

PROBABILISTIC ANALYSIS OF UNIT-DEMAND VEHICLE ROUTEING PROBLEMS

AGUSTÍN BOMPADRE* AND

MOSHE DROR,** *Operations Research Center, MIT*

JAMES B. ORLIN,*** *Sloan School of Management, MIT*

Abstract

We analyze the unit-demand Euclidean vehicle routeing problem, where n customers are modeled as independent, identically distributed uniform points and have unit demand. We show new lower bounds on the optimal cost for the metric vehicle routeing problem and analyze them in this setting. We prove that there exists a constant $\hat{c} > 0$ such that the iterated tour partitioning heuristic given by Haimovich and Rinnooy Kan (1985) is a $(2 - \hat{c})$ -approximation algorithm with probability arbitrarily close to 1 as the number of customers goes to ∞ . It has been a longstanding open problem as to whether one can improve upon the factor of 2 given by Haimovich and Rinnooy Kan. We also generalize this, and previous results, to the multidepot case.

Keywords: Vehicle routeing problem; heuristic; probabilistic analysis of algorithms

2000 Mathematics Subject Classification: Primary 68W40

Secondary 90B06

1. Introduction

1.1. Unit-demand vehicle routeing problem

We study the unit-demand Euclidean vehicle routeing problem (VRP), where the customers, x_1, \dots, x_n , and the depot, y_1 , are given as points in the plane and the distance between points is the Euclidean distance. Each customer has unit demand. There are an unlimited number of identical vehicles, and each one has capacity $Q \in \mathbb{N}$. The route of each vehicle starts and ends at the depot, y_1 . Each vehicle cannot deliver more than its capacity. The cost of a solution is the sum of the traversing costs of each vehicle. In the problems we consider the objective is to route vehicles to meet the demand of every customer while minimizing the overall cost.

Except in some special cases, the VRP is an NP-hard problem. In their seminal paper [8], Haimovich and Rinnooy Kan provided a worst-case analysis and a probabilistic analysis of the VRP (see also [9]). In [8] they proved a lower bound on the cost of the VRP with metric distance which is the maximum between the cost of a traveling salesman problem (TSP) tour and the so-called radial cost. When Q is fixed, the unit-demand Euclidean VRP admits a polynomial-time approximation scheme. This means that for any $\varepsilon > 0$ there exists a $(1 + \varepsilon)$ -approximation algorithm. The first polynomial-time approximation scheme for this

Received 14 June 2005; revision received 31 October 2006.

* Current address: Optiant Inc., 4 Van de Graaff Drive, Burlington, MA 01803, USA.

Email address: abompadr@alum.mit.edu

** Current address: Department of Management Information Systems, Eller College of Management, University of Arizona, 430M McClelland Hall, 1130 E. Helen Street, PO Box 210108, Tucson, AZ 85721-0108, USA.

*** Postal address: Sloan School of Management, MIT, 50 Memorial Drive, Cambridge, MA 02142, USA.

case appeared in [8]. Subsequently, Asano *et al.* [3] improved its running time using the polynomial-time approximation scheme for the Euclidean TSP (see [2] or [11]).

In [8] the authors also analyzed the problem from a probabilistic point of view when the locations of the customers are modeled as independent, identically distributed (i.i.d.) points in the plane and Q is a function of the number of customers. Their analysis showed that the lower bound they propose is asymptotically optimal when $Q = o(\sqrt{n})$ or $n = o(Q^2)$. As a result, the iterated tour partitioning (ITP) heuristic of [8] becomes asymptotically optimal in both cases. In the rest of the cases, the ITP heuristic is within a factor of 2 of the optimal cost. That is, under no assumptions on Q as the number of customers goes to ∞ , the probabilistic analysis of the ITP heuristic is no better than its worst-case analysis.

In this paper we show that the ITP heuristic is indeed better on average than is suggested by its worst-case analysis, by proving that its approximation ratio is strictly smaller than 2. The primary contributions of this paper are as follows.

1. We improve the approximation ratio of the ITP heuristic of [8] for the VRP. This improvement, though slight, does resolve a longstanding open question of whether any improvement was possible.
2. We provide nonlinear inequalities (Lemma 3) that are useful for improving bounds for VRP problems.
3. We show that, for n points uniformly distributed in the unit square and for every p with $0 < p \leq 1$, there is a constant $c(p)$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\text{at least } pn \text{ points have a neighbor at distance less than or equal to } \frac{c(p)}{\sqrt{n}} \right) = 0.$$

This result implies nontrivial lower bounds on the cost of combinatorial optimization problems such as the minimum latency problem.

4. We extend the probabilistic analysis of VRP to the multidepot case.

To be more precise, we first improve the radial cost lower bound on the cost of the metric VRP given by [8]. The improvement we present is at least a fraction of \sqrt{n} with probability converging to 1 as n tends to ∞ , when the n customers are given as uniformly distributed points in the square $[0, 1]^2$ and the distance is the Euclidean distance. As a result, the approximation ratio of the ITP heuristic is strictly better than 2 with probability 1 as the number of customers goes to ∞ . As a side effect, our analysis also shows a further improvement on the approximation ratio when both $Q = O(\sqrt{n})$ and $\sqrt{n} = O(Q)$. This parametric case is the border line between the cases analyzed by Haimovich and Rinnooy Kan [8].

In the second part of the paper, the results proven for the unit-demand Euclidean VRP are generalized to the multidepot case where the number and location of the depots is fixed in advance. Related works are [10] and [13]. Li and Simchi-Levi [10] performed a worst-case analysis of the multidepot VRP when the distances satisfy the triangle inequality, and showed how to reduce this problem to a single-depot case with distances satisfying the triangle inequality. Their analysis is valid for metric spaces and, thus, cannot take advantage of the properties of the Euclidean metric in the plane. Stougie [13] studied a two-stage multidepot problem where, at the first stage, the central planner decides how many depots to build (and where to build them) and, at the second stage, deals with a multidepot VRP. Stougie's analysis was probabilistic, since the customers of the second stage are i.i.d. points in the unit square.

The objective in this problem is to minimize the sum of the costs of both stages. Unlike in our approach, the number and locations of depots are variables. Stougie’s approach does not translate well to our context, since the optimal solution of the first stage simplifies the analysis of the second stage problem. We introduce a natural generalization of the ITP heuristic to the multidepot case and show that the results proved for the single-depot case carry through.

The rest of the paper is organized as follows. In Section 2 we present both new and known lower bounds on the optimal cost of the VRP. In Section 3 we analyze the value of the lower bound in a probabilistic setting. We present lower and upper bounds on the cost of combinatorial optimization problems such as the minimum latency problem. We also prove our main results about the approximation ratio of the ITP heuristic. We discuss the multi-depot VRP (MDVRP) in Section 4. In Section 5 we present an algorithm for the MDVRP that generalizes the ITP heuristic. In Section 6 we present lower bounds for the MDVRP. In Section 7 we analyze the lower bounds and the algorithm for MDVRP in a probabilistic setting. We summarize the results presented in the paper in Section 8.

1.2. Notation

Unless otherwise stated, the customers and the depot are modeled as points in the plane. The location of the i th customer is denoted by x_i , for any $1 \leq i \leq n$. The depot is located at y_1 . The set of customers is denoted by $X^{(n)}$. The distance between customers i and j is denoted by $c_{i,j}$ or by c_{x_i,x_j} , and the distance between a customer i and depot y_1 is denoted by $c_{y_1,i}$. A solution to a VRP is denoted by a pair (K, V) , where K is the number of vehicles used and $V = \{v_k : k \in \{1, \dots, K\}\}$ is the set of vehicles (or routes). Given a solution (K, V) , we let $d_i^k = 1$ when vehicle $v_k \in V$, visits customer i and let $d_i^k = 0$ otherwise. The routing cost of vehicle v_k is denoted by $c(v_k)$. By $R = \sum_{i=1}^n 2c_{y_1,i} / Q$ we denote the so-called radial cost, as in [8]. The radial cost is a lower bound on the cost of the VRP; see [8] or Lemma 3 below. We denote by $c(\text{VRP})$ or by $c(\text{VRP}(X^{(n)}))$ the cost of an optimal VRP, and we denote by $c(\text{TSP})$ or $c(\text{TSP}(X^{(n)}))$ the cost of an optimal TSP tour. Given a probability space and a probability measure, the probability that event A occurs is denoted by $P(A)$. The probability of event A conditioned on event B is denoted by $P(A \mid B)$. The complement of event A is denoted by \bar{A} . We use upper-case letters (e.g. X) to denote random variables and lower-case letters (e.g. x) to denote realizations of random variables.

For a real number a , we denote by $\lfloor a \rfloor = \sup\{k \in \mathbb{Z} : k \leq a\}$ and $\lceil a \rceil = \inf\{k \in \mathbb{Z} : k \geq a\}$ its floor and ceiling values, respectively. For a finite set A , we denote by $|A|$ its cardinality. For two real-valued functions f and g , we say that $f(n) = O(g(n))$ if, for some numbers $c > 0$ and n_0 , $f(n) \leq cg(n)$ for all $n \geq n_0$. Similarly, we say that $f(n) = \Omega(g(n))$ if, for some numbers $c > 0$ and n_0 , $f(n) \geq cg(n)$ for all $n \geq n_0$. Finally, we say that $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

2. Lower bounds on the optimal cost of the VRP

We assume in this section that all distances satisfy the triangle inequality. We refer to such problems as ‘metric’. In this section we present both new and known lower bounds on the cost of a metric VRP. The main result of this section is summarized in Lemma 5, which improves the lower bound of [8]. We here state this lower bound on the cost of a metric VRP.

Lemma 1. ([8].) *The cost of the metric VRP is at least $\max\{R, c(\text{TSP})\}$.*

The ITP heuristic $\text{ITP}(\alpha)$ for the unit-demand VRP defined in [8] (see also [1]) receives an α -optimal TSP as part of its input and outputs a solution with cost at most $R + \alpha(1 - 1/Q)c(\text{TSP})$.

We briefly describe this heuristic. It receives a tour $(y_1, i_1, i_2, \dots, i_n, y_1)$ of cost at most $\alpha c(\text{TSP})$ as part of the input, and selects the best valid route among Q solutions constructed as follows. For each $t, 1 \leq t \leq Q$, the solution, route_t , to the VRP is the union of the $K_t = \lceil (n - t)/Q \rceil + 1$ routes

$$\begin{aligned} v_1^t &= (y_1, i_1, \dots, i_t, y_1), \\ v_2^t &= (y_1, i_{t+1}, \dots, i_{t+Q}, y_1), \\ v_3^t &= (y_1, i_{t+Q+1}, \dots, i_{t+2Q}, y_1), \\ &\vdots \\ v_{\lceil (n-t)/Q \rceil + 1}^t &= (y_1, i_{(\lceil (n-t)/Q \rceil - 1)Q + t + 1}, \dots, i_n, y_1). \end{aligned}$$

That is, route_t transforms the original tour, $(y_1, i_1, \dots, j_n, y_1)$, into routes with Q customers each (except possibly the first and the last routes). The average cost of these solutions is at most $R + (1 - 1/Q)\alpha c(\text{TSP})$ and, therefore, one of them has cost at most $R + (1 - 1/Q)\alpha c(\text{TSP})$. By Lemma 1, $\max\{R, c(\text{TSP})\}$ is a lower bound on the optimal cost of the VRP, and one of the solutions considered route_t is therefore within $1 + (1 - 1/Q)\alpha$ times the optimal cost.

In the Euclidean setting there exists a polynomial-time approximation scheme for the TSP (see [2] or [11]). Therefore, the ITP heuristic is a $(2 - 1/Q)$ -approximation algorithm in this case. We denote by $c(\text{VRP}^{\text{ITP}})$ the cost of the solution generated by the ITP heuristic when it receives an optimal TSP as part of the input. When Q is not fixed, that is, it is part of the input, the approximation ratio is asymptotically 2, as was shown for a family of examples in [10]. We state this result as a lemma.

Lemma 2. ([10].) *The ITP heuristic ITP(1) is a 2-approximation algorithm.*

Corollary 2 of [6] presents a lower bound on the cost of the VRP. The impact of this lower bound on the worst-case analysis of the VRP is discussed in that paper. In Lemma 3 below we state a stronger version of this bound. The new lower bound is more suitable for the probabilistic analysis we perform in this paper.

Definition 1. Given a route $v_k = (y_1, i_1^k, i_2^k, \dots, i_{s_k}^k, y_1)$ that starts and ends at depot y_1 , we associate with it the sequence of customers $l^k = (i_1^k, i_2^k, \dots, i_{s_k}^k)$. For any customers i and j visited by v_k , let l_{ij}^k be the length of the unique path from i to j determined by v_k that does not pass through the depot y_1 . If i or j is not visited by v_k , we let $l_{ij}^k = 0$. Given a solution (K, V) to the VRP, its associated sequences l^1, \dots, l^K are the sequences associated with $v_1, \dots, v_K \in V$, respectively.

Since the distance matrix is symmetric, $l_{ij}^k = l_{ji}^k$. For example, let

$$v_k = (y_1, 1, 2, 3, 4, 5, 6, y_1)$$

with $c_{y_1,1} = c_{y_1,6} = c_{s,s+1} = 2$ for all $s, 1 \leq s \leq 6$. In this example $l_{2,6}^k = l_{6,2}^k = 8$.

Lemma 3. *Let the pair (K, V) be a solution to the metric VRP. For each $k, 1 \leq k \leq K$, and $i, 1 \leq i \leq n$, let the value of $d_i^k \in \{0, 1\}$ determine whether or not vehicle $v_k \in V$ visits customer i . The cost of (K, V) is at least*

$$\sum_{k=1}^K \sum_{i=1}^n 2c_{y_1,i} \frac{d_i^k}{\sum_{t=1}^n d_t^k} + \sum_{k=1}^K \sum_{i,j \in \{1, \dots, n\}} l_{ij}^k \frac{d_i^k d_j^k}{(\sum_{t=1}^n d_t^k)^2}. \tag{1}$$

Proof. To simplify our notation, let us define $z_i^k = d_i^k / \sum_{t=1}^n d_t^k$. Since $\sum_{j=1}^n z_j^k = 1$, we can rewrite (1) as

$$\begin{aligned} \sum_{k=1}^K \sum_{i=1}^n 2c_{y_1,i} z_i^k + \sum_{k=1}^K \sum_{i,j \in \{1,\dots,n\}} l_{ij}^k z_i^k z_j^k &= \sum_{k=1}^K \sum_{i=1}^n 2c_{y_1,i} z_i^k \sum_{j=1}^n z_j^k + \sum_{k=1}^K \sum_{i,j \in \{1,\dots,n\}} l_{ij}^k z_i^k z_j^k \\ &= \sum_{k=1}^K \sum_{i,j \in \{1,\dots,n\}} 2c_{y_1,i} z_i^k z_j^k + \sum_{k=1}^K \sum_{i,j \in \{1,\dots,n\}} l_{ij}^k z_i^k z_j^k \\ &= \sum_{k=1}^K \sum_{i,j \in \{1,\dots,n\}} (c_{y_1,i} + l_{ij}^k + c_{y_1,j}) z_i^k z_j^k \\ &\leq \sum_{k=1}^K \sum_{i,j \in \{1,\dots,n\}} c(v_k) z_i^k z_j^k. \end{aligned}$$

The last inequality holds because if $z_i^k z_j^k = 0$ then $(c_{y_1,i} + l_{ij}^k + c_{y_1,j}) z_i^k z_j^k = c(v_k) z_i^k z_j^k = 0$; otherwise, v_k starts and ends at depot y_1 and visits customers i and j , and, thus,

$$(c_{y_1,i} + l_{ij}^k + c_{y_1,j}) z_i^k z_j^k \leq c(v_k) z_i^k z_j^k$$

by the triangle inequality. We use the equality $\sum_{j=1}^n z_j^k = 1$ again, to simplify the last expression:

$$\sum_{k=1}^K \sum_{i,j \in \{1,\dots,n\}} c(v_k) z_i^k z_j^k = \sum_{k=1}^K c(v_k) \sum_{i=1}^n z_i^k \sum_{j=1}^n z_j^k = \sum_{k=1}^K c(v_k).$$

Remark 1. Let the pair (K, V) be a solution to the metric VRP, and let the value of $d_i^k \in \{0, 1\}$ determine whether or not vehicle $v_k \in V$ visits customer i . Therefore, since any vehicle v_k delivers $\sum_{i=1}^n d_i^k \leq Q$ units of demand, the term $\sum_{k=1}^K \sum_{i=1}^n 2c_{y_1,i} d_i^k / \sum_{i=1}^n d_i^k$ is at least the radial cost, $R = \sum_{i=1}^n 2c_{y_1,i} / Q$.

The lower bound on the cost of a solution to the metric VRP given by Lemma 3 contains two sums. One sum is at least the radial cost, as observed in Remark 1. The second sum is an improvement over the radial cost. This improvement will be analyzed in a probabilistic setting in the next section. We give some definitions first.

Definition 2. Given a solution (K, V) to a VRP, a vehicle v_k is said to be half-full if it visits at least $Q/2$ customers. Let $A \subseteq X^{(n)}$ be the set of customers visited by half-full vehicles. A solution is said to have the fullness property if $|A| \geq n - Q/2$.

The following lemma states that there always exists an optimal solution that has the fullness property.

Lemma 4. *There exists an optimal solution (K, V) such that $|A| \geq n - Q/2$.*

Proof. Let (K, V) be an optimal solution such that the associated set A has maximal cardinality. Either there is at most one vehicle that visits at most $Q/2$ of the customers or there are at least two. In the first case $|A| \geq n - Q/2$. In the second case we can replace two vehicles that visit at most $Q/2$ of the customers by one vehicle without increasing the routing cost. In this case we find an optimal solution with an associated set, A' , bigger than A , contradicting the maximality of A .

We give a name to the quadratic term in (1).

Definition 3. Given a solution (K, V) , let

$$QC(K, V) = \sum_{k=1}^K \sum_{i,j \in \{1, \dots, n\}} l_{ij}^k \frac{d_i^k d_j^k}{(\sum_{t=1}^n d_t^k)^2}. \tag{2}$$

Let QC be the maximum value of $QC(K, V)$ among all optimal solutions (K, V) that have the fullness property of Definition 2.

It is apparent that the radial cost plus the quadratic term QC is a lower bound on the cost of a solution (K, V) that has the fullness property. Therefore, the following lemma holds.

Lemma 5. *The cost of the VRP is at least $\max\{R + QC, c(\text{TSP})\}$.*

3. Probabilistic analysis

The main technical result of this section is Lemma 7, which shows that the quadratic term (2) is $\Omega(\sqrt{n})$ in a probabilistic model. This in turn implies our main result, Theorem 2, which states that the approximation ratio of the ITP heuristic is strictly better than 2 in a probabilistic model.

We start by giving the classic result of [4] concerning the asymptotic behavior of the cost of the TSP. Let X_1, X_2, \dots be a sequence of i.i.d. points having a distribution on $[0, 1]^2$. That is, the x and y coordinates of each X_i have a joint distribution on the square $[0, 1]^2$. Let f be the density of the absolutely continuous part of the distribution of each X_i . There exists a constant $\beta > 0$, independent of the distribution of the X_i , such that with probability 1 the cost of an optimal subtour through the first n points satisfies

$$\lim_{n \rightarrow \infty} \frac{c(\text{TSP}(X^{(n)}))}{\sqrt{n}} = \beta \int f^{1/2} dx. \tag{3}$$

Computing the exact value of β is an open problem. However, in [4] it was shown that $0.62 \leq \beta \leq 0.93$.

From now on we assume that the customers X_1, \dots, X_n are independent random variables with distribution $U[0, 1]^2$. Although the results proven in this paper also hold for more general random variables, the restriction to uniform random variables is made in order to simplify our proofs. In this case the ratio of the cost of the TSP to \sqrt{n} converges almost surely (a.s.) to β . The following theorem was proved in [8]. Informally, the result is that the radial cost dominates the cost of the TSP when $Q = o(\sqrt{n})$, and that the opposite happens when $n = o(Q^2)$. Therefore, the ITP heuristic is asymptotically optimal in both cases.

Theorem 1. ([8].) *Let X_1, X_2, \dots be a sequence of i.i.d. uniform random points in $[0, 1]^2$, and let $X^{(n)}$ denote the first n points of this sequence. Let μ denote the expected distance between X_1 and the depot.*

- *If $\lim_{n \rightarrow \infty} Q/\sqrt{n} = 0$ then*

$$\lim_{n \rightarrow \infty} \frac{c(\text{VRP}(X^{(n)}))Q}{n} = 2\mu \quad \text{a.s.}$$

- If $\lim_{n \rightarrow \infty} Q/\sqrt{n} = \infty$ then

$$\lim_{n \rightarrow \infty} \frac{c(\text{VRP}(X^{(n)}))}{\sqrt{n}} = \beta \quad \text{a.s.}$$

where $\beta > 0$ is the constant introduced in [4].

This theorem implies that the ITP heuristic is asymptotically optimal whenever

$$\lim_{n \rightarrow \infty} \frac{Q}{\sqrt{n}} = 0 \quad \text{or} \quad \lim_{n \rightarrow \infty} \frac{Q}{\sqrt{n}} = \infty.$$

Corollary 1. *If $\lim_{n \rightarrow \infty} Q/\sqrt{n} = 0$ or $\lim_{n \rightarrow \infty} Q/\sqrt{n} = \infty$, then $c(\text{VRP}^{\text{ITP}})/c(\text{VRP}) = 1$ a.s.*

In order to prove that the ITP heuristic is strictly better than a 2-approximation algorithm on average, we have to prove that the quadratic term (2) has nonnegligible growth as a function of n . Informally, the quadratic term takes into account part of the distance between customers that is neglected by the radial cost. This cost is related to the distance of a generic customer i to its closest neighbor $j \neq i$. If we define a threshold value and consider a customer to be ‘isolated’ whenever its distance to its closest neighbor is at least the threshold value, then the following lemma states that we can define the threshold as a function of a parameter p in order to guarantee that the proportion of isolated customers is at least $1 - p$ with probability approaching 1 as n goes to ∞ . This lemma is critical in proving that the quadratic term (2) is $\Omega(\sqrt{n})$ with probability 1.

Lemma 6. *For any p , $0 < p \leq 1$, there exists a value $c(p) > 0$ such that*

$$\lim_{n \rightarrow \infty} \text{P}\left(\text{at least } pn \text{ customers have a neighbor at distance less than or equal to } \frac{c(p)}{\sqrt{n}}\right) = 0. \tag{4}$$

Proof. First divide the unit square into $M^2 = \alpha pn$ subsquares, each with a side whose length is $1/\sqrt{\alpha pn}$. We assume that αpn is the square of an integer. This assumption is used to simplify the proof, but the proof carries through even if αpn is not the square of an integer if M is defined as the (unique) integer such that $M^2 \leq \alpha pn < (M + 1)^2$. We will soon choose α as a function of p .

Suppose that n points are distributed at random, corresponding to the selection (with replacement) of n squares at random. We will refer to a selected square as being isolated if no other selected square is a neighbor. (Neighbors include the square and its eight adjacent squares.) Otherwise, selected squares are said to be nonisolated. Then,

$$\begin{aligned} & \text{P}\left(\text{at least } pn \text{ customers have a neighbor at a distance less than or equal to } \frac{1}{\sqrt{\alpha pn}}\right) \\ & \leq \text{P}(\text{at least } pn \text{ selected squares have a selected neighbor}) \\ & = \text{P}(\text{at least } pn \text{ selected squares are nonisolated}). \end{aligned}$$

Suppose that we select points randomly one at a time. As the k th point is randomly selected, the probability that it is adjacent to a point that has already been randomly selected is at most $9(k - 1)/\alpha pn$, which is bounded from above by $9n/\alpha pn = 9/\alpha p$. This upper bound is independent of the locations of the first $k - 1$ points. Therefore, the total number of nonisolated

points is bounded by $2X$, where X is the number of successes in n Bernoulli experiments with success probability $9/\alpha p$. The factor of 2 comes from the fact that if point k is adjacent to a previously selected point $i < k$, then i is also adjacent to k . Thus,

$$\begin{aligned} &P(\text{at least } pn \text{ selected squares are nonisolated}) \\ &\leq P(\text{there are at least } pn/2 \text{ successes out of } n \text{ with probability of success of } 9/\alpha p). \end{aligned}$$

Therefore, if we choose α such that $18/\alpha p < p$ (that is, $\alpha > 18/p^2$), then, by the weak law of large numbers, $\lim_{n \rightarrow \infty} P(\text{at least } pn \text{ selected squares are nonisolated}) = 0$. Therefore, any value of $c(p)$ smaller than $\sqrt{p}/3\sqrt{2}$ satisfies (4).

This lemma is similar in style to known bounds on the minimum distance between a fixed point and n i.i.d. points. It differs from them in the fact that it is an upper bound on the probability of a set of dependent events defined from dependent random variables (the minimum distances among the n points). We will use it to prove a lower bound on the quadratic term (2). It also provides nontrivial lower bounds on the cost of combinatorial optimization problems such as minimum weighted matching, the TSP, and the minimum latency problem, as the next corollary shows. We will invest some time in analyzing the minimum latency problem, since the authors are not aware of asymptotic results for this problem.

The minimum latency problem (see, e.g. [5]) is to find a tour of minimum latency through all customers. The latency of a customer i_k , denoted by c_{i_k} , is the time taken to arrive at i_k . The latency of a tour is the sum of the waiting times of all customers. That is, given a solution $(y_1, i_1, i_2, \dots, i_n, y_1)$ of the latency problem, its cost is

$$\sum_{k=1}^n c_{i_k} = \sum_{k=0}^{n-1} (n - k)c_{i_k, i_{k+1}}$$

(we denote by i_0 the depot y_1). We observe that the lower bounds we obtain in the following corollary for a minimum weighted matching and the TSP are of the same order of magnitude as the asymptotic results proved for these respective problems in [12] and [4]. The lower bound for the minimum latency problem is also of the same order of magnitude as the optimal cost of this problem, as we show in Proposition 1.

Corollary 2. *The cost of a minimum weighted matching and the cost of the TSP when points are uniformly distributed in $[0, 1]^2$ are $\Omega(\sqrt{n})$ with probability 1. The cost of the minimum latency problem when points are uniformly distributed in $[0, 1]^2$ is $\Omega(n^{3/2})$ with probability 1.*

Proof. For a fixed p , $0 < p < 1$, and for a $c(p)$ such that (4) holds, minimum weighted matching has a cost of at least $(c(p)/2\sqrt{n})(1 - p)n = \Omega(\sqrt{n})$ with probability 1. Selecting $p = \frac{1}{3}$ and $c(p)$ less than $\sqrt{p}/3\sqrt{2}$ means that minimum weighted matching has a cost of at least $0.04\sqrt{n}$ with probability 1. This bound holds also for the TSP, since the minimum weighted matching is a relaxation of the TSP and gives a lower bound for this problem.

Given a solution $(y_1, i_1, i_2, \dots, i_n, y_1)$ of the minimum latency problem, its cost is $\sum_{k=0}^{n-1} (n - k)c_{i_k, i_{k+1}}$. Informally, Lemma 6 states that a significant percentage of customers are isolated. That is, for a fixed p , $0 < p < 1$, and for a $c(p)$ such that (4) holds, at least $(1 - p)n$ of the total line segments connecting two customers used by any solution of the latency problem have a cost of at least $c(p)/\sqrt{n}$. Since the cost function of the minimum latency problem is a weighted sum of the line segments used, where the line segments used first have greater weight,

the cost of any solution is at least

$$\sum_{k=pn}^{n-1} (n-k) \frac{c(p)}{\sqrt{n}} \geq \frac{c(p)}{\sqrt{n}} \frac{(1-p)^2 n^2}{2} = \Omega(n^{3/2}). \tag{5}$$

In particular, for $p = \frac{1}{5}$ and $c(p)$ less than $\sqrt{p}/3\sqrt{2}$, (5) implies that the minimum latency problem has a cost of at least $0.03n^{3/2}$ with probability 1.

The next proposition complements Corollary 2 for the minimum latency problem. That is, it shows that the minimum latency problem has cost $O(n^{3/2})$ when the depot and the customers are points in $[0, 1]^2$. We observe that this result holds in the worst-case sense, namely, when we do not assume any probabilistic distribution of the points.

Proposition 1. *The cost of the minimum latency problem when the customers and the depot are located in $[0, 1]^2$ is $O(n^{3/2})$.*

Proof. We will use an optimal solution to the TSP through the n customers to construct a solution to the minimum latency problem. Let $\text{route}_0 = (i_1, \dots, i_n)$ be an optimal traveling tour through the n customers. Few [7] proved that the optimal cost of the TSP through any n points in $[0, 1]^2$ is at most $\sqrt{2n}$. The cost of route_0 is thus $\sum_{k=1}^{n-1} c_{i_k, i_{k+1}} + c_{i_n, i_1} \leq \sqrt{2n}$. We consider the following n routes for the latency problem. For each l , $1 \leq l \leq n$, let route_l be $(y_1, i_l, i_{l+1}, \dots, i_n, i_1, i_2, \dots, i_{l-1}, y_1)$. That is,

$$\begin{aligned} \text{route}_1 &= (y_1, i_1, \dots, i_n, y_1), \\ \text{route}_2 &= (y_1, i_2, \dots, i_n, i_1, y_1), \\ &\vdots \\ \text{route}_n &= (y_1, i_n, i_1, \dots, i_{n-1}, y_1). \end{aligned}$$

For each l , $1 \leq l \leq n$, the latency cost of route_l is

$$nc_{y_1, i_l} + \sum_{k=l}^{n-1} (n+l-k-1)c_{i_k, i_{k+1}} + (l-1)c_{i_n, i_1} + \sum_{k=1}^{l-1} (l-k-1)c_{i_k, i_{k+1}}.$$

Since each line segment (i_k, i_{k+1}) of the tour route_0 appears in each possible place in routes $\text{route}_1, \dots, \text{route}_n$, the sum of the latency costs of these routes is

$$\sum_{l=1}^n nc_{y_1, i_l} + \sum_{l=0}^{n-1} l \left(\sum_{k=1}^{n-1} c_{i_k, i_{k+1}} + c_{i_n, i_1} \right) \leq n^2\sqrt{2} + \frac{n(n-1)}{2}\sqrt{2n}.$$

The average cost of these routes is at most $n\sqrt{2} + [(n-1)/2]\sqrt{2n} = O(n^{3/2})$. Since at least one of the routes $\text{route}_1, \dots, \text{route}_n$ has a cost of at most the average, this proves that there exists a solution to the minimum latency with cost $O(n^{3/2})$.

We will use Lemma 6 to prove that the quadratic term (2) is $\Omega(\sqrt{n})$ with probability 1 (see Lemma 7). We give some definitions first.

Definition 4. We say that a customer i is nonisolated with respect to a parameter c if it has a neighbor at a distance at most c/\sqrt{n} .

Proposition 2. Given a sequence $l = (i_1, i_2, \dots, i_s)$ of $s > 1$ customers where b of them are nonisolated with respect to c , the following inequality holds:

$$\sum_{i,j \in \{i_1, \dots, i_s\}} \frac{l_{ij}}{s^2} \geq \frac{c}{\sqrt{n}} \left(\frac{s}{6} - 2b \right).$$

Proof. We can express the sum $\sum_{i,j \in \{i_1, \dots, i_s\}} l_{ij}/s^2$ as $\sum_{t=1}^{s-1} [2t(s-t)/s^2] c_{i_t, i_{t+1}}$. Let $\delta(t) = 0$ whenever i_t and i_{t+1} are both nonisolated customers with respect to c , and let $\delta(t) = 1$ otherwise. Then $c_{i_t, i_{t+1}} \geq \delta(t)c/\sqrt{n}$ and, therefore,

$$\begin{aligned} \sum_{i,j \in \{i_1, \dots, i_s\}} \frac{l_{ij}}{s^2} &= \sum_{t=1}^{s-1} \frac{2t(s-t)}{s^2} c_{i_t, i_{t+1}} \\ &\geq \sum_{t=1}^{s-1} \frac{2t(s-t)}{s^2} \frac{\delta(t)c}{\sqrt{n}} \\ &= \left(\sum_{t=1}^{s-1} \frac{2t(s-t)}{s^2} - \sum_{\{t: \delta(t)=0\}} \frac{2t(s-t)}{s^2} \right) \frac{c}{\sqrt{n}}. \end{aligned} \tag{6}$$

In order to bound the right-hand side of this inequality from below, we will use the identities

$$\sum_{t=1}^{s-1} t = \frac{s(s-1)}{2} \quad \text{and} \quad \sum_{t=1}^{s-1} t^2 = \frac{(s-1)s(2s-1)}{6} \leq \frac{s^2(s-1)}{3}.$$

The sum $\sum_{t=1}^{s-1} 2t(s-t)/s^2$ is then

$$\sum_{t=1}^{s-1} \frac{2t}{s} - \sum_{t=1}^{s-1} \frac{2t^2}{s^2} \geq (s-1) - \frac{2}{3}(s-1) = \frac{1}{3}(s-1) \geq \frac{s}{6}. \tag{7}$$

In order to bound the sum $-\sum_{\{t: \delta(t)=0\}} 2t(s-t)/s^2$ from below, we observe that, among all possible distributions of b nonisolated customers in a sequence with s customers, this sum reaches its minimum value when the nonisolated customers are located in the center of the sequence l . The reason for this is that the weight $-2t(s-t)/s^2$ is a quadratic convex function of t that reaches its minimum at $t = s/2$. That is, this sum reaches its minimum when the nonisolated customers in the sequence l are $i_a, i_{a+1}, \dots, i_{a+b-1}$, where a equals either $\lfloor (s-b)/2 \rfloor$ or $\lfloor (s-b)/2 \rfloor + 1$. (In particular, a equals at most $(s-b)/2 + 1$.) Therefore, the sum $\sum_{\{t: \delta(t)=0\}} 2t(s-t)/s^2$ equals at most

$$\sum_{t=a}^{a+b-1} \frac{2t(s-t)}{s^2} \leq \sum_{t=a}^{a+b-1} \frac{2ts}{s^2} = \frac{b(2a+b-1)}{s} \leq \frac{b(s+1)}{s} \leq 2b. \tag{8}$$

Inequalities (7) and (8) imply that the right-hand side of (6) equals at least $(c/\sqrt{n})(s/6 - 2b)$.

For $Q \geq 2$, the quadratic term QC , defined in (2), is $\Omega(\sqrt{n})$ with probability 1, as the next lemma shows.

Lemma 7. *Assuming that $Q \geq 2$, there exists a $\bar{c} > 0$ such that $\lim_{n \rightarrow \infty} P(QC \geq \bar{c}\sqrt{n})$ exists and is equal to 1.*

Proof. Let (K, V) be any optimal solution that has the fullness property. By Lemma 4, there is always an optimal solution that has the fullness property. This implies that at most one vehicle of this solution visits only one customer. We will prove that there exists a $\bar{c} > 0$ such that $\lim_{n \rightarrow \infty} P(QC(K, V) \geq \bar{c}\sqrt{n}) = 1$ for any such (K, V) , which implies that $\lim_{n \rightarrow \infty} P(QC \geq \bar{c}\sqrt{n}) = 1$. Let $p, 0 < p < 1$, and $c > 0$ be parameters to be fixed afterwards. Each vehicle $v_k \in V$ has associated with it a sequence l_k of s_k customers, b_k of which are nonisolated with respect to c . Proposition 2 implies that

$$\begin{aligned} \sum_{k=1}^K \sum_{i,j \in \{1, \dots, n\}} l_{ij}^k \frac{d_i^k d_j^k}{(\sum_{j=1}^n d_j^k)^2} &\geq \sum_{k=1}^K \frac{c}{\sqrt{n}} \left(\frac{s_k}{6} - 2b_k \right) \\ &= \frac{c}{\sqrt{n}} \left(\frac{n}{6} - 2|\{\text{customers nonisolated with respect to } c\}| \right). \end{aligned}$$

We know that

$$\begin{aligned} P(|\{\text{customers nonisolated with respect to } c\}| \leq pn) \\ + P(|\{\text{customers nonisolated with respect to } c\}| \geq pn) &\geq 1 \end{aligned}$$

for any values of p and c . If we choose p and c such that Lemma 6 holds, then

$$\liminf_{n \rightarrow \infty} P(|\{\text{customers nonisolated with respect to } c\}| \leq pn) = 1.$$

Therefore, $\lim_{n \rightarrow \infty} P(|\{\text{customers nonisolated with respect to } c\}| \leq pn)$ exists and is equal to 1 and, thus,

$$\lim_{n \rightarrow \infty} P\left(\sum_{k=1}^K \sum_{i,j \in \{1, \dots, n\}} l_{ij}^k \frac{d_i^k d_j^k}{(\sum_{j=1}^n d_j^k)^2} \geq \frac{c}{\sqrt{n}} \left(\frac{n}{6} - 2pn \right)\right)$$

exists and is equal to 1.

We now have all the pieces necessary to prove the main result.

Theorem 2. *There exists a constant $\hat{c} > 0$ such that*

$$\lim_{n \rightarrow \infty} P\left(\frac{c(\text{VRP}^{\text{ITP}})}{c(\text{VRP})} \leq 2 - \hat{c}\right)$$

exists and is equal to 1.

Proof. We know that $c(\text{VRP}^{\text{ITP}}) \leq R + c(\text{TSP})$. Lemma 5 states that $c(\text{VRP}) \geq \max\{R + QC, c(\text{TSP})\}$. Therefore,

$$\frac{c(\text{VRP}^{\text{ITP}})}{c(\text{VRP})} \leq \frac{R + c(\text{TSP})}{\max\{R + QC, c(\text{TSP})\}} \leq 2 - \frac{QC}{c(\text{TSP})}. \tag{9}$$

We know that the ratio of the cost of TSP to \sqrt{n} converges a.s. to a constant, β . By Lemma 7, we know that $\lim_{n \rightarrow \infty} P(QC \geq \bar{c}\sqrt{n}) = 1$. With the choice $\hat{c} = \bar{c}/\beta < \sqrt{p}(\frac{1}{6} - 2p)/3\sqrt{2}\beta$, the limit

$$\lim_{n \rightarrow \infty} P\left(\frac{QC}{c(\text{TSP})} \geq \hat{c}\right) \tag{10}$$

exists and is equal to 1.

Equations (9) and (10) prove the theorem.

The statement of Theorem 2 does not claim any numerical value for the constant \hat{c} . However, it is easy to give a lower bound on the value of \hat{c} by following the proof of this theorem. For points distributed uniformly in $[0, 1]^2$, the constant $\beta > 0$ is at most 0.93 (see, e.g. [4]). For $p = \frac{1}{36}$, the value $\sqrt{p}(\frac{1}{6} - 2p)/3\sqrt{2}\beta$ (and, therefore, \hat{c}) is at least 0.0046. It is possible to prove a tighter bound for the constant $c(p)$ in Lemma 6 and to prove a tighter bound in Proposition 2: as a result, we can prove that \hat{c} is at least 0.015.

We can generalize this as follows. Let $\mu = E(c_{y_1, X})$ be the expected distance of customer X to the depot. If $\lim_{n \rightarrow \infty} Q/\sqrt{n}$ exists and is equal to a finite value $w > 0$, the strong law of large numbers implies that

$$\lim_{n \rightarrow \infty} \frac{R}{(2\mu/w)\sqrt{n}} = 1 \quad \text{a.s.}$$

We have the following result for the approximation ratio of the ITP heuristic.

Theorem 3. *Let \bar{c} be the constant defined in Lemma 7. Assume that $\lim_{n \rightarrow \infty} Q/\sqrt{n}$ exists and is equal to w , $0 < w < \infty$. The approximation ratio of the ITP heuristic is such that*

- if $2\mu/w + \bar{c} \geq \beta$ then

$$\lim_{n \rightarrow \infty} P\left(\frac{c(\text{VRP}^{\text{ITP}})}{c(\text{VRP})} \leq 1 + \frac{\beta}{2\mu/w + \bar{c}} - \frac{\bar{c}}{2\mu/w + \bar{c}}\right)$$

exists and is equal to 1, and

- if $2\mu/w + \bar{c} < \beta$ then

$$\lim_{n \rightarrow \infty} P\left(\frac{c(\text{VRP}^{\text{ITP}})}{c(\text{VRP})} \leq 1 + \frac{2\mu}{w\beta}\right)$$

exists and is equal to 1.

Proof. The results follow from the ratio

$$\frac{c(\text{VRP}^{\text{ITP}})}{c(\text{VRP})} \leq \frac{R + c(\text{TSP})}{\max\{R + QC, c(\text{TSP})\}}$$

and the limits

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{c(\text{TSP})}{R + QC} &\geq \frac{\beta}{2\mu/w + \bar{c}}, & \limsup_{n \rightarrow \infty} \frac{QC}{R + QC} &\leq \frac{\bar{c}}{2\mu/w + \bar{c}}, \\ \lim_{n \rightarrow \infty} \frac{R}{c(\text{TSP})} &= \frac{2\mu}{w\beta}. \end{aligned}$$

4. Multidepot vehicle routing problem

In this section we generalize the VRP to a multidepot scenario. There are n customers and m depots. Each vehicle starts and ends at the same depot. We assume that there is no restriction on the number of vehicles available at each depot.

4.1. Notation

We extend the notation from the previous sections. Unless otherwise stated, the customers and depots are modeled as points in the plane. The location of the i th customer is denoted by x_i for any $1 \leq i \leq n$. The location of the j th depot is denoted by y_j for any $1 \leq j \leq m$. The set of customers is denoted by $X^{(n)}$. The set of depots is denoted by $Y^{(m)}$. A solution to an MDVRP is denoted by a pair (K, V) , where K is the number of vehicles used and $V = \{v_k : k \in \{1, \dots, K\}\}$ is the set of vehicles (or routes). For each customer i , let $c_i = \min\{c_{i,y} : y \in Y^{(m)}\}$ be its minimum distance to a depot. By $R_{MD} = \sum_{i=1}^n 2c_i/Q$ we denote the *multidepot radial cost*. We denote by $c(\text{MDVRP})$ or $c(\text{MDVRP}(X^{(n)}))$ the cost of an optimal MDVRP, and by $c(\text{TSP})$ or $c(\text{TSP}(X^{(n)} \cup Y^{(m)}))$ the cost of an optimal traveling salesman tour through all customers and depots. Let $c_\infty(\text{MDVRP})$ denote the cost of the MDVRP when the capacity of vehicles is infinite. Following Definition 1, given a route $v_k = (y_{v_k}, i_1^k, i_2^k, \dots, i_{s^k}^k, y_{v_k})$ and customers i and j visited by v_k , we denote by l_{ij}^k the length of the unique path from i to j determined by v_k that does not pass through the depot y_{v_k} . If i or j is not visited by v_k , then we let $l_{ij}^k = 0$.

5. An algorithm for the MDVRP

We generalize the ITP heuristic to the multidepot case. The multidepot iterated tour partitioning (MDITP) heuristic MDITP(α) that we propose uses the heuristic ITP(α) as a subroutine. It first assigns each customer to its closest depot (see Figure 1) and then solves m independent VRP problems. More formally, the MDITP heuristic is as follows.

0. Let $S_j = \emptyset$ for all $j, 1 \leq j \leq m$.
1. For each customer x_i , find y_j , its closest depot, and replace S_j by $S_j \cup \{x_i\}$.
2. For each $j, 1 \leq j \leq m$, run the heuristic ITP(α) to solve the VRP problem VRP_j approximately, where y_j is the only depot and the set of customers is S_j .

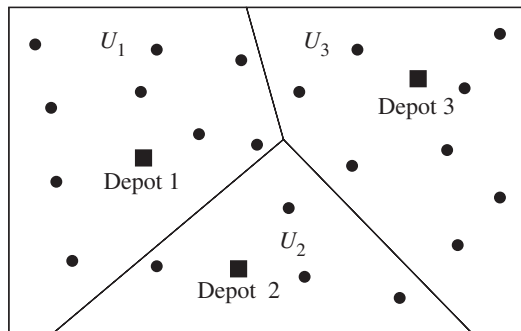


FIGURE 1: Assigning customers to their closest depot.

Let $c(\text{TSP}_j) := c(\text{TSP}(S_j \cup \{y_j\}))$ denote the cost of an optimal subtour that visits all customers of S_j and the depot y_j . The cost of the solution produced by this heuristic is the sum of the costs of the solutions produced by $\text{ITP}(\alpha)$ for each VRP_j , which is at most $R_{\text{MD}} + \alpha \sum_{j=1}^m c(\text{TSP}_j)$. In what follows, we denote by $c(\text{MDVRP}^{\text{MDITP}})$ the cost of this heuristic when we use $\text{ITP}(1)$ as a subroutine. This analysis implies the following result.

Lemma 8. *The cost of MDVRP(1) is at most $R_{\text{MD}} + \sum_{j=1}^m c(\text{TSP}_j)$.*

It is easy to see that the heuristic MDITP(1) is an $\Omega(m)$ -approximation algorithm in the worst-case analysis, where m is the number of depots. Let us consider the following example.

Example 1. There are m depots located on the unit circle and there are m customers located close to the origin, $(0, 0)$. Let $\varepsilon > 0$ and let Q , the capacity of each vehicle, be infinite. For each j , $1 \leq j \leq m$, depot y_j is located at the point $(\cos 2\pi j/m, \sin 2\pi j/m)$ and customer i_j is located at the point $(\varepsilon \cos 2\pi j/m, \varepsilon \sin 2\pi j/m)$.

In this example, the heuristic MDITP(α) will assign customer i_j to depot y_j and will output a solution with cost $2m(1 - \varepsilon)$. On the other hand, the solution that uses only one vehicle with route $(y_1, i_1, i_2, \dots, i_m, i_1, y_1)$ has a cost of at most $2(1 - \varepsilon) + \varepsilon 2\pi$. By letting ε go to 0, the ratio of the two solutions goes to m . Therefore, the heuristic MDITP(1) is an $\Omega(m)$ -approximation algorithm. However, in Section 7 we will show that the heuristic MDITP(1) performs much better on average. Namely, its approximation ratio is the same as that of the ITP heuristic for the single-depot case, on average.

6. Lower bounds on the optimal cost of the MDVRP

In this section we generalize the results of Section 2 to the multidepot case. The following lemma generalizes Lemma 1.

Lemma 9. *The cost of the MDVRP is at least $\max\{R_{\text{MD}}, c_\infty(\text{MDVRP})\}$.*

Proof. Let X_k be the set of customers visited by vehicle v_k and let y be the depot at which v_k starts and ends. Then

$$c(v_k) \geq 2 \max_{i \in X_k} \{c_{i,y}\} \geq 2 \max_{i \in X_k} \{c_i\} \geq 2 \frac{\sum_{i \in X_k} c_i}{|X_k|} \geq \frac{2}{Q} \sum_{i \in X_k} c_i.$$

Summing over all vehicles, we find that $c(\text{MDVRP}) \geq R_{\text{MD}}$. The inequality $c(\text{MDVRP}) \geq c_\infty(\text{MDVRP})$ holds since any feasible solution to the MDVRP with vehicles with capacity Q is also feasible for the infinite-capacity MDVRP.

The next lemma relates $c_\infty(\text{MDVRP})$ to the cost of a TSP.

Lemma 10. *Let $c(\text{TSP}(Y^{(m)}))$ denote the cost of an optimal subtour through the depots. Then*

$$c_\infty(\text{MDVRP}) \geq c(\text{TSP}(X^{(n)} \cup Y^{(m)})) - c(\text{TSP}(Y^{(m)})).$$

Proof. Observe that $c_\infty(\text{MDVRP}) + c(\text{TSP}(Y^{(m)}))$ is the cost of the walking tour formed by the union of the routes of an optimal solution to the infinite-capacity MDVRP plus a subtour through all the depots. This walking tour can be transformed into a TSP of lesser cost through all customers and depots by shortcutting nodes already visited. Therefore, $c_\infty(\text{MDVRP}) + c(\text{TSP}(Y^{(m)})) \geq c(\text{TSP}(X^{(n)} \cup Y^{(m)}))$, and the result follows.

The following lemma is a generalization of Lemma 3.

Lemma 11. *Let (K, V) be a solution to the MDVRP. Let $d_i^k = 1$ if vehicle v_k visits customer i and let $d_i^k = 0$ otherwise. Then the cost of (K, V) is at least*

$$\sum_{k=1}^K \sum_{i=1}^n 2c_i \frac{d_i^k}{\sum_{t=1}^n d_t^k} + \sum_{k=1}^K \sum_{i,j \in \{1, \dots, n\}} l_{ij}^k \frac{d_i^k d_j^k}{(\sum_{t=1}^n d_t^k)^2}. \tag{11}$$

We generalize the fullness property given in Definition 2 to the multidepot case. A solution to the MDVRP is said to have the fullness property if $|A| \geq n - mQ/2$. The following lemma states that there always exists an optimal solution to the MDVRP that has the fullness property.

Lemma 12. *There exists an optimal solution (K, V) to the MDVRP such that $|A| \geq n - mQ/2$.*

Proof. Given an optimal solution (K, V) , each depot has at most one vehicle that is not half-full. Otherwise, we can merge the routes of two vehicles that are not half-full and obtain another optimal solution that has the fullness property.

We give a name to the quadratic term in (11).

Definition 5. Given a solution (K, V) to the MDVRP, let

$$QC(K, V) = \sum_{k=1}^K \sum_{i,j \in \{2, \dots, n\}} l_{ij}^k \frac{d_i^k d_j^k}{(\sum_{t=1}^n d_t^k)^2}.$$

Let QC be the maximum value of $QC(K, V)$ among all optimal solutions (K, V) to the MDVRP that have the fullness property.

It is apparent that the multidepot radial cost plus the quadratic term QC is a lower bound on the cost of any solution (K, V) that has the fullness property. Since there is at least one optimal solution that has the fullness property, the following lemma holds.

Lemma 13. *The cost of the MDVRP is at least*

$$\max\{R_{MD} + QC, c(\text{TSP}(X^{(n)} \cup Y^{(m)})) - c(\text{TSP}(Y^{(m)}))\}.$$

Lemma 7 also holds in the multidepot case. The constant \bar{c} in the following lemma is the same one as in Lemma 7.

Lemma 14. *Assuming that $Q \geq 2$, there exists a $\bar{c} > 0$ such that $\lim_{n \rightarrow \infty} P(QC \geq \bar{c}\sqrt{n})$ exists and is equal to 1.*

7. Probabilistic analysis of the MDVRP

In this section we extend the main results of the single-depot case (e.g. Theorems 1 and 2) to the multidepot case.

7.1. Probabilistic analysis of lower bounds

We analyze the unit-demand Euclidean MDVRP where the depots and customers are modeled as points in the plane. The customers have unit demand. The m depots y_1, \dots, y_m are fixed in advance, whereas the locations of the n customers are i.i.d. uniform random variables in $[0, 1]^2$.

Let U_1, \dots, U_m be a disjoint partition of the square $[0, 1]^2$ such that each U_j contains exactly one depot, namely y_j . For each j , $1 \leq j \leq m$, let $n_j \geq 0$ be the number of customers that belong to U_j and let $X^{(n_j)}$ be the set of customers that belong to U_j . Let $c(\text{TSP}_j) := c(\text{TSP}(X^{(n_j)} \cup \{y_j\}))$ be the cost of an optimal subtour that visits all customers of $X^{(n_j)}$ and depot y_j . The following result holds.

Lemma 15. *We have*

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^m c(\text{TSP}_j)}{c(\text{TSP})} = 1 \quad \text{a.s.}$$

Proof. For each j , $1 \leq j \leq m$, let f_j be the restriction of the uniform density to the set U_j . That is,

$$f_j(x) = \begin{cases} 1 & \text{if } x \in U_j, \\ 0 & \text{otherwise.} \end{cases}$$

The function $f/|U_j|$ is the density function of a customer in U_j . We will prove that

$$\lim_{n \rightarrow \infty} \frac{c(\text{TSP}_j)}{\sqrt{n}} = \beta \int f_j^{1/2} dx \quad \text{a.s.} \tag{12}$$

for each j , $1 \leq j \leq m$. Let us fix j . In order to prove (12), we would like to restrict the experiment to the customers that fall inside U_j and apply the classic result of [4] that we stated at the beginning of Section 3. However, this has to be done with some care, since the number of customers that fall inside U_j is random. We observe that the distribution of customers that fall outside U_j does not affect $c(\text{TSP}_j)$. Therefore, the cost of the optimal TSP in the following experiment is probabilistically the same random variable as $c(\text{TSP}_j)$. Let the n customers be i.i.d. points with the following distribution: with probability $p_j = |U_j|$ customer i is located in U_j according to density f_j/p_j , and with probability $1 - p_j$ customer i is located in depot y_j . The main properties of the new experiment are as follows.

1. All the points fall inside U_j .
2. The cost of an optimal tour through all points and depot y_j in the new experiment has the same distribution as the cost $c(\text{TSP}_j)$ in the original experiment.

Since the absolutely continuous-density part of the distribution of a customer in the new experiment is f_j , the classic result of [4] implies that (12) holds for each j .

Since (12) holds almost surely for each j , upon summation we have

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^m c(\text{TSP}_j)}{\sqrt{n}} = \beta \int \sum_{j=1}^m f_j^{1/2} dx = \beta \int f^{1/2} dx \quad \text{a.s.} \tag{13}$$

The last equality holds because $f = \sum_{j=1}^m f_j$ and the supports of the functions f_j are mutually disjoint. Finally, (13) and (3) imply that

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^m c(\text{TSP}_j)}{c(\text{TSP})} = \lim_{n \rightarrow \infty} \frac{\sum_{j=1}^m c(\text{TSP}_j)/\sqrt{n}}{c(\text{TSP})/\sqrt{n}} = 1 \quad \text{a.s.}$$

The following theorem generalizes Theorem 1 to the multidepot case.

Theorem 4. Let X_1, X_2, \dots be a sequence of i.i.d. uniform random points in $[0, 1]^2$, and let $X^{(n)}$ denote the first n points of this sequence. Let μ denote the expected distance between X_1 and its closest depot.

- If $\lim_{n \rightarrow \infty} Q/\sqrt{n} = 0$ then

$$\lim_{n \rightarrow \infty} \frac{c(\text{MDVRP}(X^{(n)}))Q}{n} = 2\mu \quad \text{a.s.}$$

- If $\lim_{n \rightarrow \infty} Q/\sqrt{n} = \infty$ then

$$\lim_{n \rightarrow \infty} \frac{c(\text{MDVRP}(X^{(n)}))}{\sqrt{n}} = \beta \quad \text{a.s.,}$$

where $\beta > 0$ is the constant introduced in [4].

Proof. Lemmas 8, 9, and 10 imply that

$$\begin{aligned} \max\{R_{\text{MD}}, c(\text{TSP}(X^{(n)} \cup Y^{(m)})) - c(\text{TSP}(Y^{(m)}))\} &\leq c(\text{MDVRP}(X^{(n)})) \\ &\leq R_{\text{MD}} + \sum_{j=1}^m c(\text{TSP}_j). \end{aligned}$$

The ratio of R_{MD} to n/Q is

$$\frac{R_{\text{MD}}Q}{n} = \sum_{i=1}^n \frac{2c_iQ}{nQ} = \sum_{i=1}^n \frac{2c_i}{n}.$$

The law of large numbers implies that this ratio converges a.s. to 2μ . If X_1, X_2, \dots are uniformly bounded and $\lim_{n \rightarrow \infty} Q/\sqrt{n} = 0$, then

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^m c(\text{TSP}_j)Q}{n} = \lim_{n \rightarrow \infty} \frac{\sum_{j=1}^m c(\text{TSP}_j)}{\sqrt{n}} \frac{Q}{\sqrt{n}} = 0$$

since

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^m c(\text{TSP}_j)}{\sqrt{n}} = \lim_{n \rightarrow \infty} \frac{c(\text{TSP}(X^{(n)} \cup Y^{(m)}))}{\sqrt{n}}$$

is a constant. Therefore,

$$\begin{aligned} 2\mu &= \lim_{n \rightarrow \infty} \frac{\max\{R_{\text{MD}}, c(\text{TSP}(X^{(n)} \cup Y^{(m)})) - c(\text{TSP}(Y^{(m)}))\}}{n/Q} \\ &\leq \lim_{n \rightarrow \infty} \frac{c(\text{MDVRP}(X^{(n)}))}{n/Q} \\ &\leq \lim_{n \rightarrow \infty} \frac{R_{\text{MD}} + \sum_{j=1}^m c(\text{TSP}_j)}{n/Q} \\ &= 2\mu \end{aligned}$$

in this case.

Now we analyze $\lim_{n \rightarrow \infty} c(\text{MDVRP}(X^{(n)}))/\sqrt{n}$ for $\lim_{n \rightarrow \infty} Q/\sqrt{n} = \infty$. By definition of R_{MD} , we have

$$\frac{R_{\text{MD}}}{\sqrt{n}} = \sum_{i=1}^n \frac{2c_i}{Q\sqrt{n}} = \frac{\sqrt{n}}{Q} \sum_{i=1}^n \frac{2c_i}{n}.$$

Since the sum $\sum_{i=1}^n 2c_i/n$ a.s. converges to a constant as n goes to ∞ , the whole expression converges to 0 a.s. for $\lim_{n \rightarrow \infty} Q/\sqrt{n} = \infty$. The ratio of $c(\text{TSP}(X^{(n)} \cup Y^{(m)}))$ to \sqrt{n} converges to β a.s. The ratio of $c(\text{TSP}(Y^{(m)}))$ to \sqrt{n} converges to 0 since $c(\text{TSP}(Y^{(m)}))$ is a constant. The ratio of $\sum_{j=1}^m c(\text{TSP}_j)$ to \sqrt{n} also converges to β a.s., by Lemma 15. Therefore,

$$\begin{aligned} \beta &= \lim_{n \rightarrow \infty} \frac{\max\{R_{\text{MD}}, c(\text{TSP}(X^{(n)} \cup Y^{(m)})) - c(\text{TSP}(Y^{(m)}))\}}{\sqrt{n}} \\ &\leq \lim_{n \rightarrow \infty} \frac{c(\text{MDVRP}(X^{(n)}))}{\sqrt{n}} \\ &\leq \lim_{n \rightarrow \infty} \frac{R_{\text{MD}} + \sum_{j=1}^m c(\text{TSP}_j)}{\sqrt{n}} = \beta \end{aligned}$$

in this case.

7.2. Probabilistic analysis of an algorithm for the MDVRP

The following theorem is an extension of Theorems 2 and 3.

Theorem 5. *There exists a constant $\hat{c} > 0$ such that*

$$\lim_{n \rightarrow \infty} \text{P}\left(\frac{c(\text{MDVRP}^{\text{MDITP}})}{c(\text{MDVRP})} \leq 2 - \hat{c}\right) \tag{14}$$

exists and is equal to 1. Moreover,

- 1. *if $\lim_{n \rightarrow \infty} Q/\sqrt{n} = 0$ then*

$$\lim_{n \rightarrow \infty} \frac{c(\text{MDVRP}^{\text{MDITP}})}{c(\text{MDVRP})} = 1 \quad \text{a.s.},$$

- 2. *if $\lim_{n \rightarrow \infty} Q/\sqrt{n} = w > 0$ and $2\mu/w + \bar{c} \leq \beta$, then*

$$\lim_{n \rightarrow \infty} \text{P}\left(\frac{c(\text{MDVRP}^{\text{MDITP}})}{c(\text{MDVRP})} \leq 1 + \frac{\beta}{2\mu/w + \bar{c}} - \frac{\bar{c}}{2\mu/w + \bar{c}}\right)$$

exists and is equal to 1,

- 3. *if $\lim_{n \rightarrow \infty} Q/\sqrt{n} = w > 0$ and $2\mu/w + \bar{c} < \beta$, then*

$$\lim_{n \rightarrow \infty} \text{P}\left(\frac{c(\text{MDVRP}^{\text{MDITP}})}{c(\text{MDVRP})} \leq 1 + \frac{2\mu}{w\beta}\right)$$

exists and is equal to 1, and

- 4. *if $\lim_{n \rightarrow \infty} Q/\sqrt{n} = \infty$ then*

$$\lim_{n \rightarrow \infty} \frac{c(\text{MDVRP}^{\text{MDITP}})}{c(\text{MDVRP})} = 1 \quad \text{a.s.}$$

Here β is the constant introduced in [4], μ is the expected distance of a customer to its closest depot, and \bar{c} is the constant introduced in Lemma 14.

Proof. The proofs of these results follow the same lines as the proofs of Theorems 2 and 3. The proofs of (14) and item 2 follow from Lemmas 8, 13, 14, and 15. The proofs of items 1, 3, and 4 follow from Lemma 8 and Theorem 4.

8. Summary

We have studied the unit-demand Euclidean VRP using a probabilistic approach. We presented a new lower bound for the metric VRP with an unlimited number of homogeneous vehicles. This bound is at least the radial cost plus a term of the same order of magnitude as the cost of the TSP when the customers have unit demand, they are modeled as i.i.d. uniform random points in the plane, and the distance is the Euclidean distance. This lower bound improves the radial cost bound defined in [8]. We showed that this improvement over the radial cost is $\Omega(\sqrt{n})$ with probability converging to 1 as the number of customers goes to ∞ . In particular, this improvement is of the same order of magnitude as the cost of the TSP, which is also a lower bound on the cost of the unit-demand Euclidean VRP.

The improvement on the lower bound of the unit-demand Euclidean VRP we presented implies an improvement on the approximation ratio of the ITP heuristic for this problem. As a result, the approximation ratio of the ITP heuristic was shown to be strictly better than 2 with probability 1 as the number of customers goes to ∞ .

In the second part of the paper we analyzed the multidepot VRP. We gave a natural generalization of the ITP heuristic for this problem. The lower bounds presented in the first part of the paper were extended to this problem. The asymptotic results for the single-depot case from the first part were generalized to the multidepot case.

Acknowledgements

We would like to acknowledge the thorough reviews provided by the referees, which led to improvements of the paper.

References

- [1] ALTINKEMER, K. AND GAVISH, B. (1990). Heuristics for delivery problems with constant error guarantees. *Transportation Sci.* **24**, 294–297.
- [2] ARORA, S. (1996). Polynomial time approximation schemes for Euclidean TSP and other geometric problems. In *Proc. 37th Annual Symp. Foundations Comput. Sci.*, IEEE Computer Society Press, Los Alamitos, CA, pp. 2–11.
- [3] ASANO, T., KATOH, N., TAMAKI, H. AND TOKUYAMA, T. (1999). Covering points in the plane by k -tours: towards a polynomial time approximation scheme for general k . In *Proc. STOC '97*, ACM, New York, pp. 275–283.
- [4] BEARDWOOD, J., HALTON, J. L. AND HAMMERSLEY, J. M. (1959). The shortest path through many points. *Proc. Camb. Phil. Soc.* **55**, 299–327.
- [5] BLUM, A. *et al.* (1994). The minimum latency problem. In *Proc. STOC '94*, ACM, New York, pp. 163–171.
- [6] BOMPADRE, A., DROR, M. AND ORLIN, J. B. (2006). Improved bounds for vehicle routeing solutions. *Discrete Optimization* **3**, 299–316.
- [7] FEW, L. (1955). The shortest path and the shortest path through n points. *Mathematika* **2**, 141–144.
- [8] HAIMOVICH, M. AND RINNOOY KAN, A. H. G. (1985). Bounds and heuristics for capacitated routeing problems. *Math. Operat. Res.* **10**, 527–542.
- [9] HAIMOVICH, M., RINNOOY KAN, A. H. G. AND STOUGIE, L. (1988). Analysis of heuristics for vehicle routing problems. In *Vehicle Routing: Methods and Studies*, eds B. L. Golden and A. A. Assad, Elsevier, Amsterdam, pp. 47–61.
- [10] LI, C. L. AND SIMCHI-LEVI, D. (1990). Worst-case analysis of heuristics for the multidepot capacitated vehicle routeing problems. *ORSA J. Comput.* **2**, 64–73.

- [11] MITCHELL, J. S. B. (1996). Guillotine subdivisions approximate polygonal subdivisions: a simple new method for the geometric k -MST problem. In *Proc. 7th ACM-SIAM Symp. Discrete Algorithms*, ACM, New York, pp. 402–408.
- [12] PAPADIMITRIOU, C. H. (1977). The probabilistic analysis of matching heuristics. In *Proc. 15th Annual Allerton Conf. Commun. Control Comput.*, University of Illinois Press, Champaign, IL, pp. 368–378.
- [13] STOUGIE, L. (1987). *Design and Analysis of Algorithms for Stochastic Integer Programming* (CWI Tract 37). Stichting Mathematisch Centrum, Amsterdam.