Computing the Moments of Costs over the Solution Space of the TSP in Polynomial Time

Paul J. Sutcliffe, Andrew Solomon, and Jenny Edwards

Faculty of Information Technology
University of Technology, Sydney,
Sydney, Australia.
psutclif,andrews,jenny@it.uts.edu.au

Abstract. We give polynomial time algorithms to compute the third and fourth moments about the mean of tour costs over the solution space of the general symmetric Travelling Salesman Problem (TSP). These algorithms complement previous work on the population variance and provide a tractable method to compute the skewness and kurtosis of the probability distribution of tour costs. The methodology is generalisable to higher moments. Experimental evidence is given that suggests the skewness asymptotically approaches a limit point as the instance size is increased in several problem types.

1 Introduction

1.1 The TSP

The travelling salesman problem (TSP) is a classic problem in combinatorial optimization. Extensive references include [1–3]. Linear programming reductions are surveyed in [4] while the properties of frequently used local search heuristics are considered in [5]. It is natural to define the symmetric TSP in terms of a complete undirected graph \( \Gamma = (V, E) \) with the vertices \( V \) representing cities, and the edges \( E \) representing the connections between cities. We label the set of \( n \) vertices as \( \{1, 2, \ldots, n\} \), and an \( n \)-cycle permutation of these is a tour or solution, \( \pi \). The set of all tours, the solution space, is denoted \( \Theta \).

The distance between cities (or cost of an edge), is a function \( \text{cost} : E \rightarrow \mathbb{R} \) which we extend to the function \( \Omega : \Theta \rightarrow \mathbb{R} \), defined as the cost of a tour

\[
\Omega(\pi) = \sum_{i=1}^{n} \text{cost}(\{\pi(i), \pi((i \text{ mod } n) + 1)\}).
\]

The TSP is to find some \( n \)-cycle permutation \( \pi \) of \( V \) for which \( \Omega(\pi) \) is smallest. Such a permutation \( \pi^* \) is called a global minimum tour. If there are \( n \) cities then the number of tours is \( |\Theta| = (n - 1)!/2 \).

1.2 Survey of Statistical Results

Previous theoretical work on the probability distribution of the TSP is surveyed in [6, 7], these largely concern the case of the Euclidean TSP with city coordinates as \( n \) random variables in bounded subsets of \( \mathbb{R}^d \). Beardwood et. al. [8] prove that
\( \Omega(\pi^*) \) approaches a constant as \( n \to \infty \). Steele proves the variance of costs over the solution space is bounded [6]. Rhee and Talagrand prove that the tails of the cost distribution approach that of a Gaussian as the number of cities increases [9]. In the more general case Krauth and Mézard [10] extend the result of Beardwood et al. to problems with uniform random edge costs. More recently Wästlund [11] extends it to the TSP on bipartite graphs with uniform random edge costs.

Basel et al. [12] show by random sampling a remarkable linear correlation between the square root of a problem size and an estimate of the number of standard deviations between the mean tour cost and the known optimal tour cost in a real world set of approximately Euclidean problems. Sutcliffe et al. [13] give a constructive proof that the population variance of tour costs over the solution space of an instance of size \( n \) cities can be computed in \( O(n^2) \), see Theorem 1 below. Applying this, they confirm the linear relationship found by Basel et al. and show a similar, although non-linear, relationship in the case of a set of non-Euclidean real world problems.

1.3 Moments in Terms of the TSP

In terms of a TSP with solution space \( \Theta \), cost function \( \Omega \) and mean tour cost \( \mu \), the \( k \)th moment about the mean or central moment [14] can be written

\[
\text{mm}_k(\Theta) = \frac{\sum_{\pi \in \Theta} ((\Omega(\pi) - \mu)^k)}{|\Theta|}.
\] (1)

It is reported in [15] (and a simple proof follows from Lemma 1) that the mean tour cost over the solution space of a problem of size \( n \) cities with edge set \( E \) is \( \mu = \frac{2}{n-1} \sum_{e \in E} \text{cost}(e) \). The second moment or population variance is given by Theorem 1 below. Comparison of the second and third moments provides the well known statistic, the skewness, \( \alpha_3(X) = \frac{\text{mm}_3(X)}{\text{mm}_2(X)^{3/2}} \), which reflects the degree of symmetry of a probability distribution [14].

**Theorem 1.** The population variance of tour costs over the solution space of a TSP of size \( n \) cities and with edge set, \( E \) and vertex set \( V \) is

\[
\text{var} = \frac{2\beta_1}{(n-1)} - \frac{4\beta_1 + 2\beta_2}{(n-1)(n-2)}
\] (2)

with the values \( \beta_1, \beta_2 \) being defined as

\[
\beta_1 = \sum_{e \in E} c_0(e)^2
\]

\[
\beta_2 = \sum_{e = \{x,y\} \in E} [c_0(e)(S_x + S_y - 2c_0(e))]
\] (3)

where \( c_0(e) = \text{cost}(e) - \mu/n \), \( I_x \) is the set of edges incident to a vertex \( x \) with \( S_x = \sum_{e \in I_x} c_0(e) \) and similarly for \( S_y \).
2 The Third and Fourth Moment of Costs over the Solution Space

We begin with a technical lemma providing the number of tours containing various configurations of edges. Table 1 enumerates the eleven cases to be used.

**Lemma 1.** Given a TSP with graph $\Gamma$, let $P$ be a set of $m$, non-cyclic, non-singleton paths over $\Gamma$ sharing no vertices. Let $k$ be the number of vertices not appearing in any path of $P$. Then there are $2^{m-1}(k+m-1)!$ tours containing all the paths in $P$.

**Proof.** Label the paths of $P$, $p_j$ with $j \in [1 \ldots m]$. We recall that a tour is a cyclic permutation of vertices. Therefore, without loss of generality, fix $p_1$ in position and orientation and write a tour as $(p_1, i_1, i_2, \ldots, i_q, p_2, i_{q+1} \ldots, p_m \ldots, i_k)$. There are $(k+m-1)!$ orderings of the free paths and vertices. Each path is at least 2 vertices long and so each of the $m-1$ free paths has 2 orientations, implying the result. □

<table>
<thead>
<tr>
<th>case pattern</th>
<th>m</th>
<th>k</th>
<th>num. tours cities n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – – – – – – – – – – – – – –</td>
<td>$(n-2)$</td>
<td>$(n-2)!$</td>
<td>$n &gt; 2$</td>
</tr>
<tr>
<td>2 – – – – – – – – – – – – – –</td>
<td>$(n-3)$</td>
<td>$(n-3)!$</td>
<td>$n &gt; 2$</td>
</tr>
<tr>
<td>3 – – – – – – – – – – – – – –</td>
<td>$(n-4)$</td>
<td>$(n-4)!$</td>
<td>$n &gt; 3$</td>
</tr>
<tr>
<td>4 – – – – – – – – – – – – – –</td>
<td>$(n-4)$</td>
<td>$(n-4)!$</td>
<td>$n &gt; 3$</td>
</tr>
<tr>
<td>5 – – – – – – – – – – – – – –</td>
<td>$(n-5)$</td>
<td>$(n-4)!$</td>
<td>$n &gt; 4$</td>
</tr>
<tr>
<td>6 – – – – – – – – – – – – – –</td>
<td>$(n-6)$</td>
<td>$(n-4)!$</td>
<td>$n &gt; 5$</td>
</tr>
<tr>
<td>7 – – – – – – – – – – – – – –</td>
<td>$(n-5)$</td>
<td>$(n-5)!$</td>
<td>$n &gt; 4$</td>
</tr>
<tr>
<td>8 – – – – – – – – – – – – – –</td>
<td>$(n-6)$</td>
<td>$(n-5)!$</td>
<td>$n &gt; 5$</td>
</tr>
<tr>
<td>9 – – – – – – – – – – – – – –</td>
<td>$(n-6)$</td>
<td>$(n-5)!$</td>
<td>$n &gt; 6$</td>
</tr>
<tr>
<td>10 – – – – – – – – – – – – – –</td>
<td>$(n-7)$</td>
<td>$(n-5)!$</td>
<td>$n &gt; 6$</td>
</tr>
<tr>
<td>11 – – – – – – – – – – – – – –</td>
<td>$(n-8)$</td>
<td>$(n-5)!$</td>
<td>$n &gt; 7$</td>
</tr>
</tbody>
</table>

2.1 Computing the Third Moment

In order to prove our central theorem we provide some notational machinery. Let $\Theta$ be the solution space of a TSP with edge set $E$ and cost function $\Omega$. We index each $\pi$ in $\Theta$ with an integer $m \in [1 \ldots |\Theta|]$, similarly we label the edges of $E$ as
e_i with i ∈ [1 . . . |E|]. We define the function [1 . . . |Θ|] × [1 . . . |E|] : t → {0, 1} as

\[ t_{mi} = \begin{cases} 1 & \text{if edge } e_i \text{ is in tour } m \\ 0 & \text{otherwise} \end{cases} \]

Under this arrangement if m is the index of a tour π then the cost of π is

\[ \Omega(\pi) = t_{m1}\cos(e_1) + t_{m2}\cos(e_2) \ldots t_{m|E|}\cos(e_{|E|}), \]

and specializing (1) to k = 3, the third moment about the mean µ is

\[ \text{mm}_3(\Theta) = \frac{\sum_{m=1}^{\left|\Theta\right|} (t_{m1}\cos(e_1) + t_{m2}\cos(e_2) \ldots t_{m|E|}\cos(e_{|E|}) - \mu)^3}{\left|\Theta\right|}. \quad (4) \]

Now |Θ| is, of course, factorial on n and so this formulation is impractical for all but the smallest problems. In Theorem 2 we give a polynomial time solution to the problem.

Returning to notational matters, let A_p be the set of edges adjacent to edge e_p. Let N_p,q,. . . be the set of edges neither adjacent to nor equal to edges e_p, e_q, . . ., so N_p,q,. . . = E - (A_p ∪ e_p ∪ A_q ∪ e_q . . .).

**Theorem 2.** The third moment about the mean of tour costs over the solution space of a TSP with n > 3 cities, mean tour cost µ, and with edge set E is

\[ \text{mm}_3 = \frac{2\gamma_1}{(n-1)} + \frac{2(\gamma_2 + 2\gamma_3)}{(n-1)(n-2)} + \frac{2(\gamma_4 + 2\gamma_5 + 4\gamma_6)}{(n-1)(n-2)(n-3)} \]

with the values γ_1, γ_2, γ_3, γ_4, γ_5, γ_6 given by

\[
\begin{align*}
\gamma_1 &= \sum_{e \in E} c_0(e)^3 \\
\gamma_2 &= 3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q) \\
\gamma_3 &= 3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in N_p} c_0(e_q) \\
\gamma_4 &= 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_p - \{A_p \cup \{e_p\}\}} c_0(e_r) \\
\gamma_5 &= 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in N_{p,q}} c_0(e_r) \\
\gamma_6 &= \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in N_p} c_0(e_q) \sum_{e_r \in N_{p,q}} c_0(e_r)
\end{align*}
\]

where c_0(e) = cost(e) - µ/n.

**Proof.** Consider (4). Each tour has only n edges, so for a given m there are just n t_{mi} which are equal to 1, the remainder being equal to 0. So let c_0(e_i) = cost(e_i) - µ/n. Then (4) is written
\[
\text{mm}_3(\Theta) = \frac{1}{|\Theta|} \sum_{m=1}^{|\Theta|} \left( (t_{m_1}c_0(e_1) + t_{m_2}c_0(e_2) \ldots t_{m_{|E|}}c_0(e_{|E|}))^3 \right) \]

\[
= \frac{1}{|\Theta|} \sum_{m=1}^{|\Theta|} \sum_{k=1}^{|E|} \sum_{j=1}^{|E|} \sum_{i=1}^{|E|} t_{m_i}t_{m_j}t_{m_k}c_0(e_i)c_0(e_j)c_0(e_k) \]

The product \( t_{m_i}t_{m_j}t_{m_k} = 1 \), if and only if, tour \( m \) contains the edges \( e_i, e_j, e_k \) and there are six ways in which this can occur,

**case 1** All of \( e_i, e_j, e_k \) are equal. By Lemma 1 and Case 1 of Table 1 there are \((n - 2)!\) tours containing the edge.

**case 2** Two of \( e_i, e_j, e_k \) are equal and the third is adjacent. By Lemma 1 and Case 2 of Table 1 there are \((n - 3)!\) tours containing the three edges so configured.

**case 3** Two of \( e_i, e_j, e_k \) are equal and the third is non-adjacent to them. By Lemma 1 and Case 3 of Table 1 there are \(2(n - 3)!\) tours containing the two edges so configured.

**case 4** The three edges \( e_i, e_j, e_k \) form a path. By Lemma 1 and Case 4 of Table 1 there are \((n - 4)!\) tours containing the edges so configured.

**case 5** Two of \( e_i, e_j, e_k \) are adjacent and the third is non-adjacent to either. By Lemma 1 and Case 5 of Table 1 there are \(2(n - 4)!\) tours containing the three edges so configured.

**case 6** All \( e_i, e_j, e_k \) are all non-adjacent to each other. By Lemma 1 and Case 6 of Table 1 there are \(4(n - 4)!\) tours containing the three edges so configured.

For each of these six cases we write the sum of edge cost products as \( \gamma_1 \) to \( \gamma_6 \) in (2). Upon collecting like terms we have:

\[
\text{mm}_3(\Theta) = \frac{1}{|\Theta|} \left( ((n - 2)!\gamma_1 + (n - 3)!\gamma_2 + 2(n - 3)!\gamma_3 \right. \\
+ (n - 4)!\gamma_4 + 2(n - 4)!\gamma_5 + 4(n - 4)!\gamma_6) \\
\left. \right) / |\Theta| \\
= \frac{2((n - 2)!\gamma_1 + (n - 3)! \gamma_2 + 2\gamma_3 + (n - 4)! \gamma_4 + 2 \gamma_5 + 4 \gamma_6))}{(n - 1)!} \\
= \frac{2\gamma_1}{(n - 1)} + \frac{2(\gamma_2 + 2\gamma_3)}{(n - 1)(n - 2)} + \frac{2(\gamma_4 + 2\gamma_5 + 4\gamma_6)}{(n - 1)(n - 2)(n - 3)} \\
as required. \square
2.2 Reducing the Computational Complexity of Third Moment

The set $A_p$ is $O(n)$ in size, while the sets $E, N_p, N_{p,q}$ are all $O(n^2)$ in size. This implies that a naive application of Theorem 2 above would have complexity $O(n^6)$, being that of the sum $\gamma_6$. Here we show that this can be reduced to $O(n^4)$. Let $I_x$ be the set of edges incident the vertex $x$ and let $S_x = \sum_{e \in I_x} c_0(e)$, be the sum of edge costs incident to $x$. Now $|I_x| = n - 1$, so the time complexity of pre-computing all the $n$ values $S_x$ is $O(n^2)$ and the space complexity of saving them is $O(n)$.

**Lemma 2.** $\gamma_2$ can be found in $O(n^2)$

**Proof.** Recall that $\gamma_2 = 3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q)$. Consider the right most sum on $A_p$. We show this can be found in constant time. Writing each edge $e_p$, as $e_p = \{p_1, p_2\}$ and noting that $A_p = (I_{p_1} \cup I_{p_2}) - \{e_p\}$ gives,

$$\gamma_2 = 3 \sum_{e_p \in E} c_0(e_p)^2(S_{p_1} + S_{p_2} - 2c_0(e_p))$$

$$= 6\gamma_1 + 3 \sum_{e_p \in E} c_0(e_p)^2(S_{p_1} + S_{p_2})$$

This along with $|E| \in O(n^2)$ implies the result. \(\Box\)

**Lemma 3.** $\gamma_3 = -\gamma_2 - 3\gamma_1$

**Proof.** Recall that $\gamma_3 = 3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in N_p} c_0(e_q)$. Consider the right most sum, $N_p = E - (A_p \cup \{e_p\})$. So $\sum_{e \in N_p} c_0(e) = \sum_{e \in E} c_0(e) - \sum_{e \in A_p} c_0(e) - c_0(e_p)$, but $\sum_{e \in E} c_0(e) = 0$ thus

$$\gamma_3 = 3 \sum_{e_p \in E} c_0(e_p)^2 \left[ - \sum_{e \in A_p} c_0(e_q) - c_0(e_p) \right]$$

$$= -3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q) - 3 \sum_{e_p \in E} c_0(e_p)^2 c_0(e_p)$$

$$= -\gamma_2 - 3\gamma_1$$

As required. \(\Box\)

**Lemma 4.** $\gamma_4$ can be found in $O(n^3)$

**Proof.** Recall that $\gamma_4 = 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_q - (A_p \cup \{e_p\})} c_0(e_r)$. We show that the right most sum can be found in constant time given an $e_p$ and $e_q$. Let $e_p = \{s, p\}$, let $e_q = \{s, q\}$ be adjacent to it, sharing vertex $s$, and let $e_{pq} = \{p, q\}$ be adjacent to both. In addition let $I_q$ be the sets of edges incident to vertex $q$ and let $S_q$ be the pre-computed edge sum. Then $A_q - (A_p \cup \{e_p\}) = I_q - (\{e_q\} \cup \{e_{pq}\})$ and

$$\gamma_4 = 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) (S_q - c_0(e_q) - c_0(e_{pq}))$$

This along with $|E| \in O(n^2)$ and $|A_p| \in O(n)$ implies the result. \(\Box\)
Lemma 5. \( \gamma_5 \) can be found in \( O(n^3) \)

Proof. Recall that \( \gamma_5 = 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in N_{p,q}} c_0(e_r) \) We show that the right most sum can be found in constant time.

Let \( e_p = \{s,p\} \), let \( e_q = \{s,q\} \) be adjacent to it, sharing vertex \( s \), and let \( e_{pq} = \{p,q\} \) be adjacent to both. In addition let \( I_s, I_p, I_q \) be the sets of edges incident to the vertices \( s, p, q \) respectively and let \( S_s, S_p, S_q \) be the pre-computed edge sums. Now \( N_{p,q} = E - (I_s \cup I_p \cup I_q) \), but \( \sum_{e \in E} c_0(e) = 0 \) and the edges \( e_{pq}, e_q, e_p \) are each elements of two of \( I_s, I_p, I_q \) so, \( \sum_{e \in N_{p,q}} c_0(e) \) = \( c_0(e_p) + c_0(e_q) + c_0(e_{pq}) - S_s - S_p - S_q \) and

\[
\gamma_5 = 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \left[ c_0(e_p) + c_0(e_q) - S_s - S_p - S_q \right] = 6\gamma_2 + 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \left[ c_0(e_{pq}) - S_s - S_p - S_q \right].
\]

This along with \( |E| \in O(n^2) \) and \( |A_p| \in O(n) \) implies the result \( \square \)

Lemma 6. \( \gamma_6 \) can be found in \( O(n^4) \)

Proof. Recall that \( \gamma_6 = \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in N_p} c_0(e_q) \sum_{e_r \in N_{p,q}} c_0(e_r) \). We show that the right most sum can be found in constant time and that the middle sum can be rewritten over \( A_p \). Let \( e_p = \{p_1, p_2\} \) and let \( e_q = \{q_1, q_2\} \) be non adjacent. In addition let \( I_{p_1}, I_{p_2}, I_{q_1}, I_{q_2} \) be the sets of edges incident to these vertices and let \( S_{p_1}, S_{p_2}, S_{q_1}, S_{q_2} \) be the pre-computed edge sums. Now \( N_{p,q} = E - (I_{p_1} \cup I_{p_2} \cup I_{q_1} \cup I_{q_2}) \), but \( \sum_{e \in E} c_0(e) = 0 \) and the edges \( e_p, e_q, \{p_1, q_1\}, \{p_1, q_2\}, \{p_2, q_1\}, \{p_2, q_2\} \) are each elements of two of \( I_{p_1}, I_{p_2}, I_{q_1}, I_{q_2} \). Therefore write \( S_{N_{p,q}} = \sum_{e \in N_{p,q}} c_0(e) = -S_{p_1} - S_{p_2} - S_{q_1} - S_{q_2} + c_0(e_{p_1}) + c_0(e_{q_2}) + c_0(\{p_1, q_1\}) + c_0(\{p_1, q_2\}) + c_0(\{p_2, q_1\}) + c_0(\{p_2, q_2\}) \) and

\[
\gamma_6 = \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in N_p} c_0(e_q) S_{N_{p,q}},
\]

as required. \( \square \)

Theorem 3. The complexity of computing the third moment about the mean of tour costs over the solution space of a TSP with \( n \) cities is \( O(n^4) \).

Proof. This follows directly from Theorem 2, the comments at the beginning of Sect. 2.2, and Lemmas 2 to 6. \( \square \)
2.3 Computing the Fourth Moment about the Mean of Tour Costs

**Theorem 4.** The fourth moment about the mean of tour costs over the solution space of a TSP with mean tour cost \(\mu\), size \(n > 5\) cities and with edge set \(E\) is

\[
\text{mm}_4 = \frac{2\delta_1}{(n-1)} + \frac{2(\delta_{2a} + \delta_{2b} + 2\delta_{3a} + 2\delta_{3b})}{(n-1)(n-2)}
\]

\[
+ \frac{2(\delta_{4a} + \delta_{4b} + 2\delta_{5a} + 2\delta_{5b} + 4\delta_{6})}{(n-1)(n-2)(n-3)}
\]

\[
+ \frac{2(\delta_7 + 2\delta_8 + 4\delta_{10} + 8\delta_{11})}{(n-1)(n-2)(n-3)(n-4)}
\],

with the values \(\delta_1, \delta_{2a}, \delta_{2b}, \delta_{3a}, \delta_{3b}, \delta_{4a}, \delta_{4b}, \delta_{5a}, \delta_{5b}, \delta_6, \delta_7, \delta_8, \delta_9, \delta_{10}, \delta_{11}\) given by

\[
\delta_1 = \sum_{e \in E} c_0(e)^4
\]

\[
\delta_{2a} = 3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q)^2
\]

\[
\delta_{2b} = 4 \sum_{e_p \in E} c_0(e_p)^3 \sum_{e_q \in A_p} c_0(e_q)
\]

\[
\delta_{3a} = 3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q)^2
\]

\[
\delta_{3b} = 4 \sum_{e_p \in E} c_0(e_p)^3 \sum_{e_q \in A_p} c_0(e_q)
\]

\[
\delta_{4a} = 12 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_p} c_0(e_r)
\]

\[
\delta_{4b} = 6 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q)^2 \sum_{e_r \in A_p} c_0(e_r)
\]

\[
\delta_{5a} = 12 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_p} c_0(e_r)
\]

\[
\delta_{5b} = 6 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_p} c_0(e_r)^2
\]

\[
\delta_6 = 6 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_p} c_0(e_r)
\]
\[ \delta_7 = 12 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_q} c_0(e_r) \sum_{e_s \in A_r} c_0(e_s) \]
\[ \delta_8 = 12 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_q \setminus e_p} c_0(e_r) \sum_{e_s \in A_r \setminus e_q} c_0(e_s) \]
\[ \delta_9 = 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in N_p \setminus e_q} c_0(e_r) \sum_{e_s \in N_p \setminus e_r} c_0(e_s) \]
\[ \delta_{10} = 6 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in N_p \setminus e_q} c_0(e_r) \sum_{e_s \in N_p \setminus e_r} c_0(e_s) \]
\[ \delta_{11} = \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in N_p \setminus e_p} c_0(e_q) \sum_{e_r \in N_p \setminus e_q} c_0(e_r) \sum_{e_s \in N_p \setminus e_r} c_0(e_s) \]

where \( c_0(e) = \text{cost}(e) - \mu/n. \)

**Proof.** Specializing (1) to \( k = 4 \), and proceeding as we did for the third moment we have

\[
\text{mm}_4(\Theta) = \frac{\sum_{m=1}^{|E|} \sum_{l=1}^{|E|} l_m l_n l_m l_n c_0(e_i) c_0(e_j) c_0(e_k) c_0(e_l)}{|\Theta|}.
\]

The product \( l_m l_n l_m l_n = 1 \) if and only if tour \( m \) contains the edges \( e_1, e_4, e_9, e_{11} \) and there are eleven ways in which this can occur.

**case 1** All of \( e_1, e_4, e_9, e_{11} \) are equal. By Lemma 1 there are \((n - 2)!\) tours containing the edge. The value \( \delta_1 \) is the sum of terms in this case.

**case 2** From \( e_1, e_4, e_9, e_{11} \) there are 2 distinct edges and they form a path. By Lemma 1 there are \((n - 3)!\) tours containing the edges. The values \( \delta_{2a}, \delta_{2b} \) are the sums of terms in this case.

**case 3** From \( e_1, e_4, e_9, e_{11} \) there are 2 distinct edges and they are non adjacent. By Lemma 1 there are \(2(n - 3)!\) tours containing the edges. The values \( \delta_{3a}, \delta_{3b} \) are the sums of terms in this case.

**case 4** From \( e_1, e_4, e_9, e_{11} \) there are 3 distinct edges and they form a path. By Lemma 1 there are \((n - 4)!\) tours containing the edges. The values \( \delta_{4a}, \delta_{4b} \) are the sums of terms in this case.

**case 5** From \( e_1, e_4, e_9, e_{11} \) there are 3 distinct edges two of which form a path. The third is non adjacent. By Lemma 1 there are \(2(n - 4)!\) tours containing the edges. The values \( \delta_{5a}, \delta_{5b} \) are the sums of terms in this case.

**case 6** From \( e_1, e_4, e_9, e_{11} \) there are 3 distinct. All are non adjacent. By Lemma 1 there are \(4(n - 4)!\) tours containing the edges. The value \( \delta_6 \) is the sum of terms in this case.

**case 7** Each of \( e_1, e_4, e_9, e_{11} \) are distinct and form a path. By Lemma 1 there are \((n - 5)!\) tours containing the edges. The value \( \delta_7 \) is the sum of terms in this case.
case 8 Each of $e_i, e_j, e_k, e_l$ are distinct, 3 form a path, the other is non adjacent.
By Lemma 1 there are $2(n - 5)!$ tours containing the edges. The value $\delta_8$ is the sum of terms in this case.

case 9 Each of $e_i, e_j, e_k, e_l$ are distinct and form 2 non adjacent paths of 2 edges.
By Lemma 1 there are $2(n - 5)!$ tours containing the edges. The value $\delta_9$ is the sum of terms in this case.

case 10 Each of $e_i, e_j, e_k, e_l$ are distinct. Two are adjacent. The remaining are non adjacent.
By Lemma 1 there are $4(n - 5)!$ tours containing the edges. The value $\delta_{10}$ is the sum of terms in this case.

case 11 Each of $e_i, e_j, e_k, e_l$ are distinct. All are non adjacent.
By Lemma 1 there are $8(n - 5)!$ tours containing the edges. The value $\delta_{11}$ is the sum of terms in this case.

For each of these cases we write the sum of edge cost products as $\delta_1$ to $\delta_{11}$ in (4). Upon collecting like terms we have

$$\text{mm}_4(\Theta) = \frac{(n-2)!\delta_3}{|\Theta|} + \frac{(n-3)!\delta_5 + (n-3)!\delta_6 + 2(n-3)!\delta_8 + 2(n-3)!\delta_{10}}{|\Theta|}$$

Recall $|\Theta| = (n - 1)!/2$ so upon cancelation we have the result.

\[\square\]

3 Empirical Examination of the Relationship between Skewness and Problem Size

We examine four problem sets, two real world and two randomly generated. The four types are summarized in Table 2.

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Size in Cities</th>
<th>Cases</th>
<th>Problem Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Euclidean</td>
<td>10-1000</td>
<td>21</td>
<td>2 Euclidean Metric of TSPLIB [16].</td>
</tr>
<tr>
<td>VLSI</td>
<td>131-984</td>
<td>10</td>
<td>2 Euclidean Metric of TSPLIB.</td>
</tr>
<tr>
<td>Random no embed.</td>
<td>10-1000</td>
<td>21</td>
<td>Random integer edge costs from $U(0,999)$</td>
</tr>
</tbody>
</table>

Of the real world sets the first set originated in the production of very large scale integrated circuits (VLSI) and uses the 2 dimensional Euclidean metric of [16]. The second set, of 39 instances, approximately obey the triangular inequality, but are non-Euclidean. They originate in the genomics community and arise from physical mapping of canine DNA by the radiation-hybrid (RH) method. The specific data set used was obtained from the RHDF9000 dog radiation hybrid panel[17].
3.1 Results

The skewness of each instance was found using Theorems 1 and 2 in conjunction with Lemmas 2 to 6. Figure 1 shows its relationship to problem size. The relationship suggests, that in the case of the non RH data sets the skewness asymptotically approaches 0 with size. The RH data set is somewhat suggestive of convergence but to a lower limit point.

4 Conclusions and Future Work

In this paper we have given constructive proofs that the third central moment of tour costs over the solution space of any instance of a TSP of size $n$ cities can be computed in $O(n^4)$ and the that fourth central moment can be computed in $O(n^8)$. Experience with the third moment would suggest this computational complexity may be reduced to $O(n^6)$.

The method can be generalised to higher moments (at increased cost) and to variations of the problem such as the asymmetrical TSP.

Previous theoretical work on the probability distribution of the TSP was largely confined to the Euclidean case and did not extend to providing the moments. Future work will investigate the role of the third and fourth moments in refining current methods to estimate the optimal solution cost and to understanding the solution space of the problem.

Experimental evidence is given suggesting that the skewness asymptotically approaches 0 as the problem size is increased, in randomly generated non-embeddable and both random and real world 2 dimensional Euclidean instances. This implies that in these problem types, the distribution of tour costs become more symmetric as the problem size increases. This may make it possible to find bounds on the value of the odd moments of the cost distribution in certain classes of problem.
Acknowledgements

The authors would like to thank Andre Rohe, Simon de Givry, Thomas Schiex, Christophe Hitte and Jill Maddox for their assistance in obtaining the real world data sets.

References