

Moments over the Solution Space
of the
Travelling Salesman Problem

by

Paul John Sutcliffe

A thesis submitted in partial fulfillment of the requirements for
the degree of

Doctor of Philosophy in Computer Science
Faculty of Engineering and Information Technology
University of Technology, Sydney

Certificate of Authorship and Originality

I, Paul John Sutcliffe certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Paul John Sutcliffe
The candidate and author

Dedication

To my wife, my parents and sisters

Acknowledgments

Firstly, I would like to thank my supervisors Professor Jenny Edwards and Doctor Andrew Solomon. Andrew, thanks for everything you taught me. Also thanks for not groaning too loudly at the clumsiness of some of these proofs. I'm hoping our next collaboration will be Über slick. Jenny thanks for your help in all things and particularly for preventing me from “impaling” when I hoped to “imply” and “exiting” when “existing” was intended.

Sincere thanks to the staff and students of the operations research and industrial engineering group at the University of Texas at Austin for making my brief stay there so fruitful. I am particularly grateful for the guidance and encouragement of Professor J. Wesley Barnes and Doctor Bruce Colletti.

I am indebted to Simon de Givry, Thomas Schiex and Christophe Hitte, for their kind provision of the canine RH data set and for its translation into TSP instances.

I would like to extend my thanks to David Plyler, Len Hill, John Bance and Doctor Neil Thomson.

I am grateful for the financial support in the form of a scholarship from the University of Technology Sydney.

Finally, I thank my wife Lisa for her love and extreme patience.

Abstract

In this thesis we consider the statistical properties of the symmetric travelling salesman problem (TSP). Previous work on the statistical properties of the problem has been largely limited to the Euclidean case with vertex coordinates as random variables with known distribution embedded in \mathbb{R}^d , and to the case of *independent identically distributed* random edge costs. Furthermore, this previous work did not extend to computing the moments, beyond the mean. In the work presented here we consider the more general case of problem instances specified as a set of edge costs and with no (known) embedding or coordinate system available.

For an instance of the problem on n vertices with fixed edge costs we give constructive proofs that the population variance of tour costs over the solution space can be computed in $\mathcal{O}(n^2)$, the third central moment can be computed in $\mathcal{O}(n^4)$ and the fourth central moment can be computed in $\mathcal{O}(n^6)$. These results provide direct methods to compute the moments about the origin and factorial moments of these orders with the corresponding computational complexity. In addition the results provide tractable methods to compute, among other statistics, the standard deviation of tour costs, the skewness of the probability distributions of tour costs over the solution space and kurtosis of this distribution.

In the case of the stochastic TSP with *edge costs* defined as independently distributed random variables with (not necessarily identical) known mean and variance we provide a $\mathcal{O}(n^4)$ algorithm to compute the variance of tour costs.

Given a subgraph S of a tour in an n city TSP, we provide an $\mathcal{O}(n^2)$ algorithm to compute the expected tour costs over the solution space of those tours containing S . This is useful in analysing and constructing algorithms such as Gutin's greedy expectation heuristic.

We demonstrate that the probability distribution of gains over the 2-opt landscape of an n city TSP can be computed in $\mathcal{O}(n^4 \log(n))$. This result provides a tractable algorithm to compute, among other statistics the moments of gains over the landscape. The result also provides the 2-opt neutrality (the number of neighbouring solutions with identical cost) of a instance. The result has natural generalisation to the 3-opt landscape (at higher computational complexity). We relate the variance of tour costs over the solution space to that of the gains over the 2-opt landscape of a problem, providing an $\mathcal{O}(n^2)$ method to compute the variance of gains over the landscape.

We apply our method to compute the low order moments of the distribution of tour costs to several empirical studies of the solution space. Among other results we: confirm the known relationship between the standard deviation of tour costs and the optimal tour cost; we show a correlation between the skewness and the optimal tour cost; we demonstrate that the moments can be used to estimate the complete probability distribution of tour costs.

Contents

Certificate of Originality	i
Acknowledgments	ii
Abstract	iv
1 Introduction	1
1.1 Definition of the TSP	2
1.1.1 Example TSPs	2
1.2 Motivation for this Research	4
1.3 Outline and Author's Contribution	4
2 The Relationships Between the TSP and other Problems	6
2.1 Computability of the TSP	6
2.2 Approximation of the TSP	9
2.3 Types and Variations of the TSP	10
2.3.1 Euclidean and Geometric TSPs	10
2.3.2 TSP with Triangular Inequality	12
2.4 More Graph Theoretic Definitions	13
2.5 Other Problems	14
2.5.1 Minimum Spanning Trees	14
2.5.2 Hamiltonian Cycle Problem	14
2.5.3 Asymmetric TSP	15
2.5.4 Summary	15

3	Survey of TSP Algorithms	16
3.1	Tour Construction Methods	16
3.1.1	Branch and Bound	16
3.1.2	The Doubling Algorithm	18
3.1.3	Christofides' Algorithm	19
3.1.4	Nearest Neighbour	21
3.1.5	Clarke and Wright	21
3.1.6	Insertion Heuristics	22
3.1.7	The Greedy Heuristic	23
3.1.8	Gutin and Yeo Algorithm	24
3.1.9	Linear and Integer Programming	24
3.1.10	Neural Net Methods	25
3.2	Tour Improvement and Landscapes	26
3.2.1	Neighbourhoods and Landscapes	26
3.2.2	Landscapes and the TSP	27
3.2.3	Basic Tour Perturbation Operations	28
3.2.4	The k -opt Move	30
3.2.5	Distance in Landscapes	32
3.3	Tour Improvement Algorithms	32
3.3.1	Iterative Improvement	32
3.3.2	The Metropolis Algorithm	33
3.3.3	Simulated Annealing	33
3.3.4	The Lin-Kernighan Algorithm	35
3.3.5	Tabu Search	42
3.3.6	Evolutionary Algorithms	43
3.3.7	Ant Colony Optimisation	44
3.4	Summary and Relevance to this Thesis	47
4	Statistics and the TSP	49
4.1	Statistical Preliminaries	49
4.1.1	Measure Spaces	49
4.1.2	Probability Spaces	50
4.1.3	Distribution Functions and Random Variables	50

4.1.4	Independent Random Variables	52
4.1.5	Expected Value and Moments	52
4.1.6	Distributions and Processes	55
4.2	Probabilistic results on the TSP	60
4.3	The Stochastic TSP	60
4.3.1	Probability Distribution of the Euclidean TSP	62
4.3.2	Random Edge Cost Results	67
4.4	Sampling Results and the TSP	68
4.4.1	Distribution of Optimal Cost Tour	68
4.4.2	Relationship Between the Sample Variance of Tour Costs and the Optimal Tour Cost	69
4.5	Statistical Properties of Landscapes	70
4.5.1	Elementary Landscapes	70
4.5.2	Random Walks on Landscapes	72
4.6	The Empirical Result of Boese et al.	73
4.7	Conclusions	75
5	A Perfect TSP	76
5.1	Definition of a Perfect TSP	76
5.1.1	An Algorithm to Construct a Perfect TSP	77
5.1.2	Computability of the Perfect TSP	79
5.2	Tour Operations	80
5.3	Examples of the Distribution of Costs	81
5.4	Conclusion	84
6	The Expected Value of Tour Costs over Subsets of the Solution Space	86
6.0.1	The Stochastic TSP as a Generalisation of the Fixed Edge Cost TSP	87
6.0.2	The Number of Tours Containing Various Edges	87
6.1	The Expected Value of Tour Costs over the Solution Space	88
6.2	The Expected Value of Tour Costs over Subsets of the Solution Space	89

6.3	Conclusion and Further Research	92
7	The Variance of Tour Costs	94
7.1	Fixed Edge Costs	94
7.1.1	The Variance Theorem with Fixed Edge Costs	94
7.2	Random Edge Costs	99
7.3	Depth and Problem Size	102
7.3.1	Real World Problems	103
7.4	Conclusions and Future Research	105
8	The Third Moment of Costs over the Solution Space	107
8.1	Third Moment of Tour Costs	107
8.1.1	The Third Moment Theorem	108
8.1.2	Reducing the Computational Complexity	112
8.2	Empirical Results	115
8.2.1	Skewness Versus Problem Size	115
8.2.2	Skewness Versus Optimal Tour Cost	117
8.3	Conclusions and Future Research	119
9	The Fourth Moment of Costs over the Solution Space	121
9.1	The Fourth Moment of Tour Costs	121
9.2	The Fourth Moment Theorem	122
9.3	Conclusions and Future Research	127
10	Statistics over the 2-opt Landscape	129
10.0.1	The 2-opt Landscape of the TSP	129
10.1	The Probability Distribution of Gains	131
10.1.1	Equivalent Edges	131
10.2	The Variance of Gains	132
10.3	Computing the Distribution of Gains	136
10.4	Conclusions and Future Research	136
11	Future Research and Applications	138
11.1	The Moment Problem	139

11.1.1	An Algorithm to Solve the Moment Problem	140
11.1.2	Estimated Distributions	141
11.1.3	Moment Problem using Linear Programming	144
11.2	Fast Estimates of the Minimum Tour Cost	145
11.2.1	Correlation between the Optimal Tour Cost and the Mean and Standard Deviation	146
11.2.2	An Example of Optimal Tour Cost Estimation using the Mean and Standard Deviation	154
11.2.3	Conclusions	156
11.3	The Variance of Gains and Iterative Improvement in the 2-opt Landscape	158
11.3.1	Iterative Improvement and the Variance of Gains	159
11.3.2	Conclusions	162
11.4	The Covariance between Solution Spaces	162
11.5	Optimisation Algorithms	163
11.6	Generalisation to other Problems	164
11.7	Conclusion	165
	Appendix TSPLIB Numeric results	167
	Glossary	173
	Bibliography	177
	List of Figures	189
	List of Algorithms	194
	Index	195
	Colophon	200