{r(i,j)} x_{r(i,j)} + \sum_{j\in N} q^i_{r(j,i)} x_{r(j,i)} \\
&= \sum_{j\in N} \left(Q - \frac{f_{ij}}{\vec{\theta}_{ij}} \right) \vec{\theta}_{ij} \\
&\quad - \sum_{j\in N} \left(Q - \frac{f_{ji}}{\vec{\theta}_{ji}} + q_i \right) \vec{\theta}_{ji} \\
&= Q \sum_{j\in N} \vec{\theta}_{ij} - Q \sum_{j\in N} \vec{\theta}_{ji} \\
&\quad - q_i \sum_{j\in N} \vec{\theta}_{ji} + \sum_{j\in N} f_{ji} - \sum_{j\in N} f_{ij} \\
&= 0,
\end{aligned}
$$

where the last equation follows from (EC.2)-(EC.4). Thus, the solution satisfies (4). Since the corresponding solution is feasible and clearly has the same solution value as the original, SCF is at least as strong as the 1-step formulation.

We conclude that the 1-step formulation and SCF are equally strong. □

Appendix C: Set Partitioning Formulation

Although the set partitioning formulation is effectively the p-step formulation for $p = n+1$, we include the most common representation of the set partitioning formulation below, for ease of reference. Let R be the set of all feasible routes. We emphasize that each route in R is an elementary path from the starting depot 0 to the ending depot $n+1$. For each route $r \in R$ let c_r be the corresponding cost and let α_r^i be 1 if location $i \in N'$ is included on route r, and 0 otherwise. Furthermore, the binary variable x_r indicates whether route $r \in R$ is selected or not. The set partitioning formulation (SPF) of the CVRP is the following.

$$
\text{(SPF)} \quad \min \sum_{r\in R} c_r x_r \tag{EC.8}
$$

$$
\sum_{r\in R} \alpha_r^i x_r = 1 \qquad \forall i \in N' \tag{EC.9}
$$

$$
x_r \in \{0,1\} \qquad \forall r \in R. \tag{EC.10}
$$

The objective (EC.8) is to minimize the total cost. Constraints (EC.9) ensure that every customer is visited exactly once, and (EC.10) are the binarity conditions.

Appendix D: Proof of Proposition 3

We repeat Proposition 3 and provide a proof.

PROPOSITION 3 *For any $p, m \in \mathbb{N}_{>0}$ it follows that $z_p \leq z_{pm}$.*

Proof Consider an optimal solution x^* to the LP relaxation using pm-steps. Any given pm-step r can be cut into k separate p-steps $r_1, \ldots, r_k$, where k is at most m. Denoting by q_r^- the demand of the final location on p-step $r \in R^p$, the prior demands d_{r_i} are straightforwardly chosen as $d_{r_i} = d_r + \sum_{j=1}^{i-1} \left(q(r_j) - q^-_{r_j} \right)$. Note that any p-step r_i obtained this way, may occur as part of multiple pm-steps. Denote by $R(r_i)$ all pm-steps which yield p-step r_i when cutting them into separate p-steps. We define $x'_{r_i} = \sum_{r' \in R(r_i)} x^*_{r'}$ for all $1 \le i \le k$. It is easily verified that x' is a feasible solution using exclusively p-steps, while maintaining the same objective value. This shows that $z_p \le z_{pm}$. □

Appendix E: Proof of Proposition 4

Before presenting a proof of Proposition 4, we first provide another result which will be used in our proof. The following lemma is similar to what Gouveia (1995) showed for the single-commodity flow formulation of the CVRP. It states that the flow out of and into the depot is at least the total demand divided by the vehicle capacity.

LEMMA EC.2. *Denote by R_0 and R_{n+1} the p-steps connected to the starting and ending depot respectively, and let x be a feasible solution of the LP relaxation of the p-step formulation. Then*

$$\sum_{r \in R_0} x_r = \sum_{r \in R_{n+1}} x_r \ge \frac{1}{Q} \sum_{i \in N'} q_i .$$

Proof By (3), for any feasible solution, the amount of selected p-steps starting and ending at any location $i \in N'$ is equal. We can rewrite these equalities using the following notation. For every $i \in N$, let R_i^+ be the collection of p-steps that start at i, and let R_i^- be the collection of p-steps ending at i. Note that $R_0^+ = R_0$ and $R_{n+1}^- = R_{n+1}$. Using this notation and (3), it follows that

$$\begin{aligned}
\sum_{r \in R_i^+} x_r &= \sum_{r \in R_i^+} x_r - \sum_{r \in R^p} e_r^i x_r \\
&= \sum_{r \in R_i^+} \left(1 - e_r^i\right) x_r + \sum_{r \in R_i^-} \left(-e_r^i\right) x_r \\
&= \sum_{r \in R_i^-} x_r .
\end{aligned}$$

To show that the flow out of and into the depot is also the same, we first derive

$$\begin{aligned}
\sum_{r \in R_0} x_r + \sum_{i \in N'} \sum_{r \in R_i^+} x_r &= \sum_{i \in N} \sum_{r \in R_i^+} x_r \\
&= \sum_{r \in R^p} x_r \\
&= \sum_{i \in N} \sum_{r \in R_i^-} x_r \\
&= \sum_{r \in R_{n+1}} x_r + \sum_{i \in N'} \sum_{r \in R_i^-} x_r .
\end{aligned}$$

Combining the above shows that

$$\sum_{r \in R_0} x_r = \sum_{r \in R_{n+1}} x_r .$$

To derive a bound on the flow out of or into the depot, we proceed by relating the demand to the p-step variables. We use q_r^+ to denote the demand of the first location on p-step r. Note in particular that we define $q_r^+ = 0$ for all p-steps starting at the depot. Using (2), (3) and (4), it follows that

$$\begin{aligned}
\sum_{i \in N'} q_i &\leq \sum_{i \in N'} \left(\frac{1}{2} \sum_{r \in R^p} a_r^i x_r - \frac{1}{2} \sum_{r \in R^p} e_r^i x_r \right) q_i \\
&\quad + \sum_{i \in N'} \sum_{r \in R^p} q_r^i x_r \\
&= \sum_{r \in R^p} \sum_{i \in N'} q_i \frac{1}{2} \left(a_r^i - e_r^i \right) x_r + \sum_{r \in R^p} \sum_{i \in N'} q_r^i x_r \\
&= \sum_{r \in R^p} \left(q(r) - q_r^+ + \sum_{i \in N'} q_r^i \right) x_r .
\end{aligned}$$

Observe that for every $r \in R^p \setminus (R_0 \cup R_{n+1})$, that is, for every p-step that does not start nor end at the depot, q_r^i is nonzero for exactly two customers $i \in N'$. In particular, it holds that

$$\begin{aligned}
q(r) - q_r^+ + \sum_{i \in N'} q_r^i &= q(r) - q_r^+ + (d_r + q_r^+) - (d_r + q(r)) \\
&= 0.
\end{aligned}$$

We now obtain

$$\begin{aligned}
\sum_{i \in N'} q_i &\leq \sum_{r \in R^p} \left(q(r) - q_r^+ + \sum_{i \in N'} q_r^i \right) x_r \\
&= \sum_{r \in R_0 \setminus R_{n+1}} (-d_r - q_r^+) x_r \\
&\quad + \sum_{r \in R_{n+1} \setminus R_0} (d_r + q(r)) x_r \\
&\quad + \sum_{r \in R_0 \cap R_{n+1}} (q(r) - q_r^+) x_r \\
&\leq \sum_{r \in R_{n+1}} Q x_r ,
\end{aligned}$$

where we use $0 \leq d_r + q_r^+$ and $q(r) - q_r^+ \leq d_r + q(r) \leq Q$ to derive the final inequality. This shows that $\sum_{r \in R_{n+1}} x_r \geq \frac{1}{Q} \sum_{i \in N'} q_i$. □

We now repeat Proposition 4 and provide a proof.

PROPOSITION 4 *For any $p, q \in \mathbb{N}_{>0}$ such that $p < q$ and q is not an integer multiple of p, there exists an instance of the CVRP for which $z_p > z_q$.*

Proof In this proof, we first present an instance of the CVRP. Then, we provide a feasible solution to the LP relaxation of the q-step formulation for this instance. Finally, we show that any feasible solution to the LP relaxation of the p-step formulation for this instance has a higher cost, showing $z_p > z_q$.

As q is not an integer multiple of p, there exist integers $m \geq 1$ and $k \in \{1, \dots, p-1\}$ such that $q = pm + k$. Consider the following instance with $n = (m+1)\,p$ customers. In this instance, the travel costs are equal to the Euclidean distances between locations. The customers appear in $m+1$ clusters of p customers. The clusters are located on the vertices of a regular convex $(m+1)$-polygon with edges of length 1. Within a cluster the travel costs are 0 (or negligibly small), while traveling from a customer to a customer in a neighbouring cluster has cost 1 (or negligibly close to 1). Finally, the depot is located far away, and travel from any customer to the depot has a very large cost which is (negligibly close to) $C > 1$. All customers have unit demand and the vehicle capacity is $Q = p + \frac{pq}{k}$. By construction, $Q > p + q$ holds because $k < p$.

For ease of notation, let the clusters be numbered from 1 to $m+1$ as they appear along the boundary of the polygon. As a result, the cost of traveling between a customer in cluster c and $c+1$ is 1, for $1 \leq c \leq m$, as is the cost of travel between clusters $m+1$ and 1. Similarly let the customers be numbered such that customers $(c-1)p+1$ to cp are in cluster c.

Next, we construct a feasible solution for the LP relaxation of the q-step formulation of this instance. To do so, first consider the following path P_c^t visiting all customers in cluster c in order of numbering, but cyclically permuted t times, for $0 \leq t < p$.

$$\begin{aligned}P_c^t =& (c-1)p + 1 + (t \bmod p) \rightarrow \dots \\ & \rightarrow (c-1)p + 1 + ((t+p-1) \bmod p)\,.\end{aligned}$$

For example, if we consider cluster $c = 2$ in case each cluster contains $p = 4$ customers, and the customer are cyclically permuted $t = 2$ times, then $P_c^t = 7 \rightarrow 8 \rightarrow 5 \rightarrow 6$. Next, we concatenate these paths for all clusters, starting with cluster c and obtain the path $P(c,t)$, for $1 \leq c \leq m+1$ and $0 \leq t < p$.

$$\begin{aligned}P(c,t) =& P^t_{1+(c-1) \bmod (m+1)} \rightarrow P^t_{1+(c) \bmod (m+1)} \rightarrow \dots \\ & \rightarrow P^t_{1+(c+m-1) \bmod (m+1)}.\end{aligned}$$

For example, if $m+1 = 3$ and $p = 3$, $P(2,1) = 5 \rightarrow 6 \rightarrow 4 \rightarrow 8 \rightarrow 9 \rightarrow 7 \rightarrow 2 \rightarrow 3 \rightarrow 1$. Note that the cost of traversing all arcs in any path $P(c,t)$ is m.

Finally, we truncate path $P(c,t)$ by removing the last arcs, to construct a path $P(c,t,q)$ traversing exactly q arcs. Note that also the cost of traversing all arcs on any path $P(c,t,q)$ is m. Moreover, observe that for a given position between 1 and $q+1$, every customer $i \in V'$ appears at that position on exactly one path $P(c,t,q)$, among all $1 \leq c \leq m+1$ and $0 \leq t < p$. Similarly, we introduce truncated paths $P(c,t,q-1)$ traversing the first $q-1$ arcs of $P(c,t)$. Note that the total arc costs of these paths are also m.

Using the above paths, we now present a feasible solution to the LP relaxation of the q-step formulation. It selects all of the following q-steps with a positive value, for all $1 \leq c \leq m+1$ and $0 \leq t < p$:

Type 1 $\left(0 \rightarrow P_c^t, 0\right)$ with total demand p,

Type 2 $(P(c,t,q), 0)$ with total demand $q+1$,

Type 3 $(P(c,t,q-1) \to n+1, Q-q)$ with total demand q.

Observe that given the value of Q, each selected q-step r is feasible since $0 \le d_r + q(r) \le Q$. The q-steps of type 1 and 3 are selected with value $\frac{1}{Q}$ each, and the q-steps of type 2 are selected with value $\frac{Q-p-q+1}{Qq}$. We verify that this is indeed a feasible solution.

First observe that the selection of p-steps is non-negative. Next, consider (2). Observe that every customer $i \in N'$ appears once as the last node of exactly one q-step of type 1, once as the first and once as the last node of a q-step of type 2, and once as the first node of a q-step of type 3. In these cases, $a^i_r = 1$. Moreover, every customer $i \in N'$ appears as not the first or last node for $p-1$ times on a q-step of type 1, for $q-1$ times on a q-step of type 2 and $q-1$ times on a q-step of type 3. Here, $a^i_r = 2$. Therefore, it follows for all $i \in N'$ that

$$\begin{aligned}\sum_{r\in R^p} a^i_r x_r &= \frac{1}{Q} + \frac{Q-p-q+1}{Qq} + \frac{Q-p-q+1}{Qq} + \frac{1}{Q} \\ &\quad + 2\left(\frac{p-1}{Q} + (q-1)\frac{Q-p-q+1}{Qq} + \frac{q-1}{Q}\right) \\ &= 2.\end{aligned}$$

Hence, the suggested solution satisfies (2). Similarly, we verify that also (3) are satisfied. Since every customer $i \in N'$ appears as the last node of one q-step of type 1 and one of type 2, and as the first node of one q-step of type 2 and one of type 3, it follows that

$$\sum_{r\in R^p} e^i_r x_r = -\frac{1}{Q} - \frac{Q-p-q+1}{Qq} + \frac{Q-p-q+1}{Qq} + \frac{1}{Q} = 0.$$

Finally, we verify that (4) are satisfied. Note that for a q-step of type 1 with customer $i \in N'$ as last location it holds that $q^i_r = -p$. For a q-step of type 2 with i as first location $q^i_r = 1$, while if i is the last location $q^i_r = -(q+1)$. Moreover, for a q-step of type 3 with i as first location $q^i_r = (Q-q+1)$. It follows that

$$\begin{aligned}\sum_{r\in R^p} q^i_r x_r =& \frac{-p}{Q} + \frac{Q-p-q+1}{Qq} \\ & -(q+1)\frac{Q-p-q+1}{Qq} + \frac{Q-q+1}{Q} \\ =& 0.\end{aligned}$$

We conclude that the suggested solution is feasible for the LP relaxation of the q-step formulation. Moreover, observe that the cost of a q-step of type 1, 2 and 3 is C, m and $C+m$ respectively, and n of each type are selected in the solution. As a result, the cost of the provided solution is $\frac{2n}{Q}C + \frac{n(Q-p+1)}{Qq}m$. Hence, $z_q \le \frac{2n}{Q}C + \frac{n(Q-p+1)}{Qq}m$.

We complete this proof by showing that any feasible solution to the LP relaxation of the p-step formulation for this instance has a higher cost than the presented solution to the q-step formulation. Let us first separately consider the p-steps connected to a depot. Denote by R_0 the p-steps connected to the starting depot 0, and denote by R_{n+1} the p-steps connected to the ending depot $n+1$. Observe that $R_0 \cap R_{n+1}$ contains all p-steps that start at 0 and end at $n+1$.

For such p-steps $r \in R_0 \cap R_{n+1}$ it holds that $c_r \geq 2C$, and for the p-steps $r \in R_0 \cup R_{n+1} \setminus R_0 \cap R_{n+1}$ it holds that $c_r \geq C$. This allows us to derive the following bound on the costs of the selected p-steps in any feasible solution to the LP relaxation of the p-step formulation.

$$\sum_{r \in R_0 \cup R_{n+1}} c_r x_r \geq \sum_{r \in R_0} C x_r + \sum_{r \in R_{n+1}} C x_r .$$

Next, we derive a bound on $\sum_{r \in R^p} x_r$, which could be interpreted as the amount of selected p-steps, by relating it to the degree of each node in a solution. For every p-step $r \in R^p \setminus (R_0 \cup R_{n+1})$, which are the p-steps not connected to a depot, it holds that exactly p arcs are traversed. Observe that for such a p-steps r it holds that $\sum_{i \in N'} a_r^i = 2p$, and equivalently $\frac{1}{2p} \sum_{i \in N'} a_r^i = 1$. Similarly, for all $r \in R_{n+1} \setminus R_0$ it holds that $1 + \sum_{i \in N'} a_r^i = 2p$. For all $r \in R_0$, we know that at most p arcs are traversed. Hence, for all $r \in R_0 \setminus R_{n+1}$ it holds that $1 + \sum_{i \in N'} a_r^i \leq 2p$, and for $r \in R_0 \cap R_{n+1}$ it holds that $2 + \sum_{i \in N'} a_r^i \leq 2p$. Combining these observations, and using (2) yields

$$\begin{aligned} \sum_{r \in R^p} x_r &\geq \frac{1}{2p} \left(\sum_{r \in R^p} \sum_{i \in N'} a_r^i x_r + \sum_{r \in R_0} x_r + \sum_{r \in R_{n+1}} x_r \right) \\ &= \frac{n}{p} + \frac{1}{2p} \left(\sum_{r \in R_0} x_r + \sum_{r \in R_{n+1}} x_r \right). \end{aligned}$$

Finally, we derive a bound on the solution value of any feasible solution to the p-step formulation. Using that for any p-step $r \in R^p \setminus R_0 \cup R_{n+1}$ it holds that $c_r \geq 1$, it follows that

$$\begin{aligned} \sum_{r \in R^p} c_r x_r &= \sum_{r \in R^p \setminus (R_0 \cup R_{n+1})} c_r x_r + \sum_{r \in R_0 \cup R_{n+1}} c_r x_r \\ &\geq \sum_{r \in R^p \setminus (R_0 \cup R_{n+1})} x_r + \sum_{r \in R_0} C x_r + \sum_{r \in R_{n+1}} C x_r \\ &\geq \sum_{r \in R^p} x_r + \sum_{r \in R_0} (C-1) x_r + \sum_{r \in R_{n+1}} (C-1) x_r \\ &\geq \frac{n}{p} + \left(C - 1 + \frac{1}{2p} \right) \left(\sum_{r \in R_0} x_r + \sum_{r \in R_{n+1}} x_r \right) \\ &\geq \frac{n}{p} + \left(C - 1 + \frac{1}{2p} \right) \frac{2n}{Q} \\ &= \frac{2n}{Q} C + \frac{n(Q - 2p + 1)}{Qp}, \end{aligned}$$

where we use Lemma EC.2 to derive the last inequality. As this holds for all feasible solutions, the last expression is a lower bound on z_p.

Comparing the LP bounds of the q-step formulation and the p-step formulation, we observe

$$
\begin{aligned}
z_p - z_q \geq & \left(\frac{2n}{Q}C + \frac{n(Q-2p+1)}{Qp}\right) \\
& - \left(\frac{2n}{Q}C + \frac{n(Q-p+1)}{Qq}m\right) \\
= & \frac{n\left(Q(q-mp)+(p-1)(mp-q)-pq\right)}{Qpq} \\
= & \frac{n\left(Qk-pk-pq+k\right)}{Qpq}.
\end{aligned}
$$

Because $Q = p + \frac{pq}{k}$, we find for this instance that $z_p - z_q \geq nk/(Qpq) > 0$. This proves the claim. □

Appendix F: Proof of Proposition 5

We repeat Proposition 5 and provide a proof.

PROPOSITION 5 *For any $p, q \in \mathbb{N}_{>0}$ such that $p < q$, there exists an instance of the CVRP for which $z_p < z_q$ and an instance for which $z_p < z_q$.*

Proof First, consider any instance in which the demand of every individual customer equals the vehicle capacity. It follows that $z_p = z_q$. Next, to prove that there also exists an instance for which $z_p < z_q$, consider an instance with $n = p+1$ customers with unit demand. In this instance, the travel costs are equal to the Euclidean distances between locations. The customers are located at the vertices of a negligibly small regular convex $(p+1)$-polygon. The depot is located far away, such that the distance between the depot and each of the customers is (negligibly close to) C. Finally, the vehicle has capacity $2(p+1)$.

Next, we construct a feasible solution to the LP relaxation of the p-step formulation. Let P^t be the path starting with an arbitrary but fixed customer, visiting all other customers in order of appearance along the edges of the polygon, cyclically permuted t times, for $0 \leq t \leq p$. Furthermore, let $P(t, p-1)$ be path P^t truncated after traversing the first $p-1$ arcs. Using these paths, we now present a solution to the LP relaxation of the p-step formulation. It selects all of the following p-steps with a positive value, for all $0 \leq t \leq p$:

Type 1 $(0 \to P(t, p-1), 0)$ with total demand p,

Type 2 $(P^t, 0)$ with total demand $p+1$,

Type 3 $(P(t, p-1) \to n+1, p+2)$ with total demand p.

The p-steps of type 1 and 3 are selected with value $\frac{1}{2(p+1)}$ each, and the p-steps of type 2 are selected with value $\frac{3}{2p(p+1)}$. Similar to the proof of Proposition 4 found in Appendix E, one can verify that this is a feasible solution. Moreover, observe that the cost of a p-step of type 1, 2 and 3 is C, 0 and C respectively. As a result, the cost of the provided solution is $\frac{2(p+1)C}{2(p+1)} = C$.

Given that $q \geq p+1 = n$, all q-steps start or end at the depot and visit at most n customers. It follows that $c_r \geq C$ and that

$$\sum_{i \in N'} a_i^r < 2n$$

for all q-steps $r \in R^q$. For any solution x_r to the LP-relaxation of the q-step formulation, it follows from (2) that

$$\begin{aligned} 2n &= \sum_{i \in N'} \sum_{r \in R^q} a_i^r x_r \\ &= \sum_{r \in R^q} \sum_{i \in N'} a_i^r x_r \\ &< \frac{2n}{C} \sum_{r \in R^q} C x_r \\ &\leq \frac{2n}{C} \sum_{r \in R^q} c_r x_r . \end{aligned}$$

This shows that $z_q > C \geq z_p$. The claim follows. □

Appendix G: Proof of Proposition 6

We repeat Proposition 6 and provide a proof.

PROPOSITION 6 *The p-step formulation* (1)-(4) *and* (6)-(8) *is valid for the CVRP.*

Proof In this proof we demonstrate that an optimal solution to (1)-(4), and (6)-(8), is an optimal solution to the CVRP. It is easy to verify that every feasible solution of the CVRP corresponds with a selection of p-steps and a binary edge-flow satisfying (2)-(4), and (6)-(8). Next, we show that a feasible solution to (1)-(4) and (6)-(8) corresponds with a feasible solution of the CVRP.

By (2), (3), (6), and (8) the variables θ define an integer edge-flow in G. Similar to the proof of Proposition 1 in Appendix A, also in this case the edge-flow can be decomposed in paths from 0 to $n+1$ that respect the vehicle capacity, and cycles cannot exist because of (4). Observe that also in this case an elementary path from 0 to $n+1$ can only be formed if p-steps are selected which follow that path in the direction from 0 to $n+1$, and no p-steps in the reverse direction can be selected. This is due to (3) and the fact that no p-steps terminate at 0, or start at $n+1$. When applying here the same arguments of the proof of Proposition 1 in Appendix A, it should be noted that the p-steps are not necessarily selected binarily in this case, because a convex combination of p-steps that traverse the same partial path is allowed. However, it is sufficient that the flow over each partial path is binary. □

Appendix H: Proof of Proposition 7

We repeat Proposition 7 and provide a proof.

PROPOSITION 7 *The p-step formulation* (1)-(5) *has the same optimal solution value as* (1)-(4) *and* (6)-(8) *while limiting R^p to include only:*

1. *All p-steps r that start at 0 with prior demand $d_r = 0$.*
2. *All p-steps r that do not start at 0 nor end at $n+1$ with prior demand $d_r = 0$.*
3. *All p-steps r that do not start at 0 nor end at $n+1$ with prior demand $d_r = Q - q(r)$.*
4. *All p-steps r that do not start at 0 and end at $n+1$ with prior demand $d_r = Q - q(r)$.*

Proof Consider any solution x of the p-step formulation (1)-(5). Defining the variables θ by (6), it follows from binarity of x that the variables θ are integer. Moreover, by (2) and (3), the variables θ are binary. Thus, the solution

(x,θ) satisfies (6)-(8) as well. It remains to be shown that we can represent this solution using the reduced set of p-steps. In order to do so, consider any p-step r with $x_r = 1$. If r starts at 0 with $d_r > 0$, it could be replaced by p-step $r' = (P_r, 0)$, without violating any of the constraints, in particular not (4), and without changing the cost. Similarly, if r ends at $n+1$ while $d_r < Q - q(r)$, it can be replaced by $r' = (P_r, Q - q(r))$. Finally, if r does not start at 0 nor ends at $n+1$ and has prior demand $0 < d_r < Q - q(r)$, it can be represented by a convex combination of $r_1 = (P_r, 0)$ and $r_2 = (P_r, Q - q(r))$. For a proof of the reverse, we refer to the proof of Proposition 6. □

Appendix I: Remarks on the integrality conditions

It may seem natural to consider the following alternative to our integrality conditions. Let $\vec{b}^r_{ij}$ indicate whether arc $(i,j) \in A$ is used by p-step $r \in R$, and denote by $\vec{\theta}_{ij}$ the binary arc-flow variable indicating whether arc $(i,j) \in A$ is used. We replace the integrality conditions (6) and (8) by

$$\sum_{r \in R^p} \vec{b}^r_{ij} x_r = \vec{\theta}_{ij} \qquad \forall (i,j) \in A \tag{EC.11}$$

$$\vec{\theta}_{ij} \in \{0,1\} \qquad \forall (i,j) \in A \tag{EC.12}$$

and refer to the formulation (1)-(4), (7), (EC.11) and (EC.12) as the arc-based alternative.

A solution to the arc-based alternative may violate the 2-cycle elimination constraints

$$\vec{\theta}_{ij} + \vec{\theta}_{ji} \leq 1 \qquad \forall \{i,j\} : i,j \in N', i \neq j. \tag{EC.13}$$

To see this, consider the following example. We define an instance of the CVRP with two customers $N' = \{1,2\}$, their demands are $q_1 = q_2 = 1$, the vehicle capacity is $Q = 4$ and the arc costs are $c_{ij} = 1$ for all $(i,j) \in A$. We provide a solution to the LP relaxation of the arc-based alternative of the 1-step formulation. We represent the path corresponding to a 1-step r by a single arc a_r, so that we represent r as (a_r, d_r). We select the 1-steps $((0,1),0)$, $((2,1),1)$, and $((2,0),3)$ all with a value 0.5, $((1,2),1)$ with value 1, and set $\vec{\theta}_{01} = \vec{\theta}_{21} = \vec{\theta}_{20} = 0.5$ and $\vec{\theta}_{12} = 1$. This solution is feasible and has cost 2.5. This is lower than 3, which is the optimal solution value of the LP relaxation of the 1-step formulation, showing that the LP bound of the arc-based alternative is lower for this instance. In particular, this solution violates the 2-cycle elimination constraint $\vec{\theta}_{12} + \vec{\theta}_{21} \leq 1$. Adding the 2-cycle elimination constraints to the arc-based alternative yields a formulation that has the same LP bound as the p-step formulation as shown next.

PROPOSITION EC.1. *The LP bound of the arc-based alternative with 2-cycle elimination is equal to the LP bound of the p-step formulation.*

Proof Consider a solution $\left(x, \vec{\theta}\right)$ to the LP relaxation of the arc based alternative with 2-cycle elimination. By defining $\theta_{ij} = \vec{\theta}_{ij} + \vec{\theta}_{ji}$ for all $\{i,j\} \in E$, we obtain a solution (x,θ) to the LP relaxation of the p-step formulation with the same objective value. To verify feasibility, observe that (1)-(4) and (7) are trivially satisfied. Observe that because $\theta_{ij} = \vec{\theta}_{ij} + \vec{\theta}_{ji}$ it follows trivially that $\theta_{ij} \geq 0$ and by the 2-cycle elimination constraints (EC.13) that $\theta_{ij} \leq 1$. Furthermore,

$$
\begin{aligned}
\sum_{r \in R^p} b_{ij}^r x_r &= \sum_{r \in R^p} \left(\vec{b}_{ij}^r + \vec{b}_{ji}^r \right) x_r \\
&= \sum_{r \in R^p} \vec{b}_{ij}^r x_r + \sum_{r \in R^p} \vec{b}_{ji}^r x_r \\
&= \vec{\theta}_{ij} + \vec{\theta}_{ji} \\
&= \theta_{ij},
\end{aligned}
$$

showing that (6) is satisfied. We conclude that (x, θ) is a feasible solution to the LP relaxation of the p-step formulation.

Now consider a solution (x, θ) to the LP relaxation of the p-step formulation. By defining $\vec{\theta}_{ij} = \sum_{r \in R^p} \vec{b}_{ij}^r x_r$ for all $(i, j) \in A$, we obtain a solution $\left(x, \vec{\theta}\right)$ to the LP relaxation of the arc-based alternative with 2-cycle elimination with the same objective value. To verify feasibility, observe that (1)-(4) and (7) are trivially satisfied. Furthermore,

$$
\begin{aligned}
\vec{\theta}_{ij} + \vec{\theta}_{ji} &= \sum_{r \in R^p} \vec{b}_{ij}^r x_r + \sum_{r \in R^p} \vec{b}_{ji}^r x_r \\
&= \sum_{r \in R^p} \left(\vec{b}_{ij}^r + \vec{b}_{ji}^r \right) x_r \\
&= \sum_{r \in R^p} b_{ij}^r x_r \\
&= \theta_{ij} \\
&\leq 1,
\end{aligned}
$$

showing that the 2-cycle elimination constraints (EC.13) are satisfied. Observe that it follows by definition that $\vec{\theta}_{ij} \geq 0$, and by the preceding that $\vec{\theta}_{ij} \leq 1$. We conclude that $\left(x, \vec{\theta}\right)$ is a feasible solution to the LP relaxation of the arc-based alternative with 2-cycle elimination. Therefore, it follows that the LP bound of the arc-based alternative with 2-cycle elimination is equal to the LP bound of the p-step formulation. □

Compared to the p-step formulation, the arc-based alternative with 2-cycle elimination uses more variables and more constraints, and has the same LP bound.

Appendix J: Proof of Proposition 8

In this appendix, we provide a proof of Proposition 8, which states that computing any lower bound of the CVRP that is larger than or equal to the set partitioning bound is strongly NP-hard. The set partitioning formulation can be found in Appendix C. Before proving Proposition 8, we first show an intermediate result.

LEMMA EC.3. *Either there exists an optimal solution to the LP relaxation of SPF in which a single route $r \in R$ that visits all customers in N' is selected with value $x_r = 1$ and all other routes are not selected, i.e., $x_{r'} = 0$ for all $r' \in R$, $r' \neq r$, or such an optimal solution does not exist. In the latter case, every optimal solution is such that $\sum_{r \in R} x_r \geq \frac{n}{n-1}$.*

Proof We have to prove that $\sum_{r \in R} x_r \geq \frac{n}{n-1}$ if there is no optimal solution in which a single route $r \in R$ that visits all customers in N' is selected with value $x_r = 1$ and all other routes are not selected. Therefore, assume that such an optimal solution does not exist and consider any optimal solution to the LP relaxation of SPF. Assume first that this solution is such that a route $r \in R$ that visits all customers N' is selected with value $0 < x_r < 1$. We could then create a new solution $\hat{x}$ to the LP relaxation of SPF such that $\hat{x}_r = 1$ and $\hat{x}_{r'} = 0$ for all $r' \in R$, $r' \neq r$. Observe that this new solution is feasible. Furthermore, the solution value of $\hat{x}$ is the same as that of x, as otherwise optimality is contradicted. Therefore, the new solution is also optimal. This contradicts our assumption that such a solution does not exist. It follows that all routes $r \in R$ that are selected with value $x_r > 0$ visit at most $n-1$ customers in N'. By summing over all $i \in N'$, we derive from (EC.9)

$$n = \sum_{i \in N'} \sum_{r \in R} \alpha_r^i x_r = \sum_{r \in R} \sum_{i \in N'} \alpha_r^i x_r \leq \sum_{r \in R} (n-1) x_r.$$

It follows that $\sum_{r \in R} x_r \geq \frac{n}{n-1}$. □

Next, we repeat Proposition 8 and provide a proof.

PROPOSITION 8 *Computing a lower bound on the CVRP that is larger than or equal to the set partitioning bound is strongly NP-hard. Thus, no polynomial or pseudo-polynomial time algorithm exists that provides a lower bound for the CVRP which is at least the set partitioning bound, unless $P = NP$.*

Proof We provide a polynomial time reduction of the Hamiltonian cycle problem, which is known to be NP-complete (Karp 1972). We consider the version of this problem problem on a directed graph $\hat{G} = (\hat{N}, \hat{A})$, and denote $\hat{n} = |\hat{N}|$. The Hamiltonian cycle problem is to determine whether there exists a cycle that includes every node in $\hat{N}$ exactly once, yes or no.

We create an instance of the CVRP as follows. We create a complete graph $G = (N, A)$. The nodes N correspond with $\hat{N}$, where we arbitrarily select one node from $\hat{N}$ to act as starting depot 0 and ending depot $n+1$. Note that per definition of the CVRP, we maintain $A = \{(i, j) : i, j \in N, i \neq j, i \neq n+1, j \neq 0\}$. It follows that every arc in $\hat{A}$ corresponds with an arc in A. For every arc $(i, j) \in A$ such that $i \neq 0$, we set $c_{ij} = 1$ if $(i, j) \in \hat{A}$ and $c_{ij} = 2$ otherwise. For each arc $(0, j) \in A$, we set $c_{0j} = \hat{n}^2 + 1$ if $(0, j) \in \hat{A}$ and $c_{0j} = \hat{n}^2 + 2$ otherwise. Additionally we define the demand of each customer node $i \in N'$ as $q_i = 1$ and set the vehicle capacity equal to $\hat{n} - 1$, such that any route visiting all nodes is feasible.

For an instance of the Hamiltonian cycle problem to which the answer is yes, a route $r \in R$ exists which visits all customers with cost $c_r = \hat{n}^2 + \hat{n}$. From Lemma EC.3 it follows that either it is optimal to select only this route r at cost $\hat{n}^2 + \hat{n}$, or the LP bound of the set partitioning formulation is at least $(\hat{n}^2 + 1)\frac{\hat{n}-1}{\hat{n}-2}$, where we note that $n = |N'| = \hat{n} - 1$. Because $(\hat{n}^2 + 1)\frac{\hat{n}-1}{\hat{n}-2} = \hat{n}^2 + \hat{n} + 3 + \frac{5}{\hat{n}-2} > \hat{n}^2 + \hat{n}$, it follows that if the answer to the Hamiltonian cycle problem is yes, the optimal solution value of the CVRP and the set partitioning bound are equal to $\hat{n}^2 + \hat{n}$. In particular, any bound that is larger than or equal to the set partitioning bound is equal to $\hat{n}^2 + \hat{n}$.

For an instance of the Hamiltonian cycle problem to which the answer is no, every route which visits all customers has a cost of at least $\hat{n}^2 + \hat{n} + 1$. Applying Lemma EC.3 similar to before, it follows that either it is optimal to select only such a route r at a cost of at least $\hat{n}^2 + \hat{n} + 1$, or the LP bound is at least $(\hat{n}^2 + 1)\frac{\hat{n}-1}{\hat{n}-2}$. We conclude that if the answer to the Hamiltonian cycle problem is no, the set partitioning bound is at least $\hat{n}^2 + \hat{n} + 1$.

This shows that the Hamiltonian cycle problem can be solved by computing any lower bound that is larger than or equal to the set partitioning bound of the CVRP. Indeed if the lower bound is $\hat{n}^2+\hat{n}$, the answer is yes, and otherwise the answer is no. Since we have demonstrated a polynomial time reduction of the Hamiltonian cycle problem which is NP-complete, and because the largest number used to represent the instance of the CVRP is polynomial in the size of the instance, we conclude that computing any lower bound that is larger than or equal to the set partitioning bound of the CVRP is strongly NP-hard. □

Appendix K: Proof of Proposition 9

We repeat Proposition 9 and provide a proof.

PROPOSITION 9 *If $P \neq NP$, there does not exist a strongest compact or pseudo-compact formulation of the CVRP.*

Proof A compact formulation of the CVRP can be used to compute a lower bound for the CVRP in polynomial time. By Proposition 8, unless $P = NP$, no compact formulation exists that provides at least the set partitioning bound. Hence, for any given compact formulation, at least one instance of the CVRP exists for which the corresponding LP bound is strictly less than the set partitioning bound. Let n_I be the number of customers of such an instance I.

Consider a p-step formulation, for a fixed value of p such that $p \geq n_I + 1$. This formulation provides the set partitioning bound for instance I. Hence, it is a compact formulation that is not weaker, and it provides a stronger LP bound for at least one instance. Since this can be done for any compact formulation, we conclude that if $P \neq NP$, there does not exist a strongest compact formulation. By a similar argument, there also does not exist a strongest pseudo-compact formulation. □

Appendix L: Constructing the new strongest known compact formulations of the CVRP

In this appendix, we provide some details on constructing the new strongest known compact formulations of the CVRP, with the number of variables and constraints limited by a polynomial of fixed degree three and higher, as introduced in Section 3.4. For brevity, by Constraints(F3) we denote the set of constraints of the formulation F3 as found in Leggieri and Haouari (2017). In that paper, x_{ij} is the binary variable indicating whether arc $(i,j) \in A$ is selected in the solution or not. The objective function of F3 is $\sum_{(i,j)\in A} c_{ij}x_{ij}$. We additionally introduce the binary parameters $\vec{b}^r_{ij}$ indicating whether arc $(i,j) \in A$ is used by k-step $r \in R^k$. The formulation in which we combine F3 with the 1-step through p-step formulations is as follows.

$$\min \sum_{(i,j)\in A} c_{ij} x_{ij} \tag{EC.14}$$

Constraints(F3)

$$\sum_{r\in R^k} \vec{b}^r_{ij} x_r = x_{ij} \qquad \forall (i,j)\in A,\ \forall k\in\{1,\dots,p\} \tag{EC.15}$$

$$\sum_{r\in R^k} a^i_r x_r = 2 \qquad \forall i\in N',\ \forall k\in\{1,\dots,p\} \tag{EC.16}$$

$$\sum_{r\in R^k} e^i_r x_r = 0 \qquad \forall i\in N',\ \forall k\in\{1,\dots,p\} \tag{EC.17}$$

$$\sum_{r\in R^k} q^i_r x_r \geq 0 \qquad \forall i\in N',\ \forall k\in\{1,\dots,p\} \tag{EC.18}$$

$$\sum_{r\in R^k} b^r_{ij} x_r = \theta_{ij} \qquad \forall \{i,j\}\in E,\ \forall k\in\{1,\dots,p\} \tag{EC.19}$$

$$x_r \geq 0 \qquad \forall r\in R^k,\ \forall k\in\{1,\dots,p\} \tag{EC.20}$$

$$\theta_{ij} \in \{0,1\} \qquad \forall \{i,j\}\in E. \tag{EC.21}$$

The objective is modeled by (EC.14), and the constraints of F3 are included. Constraints (EC.15) link the arc-flow variables of F3 to the arc-flow defined by the p-steps. Constraints (EC.16)-(EC.21) are the equivalent to (2)-(4) and (6)-(8) for the different parameter values from 1 through p, using a single set of variables θ_{ij} for all $\{i,j\}\in E$ in (EC.19).