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*, π. The set of all tours, the *solution space*, is denoted Θ. The distance between cities (or *cost* of an edge), is a function cost : $E \to \Re$ which we extend to the function $\Omega : \Theta \to \Re$, defined as the cost of a tour $\Omega(\pi) = \sum_{i=1}^{n} \mathrm{cost}(\{\pi(i), \pi((i \mod n) + 1)\})$.

The TSP is to find some n-cycle permutation π of V for which $\Omega(\pi)$ is smallest. Such a permutation π^* 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 $\Re^d$. 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 Θ, cost function Ω and mean tour cost μ, the kth *moment about the mean* or *central moment* [14] can be written

$$\mathrm{mm_k}(\Theta) = \frac{\sum\limits_{\pi \in \Theta} ((\Omega(\pi) - \mu)^k)}{|\Theta|}. \tag{1}$$

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 $\mu = \frac{2}{n-1} \sum\limits_{e \in E} \mathrm{cost}(e)$. 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*, $\alpha_3(X) = \frac{\mathrm{mm}_3(X)}{\mathrm{mm}_2(X)^{3/2}}$, 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*

$$\mathrm{var} = \frac{2\beta_1}{(n-1)} - \frac{4\beta_1 + 2\beta_2}{(n-1)(n-2)} \tag{2}$$

with the values β_1, β_2 being defined as

$$\begin{aligned} \beta_1 &= \sum_{e \in E} c_0(e)^2 \\ \beta_2 &= \sum_{e=\{x,y\} \in E} [c_0(e)(S_x + S_y - 2c_0(e))] \end{aligned} \tag{3}$$

where $c_0(e) = \mathrm{cost}(e) - \mu/n$, 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 Γ, let P be a set of m, non-cyclic, non-singleton paths over Γ 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 $j \in [1 \dots m]$. 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 $(p_1, i_1, i_2 \dots, i_q, p_2, i_{i_q+1} \dots, p_m \dots, i_k)$. 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 1. The eleven ways that, up to four unlabelled edges, can be arranged into paths in tours of size n. The $-$ character represent an edge, so $--$ means a path with 2 edges and three vertices. The $\leftrightsquigarrow$ symbol, the set (possibly empty) of free vertices between unconnected paths. The number of paths is given by m, while k is the number of free vertices.

case	pattern	m	k	num. tours	cities n
1	$- \leftrightsquigarrow$	1	$(n-2)$	$(n-2)!$	$n>2$
2	$-- \leftrightsquigarrow$	1	$(n-3)$	$(n-3)!$	$n>2$
3	$- \leftrightsquigarrow - \leftrightsquigarrow$	2	$(n-4)$	$2(n-3)!$	$n>3$
4	$--- \leftrightsquigarrow$	1	$(n-4)$	$(n-4)!$	$n>3$
5	$-- \leftrightsquigarrow - \leftrightsquigarrow$	2	$(n-5)$	$2(n-4)!$	$n>4$
6	$- \leftrightsquigarrow - \leftrightsquigarrow - \leftrightsquigarrow$	3	$(n-6)$	$4(n-4)!$	$n>5$
7	$---- \leftrightsquigarrow$	1	$(n-5)$	$(n-5)!$	$n>4$
8	$--- \leftrightsquigarrow - \leftrightsquigarrow$	2	$(n-6)$	$2(n-5)!$	$n>5$
9	$-- \leftrightsquigarrow -- \leftrightsquigarrow$	2	$(n-6)$	$2(n-5)!$	$n>6$
10	$-- \leftrightsquigarrow - \leftrightsquigarrow - \leftrightsquigarrow$	3	$(n-7)$	$4(n-5)!$	$n>6$
11	$- \leftrightsquigarrow - \leftrightsquigarrow - \leftrightsquigarrow - \leftrightsquigarrow$	4	$(n-8)$	$8(n-5)!$	$n>7$

2.1 Computing the Third Moment

In order to prove our central theorem we provide some notational machinery. Let Θ be the solution space of a TSP with edge set E and cost function Ω. We index each π in Θ with an integer $m \in [1 \dots |\Theta|]$, similarly we label the edges of E as

e_i with $i \in [1 \ldots |E|]$. We define the function $[1 \ldots |\Theta|] \times [1 \ldots |E|] : t \rightarrow \{0,1\}$ 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

$$\Omega(\pi) = t_{m1}\text{cost}(e_1) + t_{m2}\text{cost}(e_2) \ldots t_{m|E|}\text{cost}(e_{|E|}) \ ,$$

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

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

Now $|\Theta|$ 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 $N_{p,q,\ldots}$ be the set of edges neither adjacent to nor equal to edges $e_p, e_q, \ldots$, so $N_{p,q,\ldots} = E - (A_p \bigcup \{e_p\} \bigcup A_q \bigcup \{e_q\} \ldots)$.

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 $\gamma_1, \gamma_2, \gamma_3\gamma_4, \gamma_5, \gamma_6$ given by

$$\begin{aligned}
\gamma_1 &= \sum_{e \in E} c_0(e)^3 \\
\gamma_2 &= 3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q) \\
\gamma_3 &= 3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in N_p} c_0(e_q) \\
\gamma_4 &= 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_q - (A_p \bigcup \{e_p\})} c_0(e_r) \\
\gamma_5 &= 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in N_{p,q}} c_0(e_r) \\
\gamma_6 &= \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in N_p} c_0(e_q) \sum_{e_r \in N_{p,q}} c_0(e_r)
\end{aligned}$$

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

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 $c_0(e_i) = \text{cost}(e_i) - \mu/n$. Then (4) is written

$$\mathrm{mm}_3(\Theta) = \frac{\sum_{m=1}^{|\Theta|}((t_{m1}c_0(e_1) + t_{m2}c_0(e_2)\dots t_{m|E|}c_0(e_{|E|}))^3)}{|\Theta|} \quad ,$$

$$= \frac{\sum_{m=1}^{|\Theta|}\sum_{k=1}^{|E|}\sum_{j=1}^{|E|}\sum_{i=1}^{|E|} t_{mi}t_{mj}t_{mk}c_0(e_i)c_0(e_j)c_0(e_k)}{|\Theta|} \quad .$$

The product $t_{mi}t_{mj}t_{mk} = 1$, if and only if, tour m contains the edges e_i, e_j, e_k and there are six way 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 2 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 there six cases we write the sum of edge cost products as γ_1 to γ_6 in (2). Upon collecting like terms we have:

$$\begin{aligned}\mathrm{mm}_3(\Theta) =&((n-2)!\gamma_1 + (n-3)!\gamma_2 + 2(n-3)!\gamma_3 \\ &+ (n-4)!\gamma_4 + 2(n-4)!\gamma_5 + 4(n-4)!\gamma_6)/|\Theta|\end{aligned}$$

$$= \frac{2((n-2)!\gamma_1 + (n-3)!(\gamma_2 + 2\gamma_3) + (n-4)!(\gamma_4 + 2\gamma_5 + 4\gamma_6))}{(n-1)!}$$

$$= \frac{2\gamma_1}{(n-1)} + \frac{2(\gamma_2 + 2\gamma_3)}{(n-1)(n-2)} + \frac{2(\gamma_4 + 2\gamma_5 + 4\gamma_6)}{(n-1)(n-2)(n-3)} \quad .$$

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 γ_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. *γ_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 $A_p = (I_{p1} \bigcup I_{p2}) - \{e_p\}$ gives,

$$\begin{aligned}\gamma_2 &= 3 \sum_{e_p \in E} c_0(e_p)^2 (S_{p1} + S_{p2} - 2c_0(e_p)) \\ &= 6\gamma_1 + 3 \sum_{e_p \in E} c_0(e_p)^2 (S_{p1} + S_{p2}) \ .\end{aligned}$$

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

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, $N_p = E - (A_p \bigcup \{e_p\})$. 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

$$\begin{aligned}\gamma_3 &= 3 \sum_{e_p \in E} c_0(e_p)^2 \left[- \sum_{e_q \in A_p} c_0(e_q) - c_0(e_p) \right] \\ &= -3 \sum_{e_p \in E} c_0(e_p)^2 \sum_{e_q \in A_p} c_0(e_q) - 3 \sum_{e_p \in E} c_0(e_p)^2 c_0(e_p) \\ &= -\gamma_2 - 3\gamma_1 \ .\end{aligned}$$

As required. □

Lemma 4. *γ_4 can be found in $O(n^3)$*

Proof. Recall that $\gamma_4 = 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in A_q - (A_p \bigcup \{e_p\})} c_0(e_r)$. 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 $A_q - (A_p \bigcup \{e_p\}) = I_q - (\{e_q\} \bigcup \{e_{pq}\})$ 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. □

Lemma 5. γ_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 $N_{p,q} = E - (I_s \bigcup I_p \bigcup I_q)$, 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, $\sum_{e_r \in N_{p,q}} c_0(e_r) = c_0(e_p) + c_0(e_q) + c_0(e_{pq}) - S_s - S_p - S_q$ and

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

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

Lemma 6. γ_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 the 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 $I_{p1}, I_{p2}, I_{q1}, I_{q2}$ be the sets of edges incident to these vertices and let $S_{p1}, S_{p2}, S_{q1}, S_{q2}$ be the pre-computed edge sums. Now $N_{p,q} = E - (I_{p1} \bigcup I_{p2} \bigcup I_{q1} \bigcup I_{q2})$, 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 $I_{p1}, I_{p2}, I_{q1}, I_{q2}$. Therefore write $S_{Np,q} = \sum_{e \in N_{p,q}} c_0(e) = -S_{p1} - S_{p2} - S_{q1} - S_{q2} + c_0(e_p) + c_0(e_q) + c_0(\{p_1, q_1\}) + c_0(\{p_1, q_2\}) + c_0(\{p_2, q_1\}) + c_0(\{p_2, q_2\})$ and

$$\gamma_6 = \sum_{e_p \in E} \left[c_0(e_p) \sum_{e_q \in N_p} c_0(e_q) S_{Np,q} \right],$$

as required. □

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. □

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 μ, size $n > 5$ cities and with edge set E is*

$$\begin{aligned}\mathrm{mm}_4 = \frac{2\delta_1}{(n-1)} + \frac{2(\delta_{2a}+\delta_{2b}+2\delta_{3a}+2\delta_{3b})}{(n-1)(n-2)} \\ + \frac{2(\delta_{4a}+\delta_{4b}+2\delta_{5a}+2\delta_{5b}+4\delta_6)}{(n-1)(n-2)(n-3)} \\ + \frac{2(\delta_7+2\delta_8+2\delta_9+4\delta_{10}+8\delta_{11})}{(n-1)(n-2)(n-3)(n-4)}\end{aligned} \quad ,$$

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*

$$\begin{aligned}
\delta_1 &= \sum_{e\in E} c_0(e)^4 \\
\delta_{2a} &= 3\sum_{e_p\in E} c_0(e_p)^2 \sum_{e_q\in A_p} c_0(e_q)^2 \\
\delta_{2b} &= 4\sum_{e_p\in E} c_0(e_p)^3 \sum_{e_q\in A_p} c_0(e_q) \\
\delta_{3a} &= 3\sum_{e_p\in E} c_0(e_p)^2 \sum_{e_q\in N_p} c_0(e_q)^2 \\
\delta_{3b} &= 4\sum_{e_p\in E} c_0(e_p)^3 \sum_{e_q\in N_p} c_0(e_q) \\
\delta_{4a} &= 12\sum_{e_p\in E} c_0(e_p)^2 \sum_{e_q\in A_p} c_0(e_q) \sum_{\substack{e_r\in A_q,\\ e_r\notin A_p,\\ e_r\neq e_p}} c_0(e_r) \\
\delta_{4b} &= 6\sum_{e_p\in E} c_0(e_p) \sum_{e_q\in A_p} c_0(e_q)^2 \sum_{\substack{e_r\in A_q,\\ e_r\notin A_p,\\ e_r\neq e_p}} c_0(e_r) \\
\delta_{5a} &= 12\sum_{e_p\in E} c_0(e_p)^2 \sum_{e_q\in A_p} c_0(e_q) \sum_{e_r\in N_{p,q}} c_0(e_r) \\
\delta_{5b} &= 6\sum_{e_p\in E} c_0(e_p) \sum_{e_q\in A_p} c_0(e_q) \sum_{e_r\in N_{p,q}} c_0(e_r)^2 \\
\delta_6 &= 6\sum_{e_p\in E} c_0(e_p)^2 \sum_{e_q\in N_p} c_0(e_q) \sum_{e_r\in N_{p,q}} c_0(e_r)
\end{aligned}$$

$$\delta_7 = 12 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{\substack{e_r \in A_q, \\ e_r \notin A_p, \\ e_r \neq e_p}} c_0(e_r) \sum_{\substack{e_s \in A_r, \\ e_s \notin A_q, \\ e_s \notin A_p}} c_0(e_s)$$

$$\delta_8 = 12 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{\substack{e_r \in A_q, \\ e_r \notin A_p, \\ e_r \neq e_p}} c_0(e_r) \sum_{e_s \in N_{p,q,r}} c_0(e_s)$$

$$\delta_9 = 3 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in Np,q} c_0(e_r) \sum_{\substack{e_s \in A_r, \\ e_s \in N_{p,q}}} c_o(e_s)$$

$$\delta_{10} = 6 \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in A_p} c_0(e_q) \sum_{e_r \in Np,q} c_0(e_r) \sum_{e_s \in N_{p,q,r}} c_0(e_s)$$

$$\delta_{11} = \sum_{e_p \in E} c_0(e_p) \sum_{e_q \in N_p} c_0(e_q) \sum_{e_r \in Np,q} c_0(e_r) \sum_{e_s \in N_{p,q,r}} c_0(e_s)$$

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

$$\text{mm}_4(\Theta) = \frac{\sum_{m=1}^{|\Theta|} \sum_{l=1}^{|E|} \sum_{k=1}^{|E|} \sum_{j=1}^{|E|} \sum_{i=1}^{|E|} t_{mi} t_{mj} t_{mk} t_{ml} \text{c}_0(e_i) \text{c}_0(e_j) \text{c}_0(e_k) \text{c}_0(e_l)}{|\Theta|} .$$

The product $t_{mi} t_{mj} t_{mk} t_{ml} = 1$ if and only if tour m contains the edges e_i, e_j, e_k, e_l and there are eleven ways in which this can occur.

case 1 All of e_i, e_j, e_k, e_l are equal. By Lemma 1 there are $(n-2)!$ tours containing the edge. The value δ_1 is the sum of terms in this case.

case 2 From e_i, e_j, e_k, e_l there are 2 distinct edges and they form a path. By Lemma 1 there are $(n-3)!$ tours containing the edges. The values δ_{2a}, δ_{2b} are the sums of terms in this case.

case 3 From e_i, e_j, e_k, e_l there are 2 distinct edges and they are non adjacent. By Lemma 1 there are $2(n-3)!$ tours containing the edges. The values δ_{3a}, δ_{3b} are the sums of terms in this case.

case 4 From e_i, e_j, e_k, e_l there are 3 distinct edges and they form a path. By Lemma 1 there are $(n-4)!$ tours containing the edges. The values δ_{4a}, δ_{4b} are the sums of terms in this case.

case 5 From e_i, e_j, e_k, e_l 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 δ_{5a}, δ_{5b} are the sums of terms in this case.

case 6 From e_i, e_j, e_k, e_l there are 3 distinct. All are non adjacent. By Lemma 1 there are $4(n-4)!$ tours containing the edges. The value δ_6 is the sum of terms in this case.

case 7 Each of e_i, e_j, e_k, e_l are distinct and form a path. By Lemma 1 there are $(n-5)!$ tours containing the edges. The value δ_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 δ_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 δ_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 δ_{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 δ_{11} is the sum of terms in this case.

For each of these cases we write the sum of edge cost products as δ_1 to δ_{11} in (4). Upon collecting like terms we have

$$\begin{aligned}\mathrm{mm}_4(\Theta) = & \tfrac{(n-2)!\delta_1}{|\Theta|} \\ & + \tfrac{(n-3)!\delta_{2a}+(n-3)!\delta_{2b}+2(n-3)!\delta_{3a}+2(n-3)!\delta_{3b}}{|\Theta|} \\ & + \tfrac{(n-4)!\delta_{4a}+(n-4)!\delta_{4b}+2(n-4)!\delta_{5a}+2(n-4)!\delta_{5b}+4(n-4)!\delta_{6}}{|\Theta|} \\ & + \tfrac{(n-5)!\delta_{7}+2(n-5)!\delta_{8}+2(n-5)!\delta_{9}+4(n-5)!\delta_{10}+8(n-5)!\delta_{11}}{|\Theta|} \ .\end{aligned}$$

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

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 2. Problem types

ProblemType	Size in Cities	Cases	Problem Description
Random Euclidean	10-1000	21	2 Euclidean Metric of TSPLIB [16].
VLSI	131-984	10	2 Euclidean Metric of TSPLIB.
Random no embed.	10-1000	21	Random integer edge costs from $U(0, 999)$
RH Data	68-662	39	Non Euclidean. Genomics problems.

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].

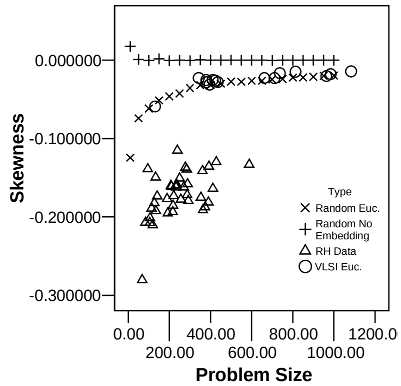

Fig. 1. The skewness versus the problem size in four problem types.

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

1. Gutin, G., Punnen, A.P.: Traveling Salesman Problem and Its Variations. Kluwer Academic Publishers (2002)
2. Ausiello, G., Protasi, M., Marchetti-Spaccamela, A., Gambosi, G., Crescenzi, P., Kann, V.: Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1999)
3. Reinelt, G.: The traveling salesman: Computational solutions for TSP applications. Springer Verlag (1994) LNCS 840.
4. Orman, A.J., Williams, H.P.: A survey of different integer programming formulations of the travelling salesman problem. Working Paper LSEOR 04.67, Department of Operational Research, London School of Economics and Political Science, London (2004)
5. Colletti, B., Barnes, J.: Local search structure in the symmetric travelling salesperson problem under a general class of rearrangement neighborhoods. Applied Mathematics Letters. **14**(1) (2001) 105–108
6. Steele, J.M.: Probability Theory and Combinatorial Optimization. SIAM (1997)
7. Yukich, J.E.: Probability Theory of Classical Euclidean Optimization Problems. Volume 1675 of Lecture Notes in Mathematics. Springer (1998)
8. Beardwood, J., Halton, J., Hammersley, J.: The shortest path through many points. Proc. Cambridge Philos. Soc. **55** (1959) 299–327
9. Rhee, W.T., Talgrand, M.: A sharp deviation inequality for the stochastic traveling salesman problem. Annals of Probability **17** (1989) 1–8
10. Krauth, W., Mézard, M.: The cavity method and the travelling-salesman problem. Europhys. Lett. **8**(3) (1988)
11. Wästlund, J.: The limit in the mean field bipartite travelling salesman problem. unpublished 2006, wwww.mai.liu.se/jowas
12. John Basel, I., Willemain, T.R.: Random tours in the traveling salesman problem analysis and application. Comput. Optim. Appl. **20**(2) (2001) 211–217
13. Sutcliffe, P., Solomon, A., Edwards, J.: Finding the population variance of costs over the solution space of the tsp in polynomial time. In Psarris, K., Jones, A.D., eds.: Math 07, Proceedings of the Eleventh WSEAS International Conference on Applied Mathematics, WSEAS (March 22-24, 2007) 23–28
14. Freund, J.E.: Mathematical Statistics. First edn. Prentice Hall (1972)
15. Punnen, A., Margot, F., Kabadi, S.: Tsp heuristics: Domination analysis and complexity. Technical report, Dept. of Mathematics, Univ. of Kentucky (2001)
16. Reinelt, G.: TSPLIB — a traveling salesman problem library. ORSA Journal on Computing **3**(4) (1991) 376–384
17. Faraut, T., de Givry, S., Chabrier, P., Derrien, T., Galibert, F., Hitte, C., Schiex, T.: A comparative genome approach to marker ordering. In: Proc. of ECCB-06. (2007) 7p.