The Stochastic Capacitated Traveling Salesmen Location Problem: a computational comparison for a United States instance

Luca Bertazzi\textsuperscript{a}, Francesca Maggioni\textsuperscript{b}*

\textsuperscript{a} Department of Economics and Management, University of Brescia, Contrada S. Chiara 50, Brescia 25122, Italy
\textsuperscript{b} Department of Management, Economics and Quantitative Methods, University of Bergamo, 24121, Italy

Abstract

We study a problem in which a facility has to be located in a given area to serve a given number of customers. The position of the customers is not known. The service to the customers is carried out by several traveling salesmen. Each of them has a capacity in terms of the maximum number of customers that can be served in any tour. The aim is to determine the service zone (in a shape of a circle) that minimizes the expected cost of the traveled routes. The center of the circle is the location of the facility. Once the position of the customers is revealed, the customers located outside the service zone are served with a recourse action at a greater unit cost. For this problem, we compare the performance of two different approaches. The first is a solution based on a heuristic proposed for a similar well known problem and the second is a solution based on a stochastic second–order cone model. An illustrative example on a United States instance with 13509 nodes shows the different solutions and expected costs obtained by the two approaches.

Keywords: Facility Location, Traveling Salesmen Location Problem, Stochastic Second–Order Cone, Computational Results.
1. Introduction

One of the most important classes of problems in the strategic design of logistic networks is given by the Facility location problems. We refer to Klose and Drezl (2005), ReVelle and Eiselt (2005) and Melo et al (2009) for recent surveys on facility location problems and to Snyder (2006) and Nagy and Salhi (2007) for surveys on facility location problems under uncertainty.

Location-Routing Problems (LRP) is a relatively new class of problems, in which the classical facility location and the vehicle routing problems are integrated. We refer to Laporte (2009) for a recent survey on vehicle routing problems. Although it is well known that facility location and routing are often interrelated, LRP received little attention in the past. However, in the last years, several papers about deterministic location–routing problems have been published. Instead, a few papers have been devoted to location–routing problems with stochastic demand.

The Traveling Salesman Location Problem (TSLP) is the simplest location–routing problem with stochastic demand. A set of customers is served by a single facility. At each time, only a subset of customers has to be served. A TSP is built to serve this subset of customers. The aim is to determine where to locate the facility in order to minimize the expected cost of the TSP. This problem has several applications in many service systems, such as delivery services, customer pickup services, repair vehicles. It has been introduced by Burness and White (1976) and studied by Berman and Simchi-Levi (1986), Simchi-Levi and Berman (1987, 1988) and McDiarmid (1992). It has been extended to the case, referred to as the Capacitated Traveling Salesmen Location Problem, with several capacitated salesmen, by Simchi-Levi (1991). The main difficulty of these problems is that there exist \(2^n - 1\) different subsets of \(n\) customers that can require to be served and therefore an exponential number of TSP has to be solved. A polynomial time heuristic algorithm, with known relative worst-case error, has been proposed in Simchi-Levi (1991). A different case is given by the Probabilistic Traveling Salesman Location Problem (PTSLP). In this case, first an a priori TSP visiting all customers is computed. Then, for each time, the TSP to serve the subset of customers that have to be visited is obtained by just skipping in the a priori tour the customers that have not to be served at that time. The aim is to simultaneously determine where to locate the facility and the a priori TSP that minimize the expected cost. This problem has been introduced by Berman and Simchi-Levi (1986) and has been studied by Bertsimas (1989). Recent papers are Klibi et al. (2010), Santoso et al (2005) (see also Beraldi & Bruni, 2009 for an emergency service vehicle location).

We study a problem in which a single facility (typically a service station) has to be located in a given area. This facility is used to serve a given number of customers. The position of the customers is not known. The service to the customers is carried out by several traveling salesmen. Each of them has a capacity in terms of the maximum number of customers that can be served in any tour. The aim is to determine the service zone (in a shape of a circle) that minimizes the expected cost of the traveled routes. The center of the circle is the location of the facility. Once the position of the customers is revealed, the customers located outside the service zone are served with a recourse action at a greater unit cost. This problem is referred to as the Stochastic Capacitated Traveling Salesmen Location Problem with Recourse (SCTSLP–R).

Our aim is to propose stochastic optimization models (see Birge & Louveaux, 1997; Kall & Wallace, 1994; Shapiro, 2008; Maggioni et al 2009; Maggioni et al. 2013 for two logistic applications) and solution methodologies to solve the SCTSLP–R. We first consider a solution of the problem based on the heuristic proposed by Simchi-Levi (1991) for the Capacitated Traveling Salesmen Problem. Bertazzi and Maggioni (2012) proved that in the worst–case, this algorithm can give a solution infinitely worse with respect to the optimal one. Then, the problem was first modeled as a two-stage stochastic semidefinite programming problem (see Ariyawansa & Zhu, 2006; Pataki, 2003; Vandenberghe & Boyd, 1996; Ko & Vaidya 2000) and then as a two-stage stochastic second-order cone programming problem (see Alizadeh & Goldfarb, 2003; Maggioni et al. 2009; Maggioni et al. 2010; Maggioni & Wallace 2012), in which the first stage decisions are the facility location and the radius of the service zone. Each scenario is represented by a set of ellipses. Each ellipse represents the area covered by a traveling salesman to supply its customers, randomly generated by uniform distribution inside
a given region. Finally, computational results showed that the solutions obtained by exactly solving this model significantly dominate the ones obtained by the heuristic algorithm.

In this paper, the problem is described and formulated in Section 2. In Section 3 the two solution approaches are described. Finally, in Section 4 an illustrative example, based on a United States TSP instance with 13509 nodes, is provided to compare the solutions and the expected costs obtained by the different approaches.

2. Problem description and formulation

A facility has to serve $R$ customers in a given square $A$ with side length $L$. The position of the customers is not known. A set $N = \{1, 2, \ldots, N\}$ of traveling salesmen is available to serve the customers. Each traveling salesman $i \in N$ has a given initial position and a given capacity $q_i$ in terms of number of customers that can be served in a tour. A realization of the customers positions represents a scenario of the problem. Each scenario is modeled by a given number of ellipses. Each ellipse corresponds to the area covered by a traveling salesman at a scenario. Let $P = \{1, 2, 3, 4\}$ be the set of scenarios. Each scenario $P_k$ is composed of $N$ random ellipses and has probability $p_k$. The ellipse $E_k^i$ corresponds to the area covered by the traveling salesman $i$ at scenario $k$ and is composed of a maximum of $q$ customers. Therefore, the number of customers $R$ is not greater than $Knq$. The problem is to determine the service zone that minimizes the total expected cost.

The service zone is a circle $C$, with center $\bar{u} \in \mathbb{R}^2$ (with $\mathbb{R}^2$ the space of 2-dimensional real vectors) and radius $r$:

$$C = \{u \in \mathbb{R}^2: u^T u - 2\bar{u}^T u + y \leq 0\}$$

where the center $\bar{u} \in \mathbb{R}^2$ and $y$ are decision variables of the problem. The corresponding radius $r$ is $\sqrt{\bar{u}^T u - y}$ and is not greater than $r_{max} = \sqrt{2L}$. The cost to select the circle $C$ is $\alpha$ times the cost of the corresponding Capacitated Vehicle Routing Problem (CVRP), estimated by using the formula inspired to Daganzo (1984):

$$D_1(r) = 2\bar{d} \frac{r}{r_{max}} + 0.57 \sqrt{AR \frac{r}{r_{max}}} = r \left( \frac{R}{q} + 0.57 \sqrt{\pi R \frac{r}{r_{max}}} \right),$$

(1)

where $\bar{d} = r/2$ is the average distance between the customers and the facility, $Rr/r_{max}$ is the number of customers expected in the service zone $C$ and $A = \pi r^2$ is the area of the service zone $C$. This formula reduces to

$$D_1(r) = r \left( \frac{R}{q} + 0.57 \sqrt{\pi R} \right),$$

whenever the service zone is able to cover all possible customers. An example is when $u$ is equal to the center of the square $A$ and $r = \sqrt{2L}/2$.

In any scenario $k$, if all customers are covered by $C$, no further action is needed. Otherwise, a recourse action is needed to enlarge the service zone $C$ to the new circle

$$C_k = \{u \in \mathbb{R}^2: u^T u - 2\bar{u}^T u + \bar{y}_k \leq 0\},$$

with the same center $\bar{u}$ of $C$ and radius $r_k = \sqrt{\bar{u}^T u - \bar{y}_k}$ greater than $r$ enough to cover all customers in the scenario. The cost of this new circle is $\beta$ times, with $\beta > \alpha$, the cost of the corresponding CVRP, estimated by
where \( r + (r_k - r)/2 \) is the average distance between the customers and the facility, \( R(1 - r/r_{\text{max}}) \) is the number of customers expected outside the service zone \( C \) and \( \pi(r_k^2 - r^2) \) is the difference between the area of \( C_k \) and the area of the service zone \( C \).

The aim of the problem is to determine \( \bar{u}, \gamma \) and \( \bar{y}_k \) for all scenarios \( k \in K \), such that the expected cost

\[
\alpha D_1(r) + \beta \sum_{k \in K} p_k D_2(r, r_k)
\]

is minimized.

### 3. Solution Approaches

One of the simplest solutions of the Stochastic Capacitated Traveling Salesmen Location Problem with Recourse SCTSLP-R consists in fixing the service zone equal to the smallest circle covering the square \( A \). This circle has center in the center of the square \( A \) and radius \( r = \sqrt{2L}/2 \). This heuristic is referred to as Heuristic SCC. Bertazzi and Maggioni (2012) showed that this heuristic can have a performance infinitely worse than the optimal solution of the SCTSLP-R problem.

A second heuristic, referred to as Heuristic SL, is based on the heuristic proposed by Simchi–Levi (1991) for the Capacitated Traveling Salesmen Problem, which selects the customer at which to locate the facility. Given the location of the facility, the Heuristic SL selects, as service zone, the minimum enclosing circle having center at the selected location. Let \( R \) be the set of the \( R \) customers to serve, \( h_a \) be the probability that customer \( a \in R \) requires to be served at a given time. Since \( h_a \) is not a data of the SCTSLP-R problem, we assume \( h_a = \frac{1}{R} \).

The Heuristic SL can be described as follows.

**Heuristic SL**

For each customer \( a \in R \):

Order the customers \( b \in R \) in the non-decreasing order of \( d(a, b) \). Rename them as \( \gamma_1, \gamma_2, \ldots, \gamma_R \).

Note that \( \gamma_1 = a \).

Compute

\[
\Phi(a) = \frac{1}{q} \sum_{b \in R} h_b d(a, b) + \left( 1 - \frac{1}{q} \right) \left( \sum_{b=2}^{R} d(a, \gamma_b) h_{\gamma_b} \prod_{r=1}^{\delta-1} (1 - h_{\gamma_r}) \right).
\]

Select the customer \( a^* = \arg \min_{a \in R} \Phi(a) \) as facility location.

Compute the radius \( r \) of the minimum enclosing circle having center in \( a^* \).

Bertazzi and Maggioni (2012) showed that this heuristic has a cost at least 73.205% greater than the optimal cost of the SCTSLP-R problem in the instances with one scenario, while there exists a tight worst-case performance bound of 2.07 in the scenario in which this heuristic has maximum cost. In the general case, i.e. instances with several scenarios, this heuristic can have a performance infinitely worse than the one of the optimal solution of the SCTSLP-R problem.

A stochastic second–order cone model formulations of the SCTSLP–R problem is proposed in Bertazzi and Maggioni (2012). Second-order cone programming (SOCP) consists in convex optimization problems in which a
The linear function is minimized over the intersection of an affine set and the product of second-order (Lorentz) cones. The first step is to approximate \( D_1(r) \) by a function proportional to the area of the circle \( C \). Then, the second stage cost \( D_2(r) \) is approximated by a function proportional to the difference between the area of \( C_k \) and the area of \( C \). Let \( x = [d_2, \bar{u}, \gamma]^T \) and \( y = [z, \bar{y}, \delta] \) be the first stage and the second stage decision variable, respectively. In particular, \( d_2 \) is an upper bound on square of the radius of the circle \( C \), \( \bar{u} \in \mathbb{R}^2 \) is the center of circle \( C \), and \( \gamma \) is the coefficient in the equation of circle \( C \). The vector \( \delta \) \( \in \mathbb{R}^K \) is the vector of the upper bounds at each scenario \( k \) on the quantity the radius of circle \( C \) should be enlarged to cover all the ellipses \( E_k^i \) by means of an effective service zone \( C_k \) with

\[
C_k = \{ u \in \mathbb{R}^2 : u^T u - 2\bar{u}^T u + \bar{y}_k \leq 0 \},
\]

\( \gamma \in \mathbb{R}^K \) is the vector of the coefficients \( \gamma_k \) of the second stage circles \( C_k \), \( k \in K \) and \( \delta \in \mathbb{R}^K \) is a vector of nonnegative parameters (see Vandenberghe & Boyd, 1996; Sun & Freund, 2004). The coefficients of the decision variables in the objective function are given by \( c = [\alpha \theta^* \pi, 0, 0]^T \) and \( q = [\beta \phi^* \pi, 0, 0]^T \) where \( \alpha > 0 \) is the cost per unit of the area of \( C \) and \( \beta \) is the vector of identical costs per unit increase of the area of \( C \) to the area of \( C_k \), \( \theta^* \) and \( \phi^* \) obtained by solving a minimum least square problem in order to approximate \( D_1(r) \) and \( D_2(r, r_k) \) (see Bertazzi & Maggioni, 2012). The model can be formulated as follows:

\[
\begin{align*}
\min_{x \in \mathbb{R}, y \in \mathbb{R}^2} & \quad c^T x + \sum_{k=1}^K p_k q^T y \\
\text{Subject to} & \quad d_2 + \gamma \geq \|\bar{u}\|^2, \\
& \quad (s_{kj}^i + \delta_k \lambda_{kj}^i - 1)^2 \geq \left(\left(\frac{2h_{kj}^i}{s_{kj}^i - \delta_k \lambda_{kj}^i + 1}\right)^2, \quad k \in K, i \in N, j = 1,2 \right) \\
& \quad h_k^i = Q_k^i (\delta_k \lambda_k + \bar{u}), \quad k \in K, i \in N, \\
& \quad \bar{y} \leq \delta_k v_k^i - 1^T s_k^i, \quad k \in K, i \in N, \\
& \quad \delta_k \geq \frac{1}{\lambda_{\min}(\lambda_k)}, \quad k \in K, i \in N, \\
& \quad \delta_k \geq 0, \quad k \in K, \\
& \quad \bar{y} - \bar{y}_k \geq 0, \quad k \in K, \\
& \quad \bar{y} - \bar{y}_k \leq z_k, \quad k \in K, \\
& \quad d_2 \geq 0.
\end{align*}
\]

The objective function (3) expresses the minimization of the expected total cost; constraint (4) expresses an upper bound on the first stage radius, constraints (5) - (6) - (7) - (8) - (9) the inclusion of all the ellipses at scenario \( k \) in the second stage circle \( C_k \); constraints (10) - (11) represent a lower and upper-bound on the enlargement of the second stage circle \( C_k \) with respect to the first one \( C \). Finally, (12) represents a non-negativity constraint on the first-stage decision variable \( d_2 \).
4. An illustrative example: a United States TSP instance

In this section, in order to show the differences in the solutions and in the expected costs obtained by applying the Heuristic SL and SSOCP approaches, we provide an illustrative example based on the benchmark instance of the TSP having 13509 nodes to visit in the United States (see TSPLIB, instance name: usa13509.tsp).

We implement the SSOCP model in GAMS 22.5, by using the second order cone programming solver from the software package Mosek (http://www.mosek.com/), while the Heuristic SL is implemented in C++.

The two approaches are compared on the basis of a set of randomly generated problem instances as described in the following subsection. The results show that the solution obtained by the SSOCP model is significantly better than the heuristic SL solution.

4.1 Scenarios generation

The United States instance (usa13509.tsp) composed by 13509 nodes, has been divided into 4 different subareas according to the longitude of the points. Subsets of $q = 100$ nodes (possible customers) are then extracted in the 4 longitude areas by a uniform random distribution. Minimum covering ellipses

$$E_k^i = \{ u \in \mathbb{R}^2 : u^T H u + 2 g^T u + v \leq 0 \},$$

representing the traveling salesman $i$ at scenario $k$, visiting the subset of $q = 100$ random customers, have been generated under MATLAB 7.4.0 environment according to Sun and Freund (2004). Fig. 1 shows four traveling salesmen (the ellipses) supplying 400 customers randomly extracted from the United States instance (usa13509.tsp) composed by 13509 nodes. Red points represent the centres of the 4 ellipses. Notice that each customer is served by a unique traveling salesman and the $K$ scenarios are supposed to be equiprobable, with probability $1/K$.

Table 1. Region and cost solutions for an increasing number $K=10, 50, 100, 150, 200, 300, 400, 500$ of equiprobable scenarios.
In order to validate the model, we have first analyzed the sensitivity of the solutions to the number of scenarios from 50 up to 500. Fig. 2 and Table 1 show the stabilization of the objective function in the SSOCP model as the number of scenarios increases. From the results, we can deduce an in-sample stability, i.e. the optimal objective values are approximately the same when different numbers of scenarios are considered (for a definition of in-sample stability see Kaut & Wallace, 2007) and they stabilize around the value of 520.

### Table 1: Sensitivity analysis of the SSOCP model

<table>
<thead>
<tr>
<th>$K$</th>
<th>$d_2$</th>
<th>$\tilde{u}_1$</th>
<th>$\tilde{u}_2$</th>
<th>$\gamma$</th>
<th>$\tau$</th>
<th>$\varphi$</th>
<th>Cost $1^{st}$ stage</th>
<th>Cost $2^{nd}$ stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>13.54</td>
<td>-9.63</td>
<td>4.24</td>
<td>97.17</td>
<td>0</td>
<td>3.68</td>
<td>443.27</td>
<td>520.51</td>
</tr>
<tr>
<td>50</td>
<td>17.01</td>
<td>-9.74</td>
<td>4.06</td>
<td>94.18</td>
<td>0</td>
<td>4.14</td>
<td>556.89</td>
<td>592.43</td>
</tr>
<tr>
<td>100</td>
<td>13.75</td>
<td>-9.67</td>
<td>4.21</td>
<td>97.45</td>
<td>0</td>
<td>3.71</td>
<td>450.02</td>
<td>529.11</td>
</tr>
<tr>
<td>150</td>
<td>14.01</td>
<td>-9.74</td>
<td>4.16</td>
<td>98.10</td>
<td>0</td>
<td>3.74</td>
<td>458.40</td>
<td>521.58</td>
</tr>
<tr>
<td>200</td>
<td>14.34</td>
<td>-9.73</td>
<td>4.18</td>
<td>97.93</td>
<td>0</td>
<td>3.79</td>
<td>469.41</td>
<td>533.40</td>
</tr>
<tr>
<td>300</td>
<td>15.21</td>
<td>-9.73</td>
<td>4.19</td>
<td>97.97</td>
<td>0</td>
<td>3.77</td>
<td>465.16</td>
<td>528.97</td>
</tr>
<tr>
<td>400</td>
<td>13.70</td>
<td>-9.70</td>
<td>4.19</td>
<td>97.91</td>
<td>0</td>
<td>3.70</td>
<td>448.32</td>
<td>518.57</td>
</tr>
<tr>
<td>500</td>
<td>13.86</td>
<td>-9.71</td>
<td>4.19</td>
<td>97.96</td>
<td>0</td>
<td>3.72</td>
<td>453.55</td>
<td>521.73</td>
</tr>
</tbody>
</table>

Fig. 2 Sensitivity analysis (see Maggioni et al., 2010) of the first stage (dotted line) and total costs (solid line) values versus an increasing number of scenarios for the SSOCP model.

### 4.2 Comparison with the Heuristic SL

We now compare the SSOCP solution with the Heuristic SL in the case of 4 traveling salesmen ($N = 4$) and capacity $q = 100$ (see Table 2). The number of scenarios is $K = 34$, such that the total number of customers is $R = KNq = 13509$ as given in the TSPLIB usa13509.tsp. In order to make a consistent comparison between the two solutions approaches, the costs $\theta$ and $\phi$ in the SSOCP model are now computed according to the cost of the corresponding Capacitated Vehicle Routing Problem (CVRP), by solving the minimum least square problem explained in Bertazzi and Maggioni, (2012) (see Fig. 3). Table 2 shows the out-of-sample costs of the Heuristic SL solutions, computed as follows: different subsets of $q = 100$ possible customers are uniformly extracted in 4 different longitude areas from the South to the Nord of United States and the corresponding minimum covering ellipses are generated according to the procedure described in 4.1. The process is repeated for $K = 34$ scenarios.
such that all the 13509 customers are visited. The solution obtained by the **Heuristic SL** is then evaluated in terms of SSOCP costs (out-of-sample analysis, obtained by fixing the variables $\tilde{u} \in \mathbb{R}^k$ obtained in the **Heuristic SL** solution) and compared with the corresponding SSOCP total costs (see sixth and eighth column of Table 2). The result shows that the solution obtained by the **Heuristic SL** performs worse than the SSOCP formulation, forcing the facility to be fixed in one of the customers location in Indiana (see Fig. 4) instead of Nebraska as obtained in the SSOCP solution. Notice that, in order to reach all the customers, the service zone radius in the **Heuristic SL** has to be larger (4.47 instead of 3.82).

Fig. 3 Comparison of the costs $D_1(r)$, (a) and $D_2(r,r_k)$, (b) of the corresponding Capacitated Vehicle Routing Problem (CVRP), estimated by using the Daganzo formulas (1) - (2) (diamons) and the approximation by a minimum least square problem (squares) as described in Bertazzi and Maggioni (2012).

Fig. 4 Comparison of the SSOCP approach (-9.74,4.15) and Heuristic SL (-8.72,3.9) facility location solutions for the 13509 nodes instance for the United States. The SSOCP solution refers to the case of K=150 scenarios, N=4 traveling salesmen with capacity q=100.
Table 2. Comparison between the SSOCP model solution and out-of-sample evaluation of the Heuristic SL in the case of $R = 13509$ customers served by $N = 4$ traveling salesmen with capacity $q = 100$ and $K = 34$.

<table>
<thead>
<tr>
<th>$K$</th>
<th>$d_2$</th>
<th>$u_1$</th>
<th>$u_2$</th>
<th>$\gamma$</th>
<th>$\tau$</th>
<th>$R$</th>
<th>Cost 1$^{st}$ stage</th>
<th>Cost 2$^{nd}$ stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSOCP</td>
<td>14.62</td>
<td>- 9.74</td>
<td>4.14</td>
<td>97.32</td>
<td>0</td>
<td>3.82</td>
<td>478.46</td>
<td>530.22</td>
</tr>
<tr>
<td>SL</td>
<td>20</td>
<td>- 8.73</td>
<td>3.91</td>
<td>71.41</td>
<td>0</td>
<td>4.47</td>
<td>654.66</td>
<td>786.37</td>
</tr>
</tbody>
</table>

5. Conclusion

We consider the problem of a single facility serving a given number of customers in a given area. The position of the customers is not known. The service to the customers is carried out by several capacitated traveling salesmen. The aim is to determine the service zone (in a shape of a circle) that minimizes the expected cost of the traveled routes. The center of the circle is the location of the facility. Once the position of the customers is revealed, the customers located outside the service zone are served with a recourse action at a greater unit cost.

We compare the performance of two different approaches: the first one based on the heuristic proposed by Simchi-Levi (1991) and the second one as a two-stage stochastic second-order cone programming (SSOCP). The two models are then compared on a United States TSP instance with 13509 nodes in terms of solutions and costs and the advantages of the SSOCP formulation are discussed.

References


