MAXIMUM-ENTROPY PRINCIPLE APPROACH TO THE MULTIPLE TRAVELLING SALESMAN PROBLEM AND RELATED PROBLEMS

BY

BRIAN A. ROEHL

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Mechanical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Adviser:

Professor Srinivasa Salapaka

ABSTRACT

This thesis presents an investigation into the applications of the maximum-entropy principle as a heuristic for the multiple travelling salesman problem. This is a computationally complex problem which requires special treatment by conventional optimization techniques. Specific focus is given to developing a generalized framework for this problem that can be applied to any number of variants on the basic formulation. Additional consideration is given to the applications of this generalized framework to other variants on the travelling salesman problems such as the close enough travelling salesman problem.

The heuristic framework developed here is shown to provide flexibility in addressing the multiple salesman variation on the travelling salesman problem as well as a several other variants on the travelling salesman problem. Additionally, this framework is shown to be effective in determining solutions to this class of problems, and it is especially effective for the close-enough travelling salesman problems which is particularly challenging for most conventional combinatorial algorithms.

Concrete steps are presented by which to further extend and improve this framework to become both more widely applicable to variants on the travelling salesman problem, and more computationally efficient in solving such problems.

ACKNOWLEDGEMENTS

I'd like to thank my adviser Professor Srinivasa Salapaka for all of his help and support. He has been absolutely fantastic in guiding me through the difficulties and challenges that I have faced in this process. Without his insight I doubt I'd have nearly so much to show for my work. I have to thank my parents for years of guidance and support, and most importantly, for giving me a great deal of independence to go off and do things my own way sometimes. I'd also like to mention my younger sister Michelle who is turning out to be a fine engineer in her own right and who I'm quite proud of. I look forward to seeing everything she accomplishes in the future.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

THE TRAVELLING SALESMAN PROBLEM

The Travelling Salesman Problem (TSP) is one of the most extensively studied optimization problems. The basic formulation is that a salesman must visit a series of cities in such a way to minimize the total distance covered. Therefore a TSP is defined by a set of nodes or cities, and the edges between them which define the cost of travel between each city. Each solution to a TSP is called a tour, made up of a combination of edges such that each node is visited sequentially. The optimal tour is the combination of edges that minimizes the total cost to the salesman.

The applications of the TSP to real world problems are quite comprehensive. Common applications of the TSP are vehicle delivery route planning and toolhead path planning for drilling VSLI circuit boards. Junger et al. as well as Bektas have explored many more applications of the TSP to more specialized problems [1][2], demonstrating the real world value and importance of developing effective solutions to the TSP.

The travelling salesman problem is among the class of NP-Hard problems that are computationally intensive to solve [2]. Generally, finding the optimal solution requires a calculation time at least proportional to $\Theta(n!)$ where n is the number of nodes in the system. For relatively small data sets linear solving programs can find the optimal solution within a short period of time, but for larger data sets the computations can become extremely time intensive. Only with the development of more powerful computers has the optimal solution been discovered for larger data sets, as computation times in the hundreds of CPU-years had made those solutions infeasible in the past [3]. There are two main online sources for large test data sets along with the respective lower bound and current best known to [4][5]. Many heuristics have been developed for the TSP [1][2][6] and they can offer significant runtime improvements conventional optimization for large data sets, at the sacrifice of some deviation from the optimal solution.

VARIANTS ON THE TRAVELLING SALESMAN PROBLEM

Despite the success of heuristics for the basic TSP, there are a quite a few variants to the TSP that can pose a challenge to some of these methodologies. Variants to the basic TSP are necessary to appropriately model realistic situations. The multi-Travelling Salesmen Problem (mTSP) allows more than one 'salesman' to operate between the cities, such that the solution to the mTSP is comprised of several routes, one for each salesman, and the optimal tour would be the set of routes such that the total distance travelled is minimized. Additional constraints may be imposed on the system such as requiring each salesman to start at the same point, representing a warehouse or depot. In order to use the conventional heuristics on a mTSP, transformations can be applied in order to generate a TSP which represents the mTSP [7].

The Close Enough Travelling Salesman Problem (CETSP) is a variant where the salesman must only come within a certain radius of each city on the tour. This adds great complexity to the problem as this creates an infinite number of routes between each node. Because of the significant increase in the number of edges, many conventional heuristics are unable to address this variant. Special formulations have been developed to address this problem [8][9].

Another variant is whether the connections between each node are symmetric or asymmetric. A symmetric TSP means that the cost function between a pair of nodes will be the same regardless of the direction of travel between them, whereas an asymmetric TSP imposes varied cost functions between nodes that are based on the direction of travel. This is useful for modeling real world situations such as one-way roads for deliveries, or the jet stream for air travel [10]. Other variants on the TSP that will not be considered here are the multiple depots multiple travelling salesmen problem, capacity constraints on each salesman, or time windows for visiting each city [11]. These four variations are not considered in the framework presented here, but they present opportunities for further extension of this generalized framework at a later time.

Currently the wide field of variants on the basic TSP necessitates the existence of specialized heuristics to solve each variant or class of variant. This paper presents a more generalized framework for a heuristic for solving variants on the TSP.

MAXIMUM ENTROPY PRINCIPLE

The maximum-entropy principle was developed at the intersection of statistical mechanics and information theory. It is a methodology that allows the greatest amount of information about a system to be inferred from on a limited amount of given information [12]. This principle is capable of considering all system states by assigning each a probability within a probability distribution for the system. In this paper the maximum-entropy principle is used to consider every potential tour of the cities, and through the optimization process of deterministic annealing the shortest tour through every city is determined.

This deterministic annealing procedure is explained by Rose in the context of the clustering problem where m codevectors are used to determine clusters within n data points [13]. Rose then applies this deterministic annealing framework to the basic TSP as a case of constrained clustering, which serves as the foundation for this extension to the multiple travelling salesman problem.

NOVEL WORK

This thesis presents a new heuristic framework for approximating the solutions to the multiple travelling salesmen problem and other variants on the TSP. This approach is an extension of the maximum-entropy principle and deterministic annealing to a new form of the TSP. This framework is presented as a general tool that can be adapted to a number of variants on the basic TSP. This is advantageous as it provides greater flexibility than may be available with some of the current exact solutions or heuristics for the TSP. Additionally, because of the way this framework is formulated, unlike most approaches to the TSP, this framework is independent of the edges defined between pairs of nodes. This makes it particularly suited for variants such as the close-enough travelling salesman problem which are challenging due to the infinite number of possible tours available which is a serious problem for standard combinatorial algorithms. Through this thesis it will be shown that this new framework is an effective new tool for use in the travelling salesman problem and many variants thereof.

THESIS STRUCTURE

The problems that this research has addressed are laid out in Ch. 2: Problem Formulation. These three variations on the basic travelling salesman problem will be discussed throughout this paper. A formal review of the current state of the solutions to travelling salesman problem, both exact and heuristic, is presented in Ch. 3: Literature Review. Additionally the foundation of the heuristic framework presented in this paper is explained in the literature review; this consists of both the maximum-entropy principle and deterministic annealing. Ch. 4: Methodology is a thorough mathematical derivation of the heuristic framework for solving the each of the variants of the TSP presented in the Problem Formulation. The results of the implementation of the heuristic framework are presented in Ch. 5: Results and Analysis. This leads directly into Ch. 6: Discussion where the outcomes and implications of this research are considered and the next steps in the development of this framework are identified and explained, laying out the opportunities

CHAPTER 2

PROBLEM FORMULATION

MULTI-TRAVELLING SALESMEN PROBLEM: NON-RETURNING

Given a set of N nodes (x_i), and a set of M salesmen (y_j), determine the optimal tour such that each node is visited by one salesman and the total distance travelled by all salesmen is minimized. The starting and ending position of each salesman is not constrained and they do not have to be coincident. This is applicable to problems for non-recurring events, such as the scheduling of orders at a steel rolling company [14].

MULTI-TRAVELLING SALESMEN PROBLEM: RETURNING

Given a set of N nodes (x_i), and a set of M salesmen (y_j), determine the optimal tour such that each node is visited by one salesman and the total distance travelled by all salesmen is minimized. The starting and ending position of each salesman must be coincident. This is applicable to problems for recurring events, such as the scheduling of jobs at a commercial printing company with monthly orders [15].

Alternatively, salesmen may be constrained to start and end at a depot, which models the real world situation of a recurring routing problem such as the School Bus Routing Problem [1] or delivery problems. A further variant allows the salesmen to start at any one of D depots. This additional generalization will not be addressed in this paper.

CLOSE ENOUGH TRAVELLING SALESMAN PROBLEM

Given a set of N nodes (x_i), each with a specified radius ρ_i, and a set of M salesmen (y_j), determine the optimal tour such that at least one salesman comes within ρ_i of each node. A CETSP may be used to represent problems such as the wireless meter reader [8], or aerial reconnaissance [9]. The CETSP variant may be applied to any of the TSP class of problems. The most significant difference between point-based TSPs and the close-enough travelling salesman

problem is that due to the radius associated with each node, the CETSP does not define a specific edge between a pair of nodes, rather there is a continuous field of possible edges between a pair of nodes. The most significant result of this change is that it creates an infinite number of possible solutions to this problem.

CHAPTER 3

LITERATURE REVIEW

MAXIMUM ENTROPY PRINCIPLE

The maximum-entropy principle (MEP) has foundations in statistical mechanics and information theory, and shares significant similarities with deterministic annealing. The basic formulation of the MEP is outlined by Jaynes [12][16].

Given a function of x_i, $f(x_i)$, where x_i may take any of n discrete values, or states. There is a certain probability p_i of x_i taking each of n values, but that information is not available. Therefore the expected value of the function $f(x)$ is

$$F = \langle f(x) \rangle = \sum_{i=1}^{n} p(x_i)f(x_i) \tag{1}$$

Clearly the system is subject to a constraint on the sum of the probabilities p_i such that

$$\sum_{i=1}^{n} p(x_i) = 1 \tag{2}$$

The Shannon entropy of the probability distribution p_i is given by

$$H\big(p(x_1) \dots p(x_n)\big) = -\sum_{i=1}^{n} p(x_i) \log\big(p(x_i)\big) \tag{3}$$

In order to maximize the Shannon entropy Eq. (3) given the constraints Eq. (1) and Eq. (2) Lagrangian multipliers are employed as follows:

$$L = H - (\lambda_0 - 1)\left(\sum_{i=1}^{n} p(x_i) - 1\right) - \beta\left(\sum_{i=1}^{n} p(x_i)f(x_i) - F\right) \tag{4}$$

$$\frac{\partial L}{\partial p_i} = 0 = -(1 + \log(p(x_i))) - (\lambda_0 - 1) - \beta f(x_i)$$

$$\log(p(x_i)) = -\lambda_0 - \beta f(x_i)$$

$$p(x_i) = e^{-\lambda_0 - \beta f(x_i)} \tag{5}$$

Applying the constraint Eq. (2) to Eq. (5) yields

$$\sum_{i=1}^{n}\left(e^{-\lambda_0-\beta f(x_i)}\right) = 1$$

$$e^{-\lambda_0} = \frac{1}{\sum_{i=1}^{n} e^{-\beta f(x_i)}} \tag{6}$$

Substituting Eq. (6) back into Eq. (5) reveals that in order to maximize entropy, the probability of any state x_i is described by the Gibbs distribution

$$p(x_i) = \frac{e^{-\beta f(x_i)}}{\sum_{i=1}^{n} e^{-\beta f(x_i)}} \tag{7}$$

Finally, the Lagrange multiplier β may be determined by substituting Eq. (7) into Eq. (1) such that the value of β will achieve the appropriate expected value of the function $f(x)$.

DETERMINISTIC ANNEALING

Using the foundation of the maximum-entropy principle to develop the probabilistic representation of each potential system state, Rose deterministic annealing (DA) uses the Lagrange multiplier for the constrained Shannon entropy as a parameter in an optimization [17]. Varying β is equivalent to varying the expected value of the function $f(x)$. A very low β will lead to a uniform distribution, as evidenced by Eq. (7), and at the other extreme, for a very large β the probabilities will approach binary values of 1.0 or 0.0. Rose proposes an iterative solving process that slowly modifies β so that the location of the codevectors will gradually approach the optimum.

The method that allows the global optimum to be determined is the gradual modification of the Lagrange multiplier β. The starting condition of $\beta = 0$ causes the position of all codevectors to converge to the same point due to the uniform probability distribution from Eq. (7). As β is modified this global minimum changes, and by incrementing β in sufficiently small steps the solution for the codevectors will track this global minimum and will not solve to local minima. As β becomes sufficiently large the optimal solution will emerge.

The existing challenges in this procedure are two-fold. First, the rate of change of β must be sufficiently small so that the solution will still track the global minimum. This has the tradeoff

that smaller changes lead to more iterations, and therefore longer computation time. Additionally, there is no clear indication when β is sufficiently large such that the optimum solution has been reached.

Rose has explored how to apply DA to the TSP, which is similar to the case of a constrained clustering problem [13]. In order to use the DA algorithm to solve the TSP, the number of codevectors used in is set equal to the number of nodes in the TSP. The distortion function is modified to include the tour length. This requires then a second Lagrange multiplier for the tour length component of the distortion function, in addition to the first Lagrange multiplier for the original component of the distortion function between codevectors and nodes. Solving through the gradual change in both Lagrange multipliers leads to a solution of the TSP. Rose does not provide an optimal method for varying the two Lagrange multipliers in conjunction, though he suggests that the second multiplier θ be minimized before each increment of β is applied. There are opportunities to develop a more effective methodology for the changes in the values of β and θ during the optimization.

TRAVELLING SALESMAN PROBLEM EXACT SOLUTIONS

There are a number of methodologies to determine the optimal solution to a TSP, many of which are outlined in a survey of the state of TSP solutions by Junger [2]. Of specific interest is the Concorde computer code developed by Applegate et al. which is a hybrid implementation of the Lin-Kernighan heuristic to the TSP[18]. This code has been extensively developed in order to achieve minimal run time while still determining optimal results for TSPs. The Concorde solution is arguably the most successful optimal TSP solver; the largest TSP currently solved by the Concorde code is 85,900 cities, which required 136 CPU years [3]. Although the exact solution was determined, this clearly demonstrates the practical limitations on determining exact solutions to large data sets and motivates the need for heuristics to address the TSP class of problems.

TRAVELLING SALESMAN PROBLEM HEURISTICS

Johnson et al. performed a fairly comprehensive survey of the existing heuristic approaches to the travelling salesman problem [6]. Many of these heuristics are based in graph theory, such as the nearest neighbor heuristic. This is a construction heuristic which successively builds a solution tour. For this heuristic the nearest neighbor to the starting node is determined and a tour is constructed by continuing to add the nearest neighbor to the newest node until all have been reached. Johnson et al. showed that this heuristic generally produced solutions that were 24.2% beyond optimal. Other heuristics use insertion techniques where nodes are added to an incomplete tour which does not encompass all nodes. For example, the farthest insertion would successively insert, at the best position within the tour, the node which is farthest from all nodes that exists within the incomplete tour. Johnson's research shows that this method produces results that are on average 9.9% above the optimal tour.

A separate class of heuristics explained by Johnson is the improvement heuristics such as two-opt exchange. This heuristic identifies sets of nodes where a pre-existing tour crosses itself, and swaps the position of these nodes in order to shorten the tour length. This is more time intensive than some of the previously discussed heuristics, but it was able to achieve results that were 8.3% above the optimal.

More recently there have been other heuristics that have been developed that were not included in the survey conducted by Johnson. The ant colony optimization applied by Junjie et al. uses the concept of random ant movement affected by pheromones left by other ants. Each ant leaves pheromones where it has visited and over time the optimal solution develops based on the success of each tour completed by successive ants [19].

Extended simulated annealing has been another successful heuristic in the mTSP [20]. This heuristic is somewhat similar to deterministic annealing, although there are some critical differences. Simulated annealing computes the entropy of the system and the probability of a state having a specific energy level. The system is randomly perturbed by either adding or removing a codevector, swapping two codevectors within a salesman's route, or swapping two codevectors between the routes of two different salesmen. The change in the tour length is determined, from which the probability of that state occurring is generated. The process is

repeated until equilibrium is reached, at which point the Lagrange multiplier is incremented according to the pre-determined annealing schedule. This is repeated until the final value of the Lagrange multipliers is reached. Song et. Al demonstrated that this method is capable of determining mTSP solutions, though they did not reach conclusions as to the relative effectiveness of this heuristic compared to the other existing heuristics.

The close-enough travelling salesman problem is a more recently proposed formulation, and as result, less research has been conducted in this area. Gulczynski et al. proposed a heuristic solution that first consolidates the nodes to a set of supernodes, such that if each supernode is visited, the tour will have come within the appropriate distance of each node. The supernode set is solved, after which the tour is economized with respect to each node in order to minimize the total cost of the tour [8]. An alternate, though similar, solution by Mennell first creates Steiner Zones in which the tour comes close enough to multiple nodes. A representative point on each Steiner Zone is chosen and the tour is solved. Then, fixing the sequence of visiting each Steiner Zone the optimal point on each zone is determined [9]. This heuristic was demonstrated to be a very time-effective process for solving the CETSP problem, as it avoids repetitive computation of the distance matrix between nodes. Mennell has also proposed a mixed-integer nonlinear programming heuristic for the CETSP, and carried out an extensive series of benchmark tests to characterize the performance of this heuristic [21].

FOUNDATION FOR DERIVATION PROCESS

The goal is to develop a generalized heuristic for the returning mTSP. Simpler cases are considered first, after which the generalized solution is constructed. The simplest case of a mTSP contains two salesmen and does not require the salesmen to return to the starting node. A slightly more complex formulation requires the salesmen to return to the initial node, which may either be specified at a given depot or may remain unspecified. As the unspecified starting location is more general, that is considered here. Simplifying this to the single depot multi travelling salesman problem requires a minor adjustment which is discussed.

The generalized M salesman case is developed out of the returning mTSP with two salesmen. This takes advantage of the fact that several parts of the distortion function remain unchanged regardless of the number of salesmen.

Finally, the CETSP is considered. This is done in the context of the two salesman non-returning travelling salesman problem, but clearly demonstrates how the MEP framework presented here can be applied to many variations on the TSP.

MULTI-TRAVELLING SALESMAN PROBLEM: 2 SALESMEN, NON-RETURNING

This case is presented as a system with N nodes and 2 salesmen. The objective is to minimize the total distance travelled by both salesmen while visiting every node. The

An instance is defined by three parameters, Y, V, and R. The parameter Y is a set of codevectors, V is a set of association coefficients for each node-codevector pair, which are 0 unless the node and codevector are associated with each other, and R is the location of the partition representing the break between subsequent salesmen in the chain of consecutive codevectors. The instance can be described mathematically as

$$Y = \begin{Bmatrix} y_1 \\ \vdots \\ y_n \end{Bmatrix}; V = \{v_{ij}\} = \begin{cases} 1 \; if \; y_j \; is \; associated \; with \; x_i \\ 0 \; else \end{cases}; R = k \; \begin{matrix} there \; is \; no \; link \\ between \; y_k, y_{k-1} \end{matrix}$$

As this is the non-returning version of the mTSP there is no connection between y_1 and y_n. Therefore $y_0 = y_{n+1} = 0$ when they do appear in the equations.

Applying the maximum-entropy principle to this problem will be done in several steps. First the distortion function (or cost function) D is developed for each instance. There are three main components to the distortion function

$$D(Y,V,R) = D_1(Y,V) + D_2(Y) + D_3(Y,R) \tag{8}$$

The primary component is the distance between the nodes and codevectors, Eq. (9), and is the sole component in the basic clustering problem. This is dependent only on the codevectors and the association coefficients so that

$$D_1(Y,V) = \sum_{Y'} \sum_{V'} v_{ij} d(x_i, y_j) \tag{9}$$

In order to include the tour length in the cost function to represent a TSP, the distance between adjacent codevectors calculated in Eq. (10) which is solely dependent on the location of the codevectors, and it is calculated as

$$D_2(Y) = \theta \sum_{Y'} d(y_j, y_{j+1}) \tag{10}$$

The final component represents the partition of the codevectors for the independent salesmen. Eq. (11) accounts for this as it subtracts the distance at the partition between codevector y_k and y_{k+1} from the distortion function. This component is dependent on both the location of the codevectors and the partition parameter, which leads to the form of

$$D_3(Y,R) = -\theta d(y_k, y_{k+1}) \tag{11}$$

The distance function in Eq. (12) governs the how the minimization treats the distance between the nodes and codevectors. In this case the squared norm of the distance between two points is minimized as follows

$$d(x_i, y_j) = \|x_i - y_j\|^2 \tag{12}$$

13

The probability of any given instance is given by the Gibbs distribution in (13) as previously demonstrated in the discussion of the maximum-entropy principle in Eq. (7)

$$P(Y,V,R) = \frac{e^{-\beta D(Y,V,R)}}{\sum_{Y',V'R'}[e^{-\beta D(Y,V,R)}]} \tag{13}$$

The probability of a specific set of codevectors is simply calculated by summing Eq. (13) over the parameters V and R so that

$$P(Y) = \sum_{V',R'} P(Y,V,R) = \frac{\sum_{V',R'}[e^{-\beta D(Y,V,R)}]}{\sum_{Y',V'R'}[e^{-\beta D(Y,V,R)}]} \tag{14}$$

$$Let\ Q(Y) = \sum_{V',R'}[e^{-\beta D(Y,V,R)}] = \sum_{V',R'}[e^{-\beta(D_1(Y,V)+D_2(Y)+D_3(V,R))}]$$

$$Q(Y) = \left\{\sum_{V'} e^{-\beta D_1(Y,V)}\right\}\{e^{-\beta D_2(Y)}\}\left\{\sum_{R'} e^{\beta D_3(Y,R)}\right\} \tag{15}$$

Due to the binary nature of the association parameter v_{ij}, the first term of Eq. (15) may be simplified to become

$$\left\{\sum_{Y'} e^{-\beta D_1(Y,V)}\right\} = \left\{\prod_{i=1}^{n}\sum_{j=1}^{n} e^{-\beta d(x_i,y_j)}\right\} \tag{16}$$

Substituting Eq. (16) into Eq. (15) yields

$$Q(Y) = \left\{\prod_{i=1}^{n}\sum_{j=1}^{n} e^{-\beta d(x_i,y_j)}\right\}\{e^{-\beta D_2(Y)}\}\left\{\sum_{k=1}^{n} e^{\beta \theta d(y_k,y_{k+1})}\right\} \tag{17}$$

The probability function for the set of codevectors Y can be represented by the Gibbs distribution

$$Let\ P(Y) = \frac{e^{-\beta F(Y)}}{\sum_{Y'} e^{-\beta F(Y)}} \tag{18}$$

with respect to the function $F(Y)$ which is the free energy of the system

Substituting Eq. (18) into Eq. (14) allows F(Y) to be determined as

$$F(Y) = -\frac{1}{\beta}\log(Q(Y)) \tag{19}$$

Substituting Eq. (17) into Eq. (19) allows the free energy to be represented in terms of the distortion functions

$$F(Y) = -\frac{1}{\beta} \sum_{i=1}^{n} \left\{ \log \left[\sum_{j=1}^{n} \left(e^{-\beta d(x_i, y_j)} \right) \right] \right\} + \theta \sum_{j=1}^{m} d(y_j, y_{j+1}) - \frac{1}{\beta} \log \left[\sum_{k=1}^{n} \left(e^{\beta \theta d(y_k, y_{k+1})} \right) \right] \quad (20)$$

Taking the derivative of Eq. (20) with respect to each codevector allows the determination of the set of codevectors that maximize entropy in the system.

$$\frac{\partial F(Y)}{\partial y_m} = 2\theta(2y_m - y_{m-1} - y_{m+1})$$
$$- \frac{2}{\beta} \sum_{i=1}^{n} \left[(y_m - x_i)(-\beta) \frac{e^{-\beta d(x_i, y_m)}}{\sum_{j=1}^{n} e^{-\beta d(x_i, y_m)}} \right] \quad (21)$$
$$- \frac{2}{\beta}(\beta\theta) \frac{\left[(y_m - y_{m+1}) e^{\beta \theta d(y_m, y_{m+1})} + (y_m - y_{m-1}) e^{\beta \theta d(y_{m-1}, y_m)} \right]}{\sum_{k=1}^{n} e^{-\beta d(y_k, y_{k+1})}} = 0$$

The association probability function $p(y_m/x_i)$ represents the likelihood that a particular codevector will be associated with any particular node. This function follows the Gibbs distribution, and is expressed as

$$p(y_m|x_i) = \frac{e^{-\beta d(x_i, y_m)}}{\sum_{j=1}^{n} e^{-\beta d(x_i, y_m)}} \quad (22)$$

The partition probability function $p_r(m)$ represents the likelihood that the partition between the two salesmen will occur between any two codevectors. This function follows the Gibbs distribution and represents the probability that the partition will occur between codevectors y_m and y_{m+1}. The partition probability function is written a function of the distance function as

$$p_r(m) = \frac{e^{\beta \theta d(y_m, y_{m+1})}}{\sum_{k=1}^{n} e^{\beta \theta d(y_k, y_{k+1})}} \quad (23)$$

Substituting Eq. (22) and Eq. (23) into Eq. (21) allows for the characteristic equations for each y to be determined. Solving for each y provides the solution to the system at this pair of β and θ values, so that for every codevector

$$y_m = \frac{\theta y_{m+1}(1 - p_r(m)) + \theta y_{m-1}(1 - p_r(m-1)) + \sum_{i=1}^{n} p(y_m|x_i) x_i}{\theta(2 - p_r(m-1) - p_r(m)) + \sum_{i=1}^{n} p(y_m|x_i)} \quad (24)$$

15

Comparing the characteristic equation for the non-returning two salesmen TSP in Eq. (24) to the characteristic equation for a basic TSP in Eq. (25) shows how the partition probability function plays a critical role in distinguishing the solution to the mTSP from the basic TSP. It should be evident that the mTSP presented in Eq. (24) is only slightly more complex than the basic TSP in Eq. (25). In fact, the mTSP is transformed to the basic TSP when $p_r(m) = 0$ for all m

$$y_m = \frac{\theta y_{m+1} + \theta y_{m-1} + \sum_{i=1}^{n} p(y_m|x_i)x_i}{2\theta + \sum_{i=1}^{n} p(y_m|x_i)} \tag{25}$$

MULTI-TRAVELLING SALESMAN PROBLEM: 2 SALESMEN, RETURNING

When considering the case where the TSP must solve for a returning tour, there is some added complexity in the mTSP variant. For the non-returning case the partition function identifies the longest distance between codevectors as the location of the partition. In this case, not only does the partition function consider the distance between the codevectors where the partition occurs, but it must also account for the distance incurred in completing the continuous tour by reconnecting to the other end of the loop. In order to take a step towards a more general solution both partitions will be described by the partition parameter R and as a result the starting and ending codevectors are connected.

$$R = k, \ell \quad \begin{matrix} there\ is\ no\ link \\ between\ y_k, y_{k+1} \end{matrix} ; \begin{matrix} there\ is\ no\ link \\ between\ y_\ell, y_{\ell+1} \end{matrix} ; \begin{matrix} there\ is\ a\ link \\ between\ y_k, y_{\ell-1} \end{matrix} ; \begin{matrix} there\ is\ a\ link \\ between\ y_\ell, y_{k-1} \end{matrix}$$

It bears noting that the chain of codevectors has a distinct start y_1, and end y_n, but for the purpose of this framework, y_1 and y_n are considered to be adjacent such that

$$y_0 = y_n \ ; \ y_{n+1} = y_1$$

One component of the cost function must be reformed to include the extra information held in the revised partition parameter. Not only does the partition parameter describe where links will removed, it now also describes where new links will be introduced

$$D_3(Y, R) = -\theta[d(y_k, y_{k+1}) + d(y_\ell, y_{\ell+1}) - d(y_k, y_{\ell+1}) - d(y_\ell, y_{k+1})] \tag{26}$$

16

Because the terms containing each of D_1, D_2, and D_3 are independent in the free energy function Eq. (20), the changes in D_3 between Eq. (11) and Eq. (26) have no effect on the other terms. This allows the affected component of the free energy to be isolated and considered alone as $F_3(Y)$.

$$F(Y) = -\frac{1}{\beta}\sum_i\left\{log\left(\sum_j\left[e^{-\beta d(x_i,y_j)}\right]\right)\right\} + \theta\sum_j d(y_j, y_{j+1}) - \frac{1}{\beta}log\sum_k\sum_\ell e^{-\beta D_3(Y,R)}$$

$$Let\ F_3(Y) = -\frac{1}{\beta}log\sum_k\sum_\ell e^{\beta\theta[d(y_k,y_{k+1})+d(y_\ell,y_{\ell+1})-d(y_k,y_{\ell+1})-d(y_\ell,y_{k+1})]} \tag{27}$$

For this formulation the partition probability function must be adjusted to account for the new terms introduced to represent the continuous tour. These changes modify the partition probability function so that it is now

$$p_r(k,\ell) = \frac{e^{\beta\theta[d(y_k,y_{k+1})+d(y_\ell,y_{\ell+1})-d(y_k,y_{\ell+1})-d(y_\ell,y_{k+1})]}}{\sum_k\sum_\ell e^{\beta\theta[d(y_k,y_{k+1})+d(y_\ell,y_{\ell+1})-d(y_k,y_{\ell+1})-d(y_\ell,y_{k+1})]}} \tag{28}$$

The derivative of Eq. (27) with respect to the codevector y_m allows the characteristic equation of each codevector to be determined.

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta\sum_k\sum_\ell[p_r(k,\ell)\{(y_m - y_{m+1})\delta_{k-m} + (y_m - y_{m-1})\delta_{k-(m-1)} + (y_m - y_{m+1})\delta_{\ell-m}$$
$$+ (y_m - y_{m-1})\delta_{\ell-(m-1)} - (y_m - y_{\ell+1})\delta_{k-m} - (y_m - y_k)\delta_{\ell-(m-1)} - (y_m - y_{k+1})\delta_{\ell-m}$$
$$- (y_m - y_\ell)\delta_{k-(m-1)}\}] = 0$$

$$Where\ \delta_{k-m} = \begin{cases} 1 & if\ k - m = 0 \\ 0 & else \end{cases}$$

Like terms are grouped and the y_m terms cancel each other to yield

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta\sum_k\sum_\ell[p_r(k,\ell)\{(y_{\ell+1} - y_{m+1})\delta_{k-m} + (y_\ell - y_{m-1})\delta_{k-(m-1)} + (y_{k+1} - y_{m+1})\delta_{\ell-m}$$
$$+ (y_k - y_{m-1})\delta_{\ell-(m-1)}\}] = 0$$

The nested sums are reduced to several non-nested sums by separating the terms that are invariant with either k or l due to the delta functions. The simplified equality is

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta \left\{ \sum_\ell p_r(m, \ell)\{ y_{\ell+1} - y_{m+1}\} + \sum_k p_r(k, m)\{y_{k+1} - y_{m+1}\} \right.$$

$$\left. + \sum_\ell p_r(m - 1, \ell)\{ y_\ell - y_{m-1}\} + \sum_k p_r(k, m) \{ y_{k+1} - y_{m-1}\} \right\}$$

The equation is once again simplified changing the summation variables so that they are the same, then grouping like terms, resulting in the equality

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta \left\{ \sum_\ell \{p_r(m, \ell) + p_r(\ell, m)\} \{ y_{\ell+1} - y_{m+1}\} \right.$$

$$\left. + \sum_\ell \{p_r(m - 1, \ell) + p_r(\ell, m - 1)\} \{ y_\ell - y_{m-1}\} \right\} = 0 \tag{29}$$

By Eq. (28) it can be shown that

$$p_r(m, \ell) = p_r(\ell, m) \tag{30}$$

Eq. (30) is substituted into Eq. (29) and the terms with y_m are removed from the sums so that the sums are independent of y_m, and the equality can be expressed as

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta \left\{ 2p_r(m, m - 1)(y_m - y_{m+1}) + \sum_{\ell \neq m-1} 2p_r(m, \ell) \{ y_{\ell+1} - y_{m+1}\} \right.$$

$$\left. + 2p_r(m, m - 1)(y_m - y_{m-1}) + \sum_{\ell \neq m} 2p_r(m - 1, \ell) \{ y_\ell - y_{m-1}\} \right\} = 0 \tag{31}$$

Now that the derivative of each of the new components of the free energy equation is known, it can be reconstructed. Eq. (31) is the derivative of Eq. (20) with the modified distortion function D_3. This new equation is

$$\frac{\partial F(Y)}{\partial y_m} = 2\theta(2y_m - y_{m-1} - y_{m+1})$$

$$- 2 \sum_{i=1}^n [p(y_m|x_i)(y_m - x_i)]$$

$$- 2\theta \left\{ 2p_r(m, m - 1)(y_m - y_{m+1}) + \sum_{\ell \neq m-1} 2p_r(m, \ell) \{ y_{\ell+1} - y_{m+1}\} \right.$$

$$\left. + 2p_r(m, m - 1)(y_m - y_{m-1}) + \sum_{\ell \neq m} 2p_r(m - 1, \ell) \{ y_\ell - y_{m-1}\} \right\} = 0 \tag{32}$$

The characteristic equation for the returning two salesman problem is thus determined to be

$$\theta(y_{m+1} + y_{m-1})\{1 - 2p_r(m, m-1)\}$$
$$+\sum_{i=1}^{n} p(y_m|x_i)x_i$$
$$+\theta \sum_{\ell \neq m-1}[2p_r(m, \ell)\{y_{\ell+1} - y_{m+1}\}]$$
$$y_m = \frac{+\theta \sum_{\ell \neq m}[2p_r(m-1, \ell)\{y_\ell - y_{m-1}\}]}{2\theta\{1 - 2p_r(m, m-1)\} + \sum_{i=1}^{n} p(y_m|x_i)} \tag{33}$$

This characteristic equation shares a similar structure to Eq. (24) for the non-continuous tour. This provides intuition into how the characteristic equation responds to adjustments in the problem formulation which will be useful in generalizing this framework to the *M*-salesmen case.

MULTI-TRAVELLING SALESMAN PROBLEM: M SALESMEN, RETURNING

Clearly there is a necessity to develop a more generalized case than the two salesmen TSP. In order to achieve this, the framework presented here allows for *M* salesmen to be used in the MEP optimization. The definition of an instance must be modified to account for the existence of up to *M* salesmen. The change is applied to the partition parameter *R* such that

$$R = \{k_1, k_2 \dots k_M\} \text{ such that } \begin{array}{l} for\ each\ k_i\ there\ is\ no\ link\ between\ y_{k_i}, y_{k_i+1} \\ and\ there\ is\ a\ link\ between\ y_{k_i}, y_{(k_{i-1})+1} \end{array}$$

One advantage in this formulation provides is that it automatically checks whether a smaller number of salesmen is optimal. When $k_q = k_{q+1}$ the result is that this mimics a partition parameter

$$R = \{k_1, k_2 \dots k_i \dots k_M | i \neq q\} \text{ such that } \begin{array}{l} for\ each\ k_i\ there\ is\ no\ link\ between\ y_{k_i}, y_{k_i+1} \\ and\ there\ is\ a\ link\ between\ y_{k_i}, y_{k_{(i-1)}+1} \end{array}$$

This partition parameter results in *M-1* salesmen. This can be extended to show that if all k_i are equivalent the result is the single salesman TSP.

The distortion function Eq. (8) is still valid for this generalized definition of an instance. Just as in the two salesmen returning TSP, the only component that changes is the distortion function based on the partition parameter

$$D_3(Y, R) = -\theta \sum_{R'} \left\{ d(y_{k_i}, y_{k_i+1}) - d\left(y_{k_i}, y_{k_{(i-1)}+1}\right) \right\} \tag{34}$$

19

Substituting Eq. (34) into Eq. (8) and carrying out the same operations described previously yields the new free energy equation

$$F(Y) = \theta \sum_{j=1}^{m} d(y_j, y_{j+1}) - \frac{1}{\beta} \sum_{i=1}^{n} \left\{ log \left[\sum_{j=1}^{n} \left(e^{-\beta d(x_i, y_j)} \right) \right] \right\}$$
$$- \frac{1}{\beta} log \sum_{R'} e^{\beta\theta \left[d\left(y_{k_i}, y_{k_i+1}\right) - d\left(y_{k_i}, y_{k_{(i-1)}+1}\right) \right]}$$

(35)

Using the same methodology that was applied when making adjustments for the returning two salesman case, only the components of *F(Y)* that changed due to the modification of the cost function will be recalculated.

$$F_3(Y) = -\frac{1}{\beta} log \sum_{R'} e^{\beta\theta \left[d\left(y_{k_i}, y_{k_i+1}\right) - d\left(y_{k_i}, y_{k_{(i-1)}+1}\right) \right]}$$

(36)

Additionally, the partition probability function must be adjusted to account for this generalized case. It is still derived from the Gibbs distribution, only a few terms have changed so that it becomes

$$p_r(R) = \frac{\sum_{i=1}^{M} e^{\beta\theta \left[d\left(y_{k_i}, y_{k_i+1}\right) - d\left(y_{k_i}, y_{k_{(i-1)}+1}\right) \right]}}{\sum_{R'} \sum_{i=1}^{M} e^{\beta\theta \left[d\left(y_{k_i}, y_{k_i+1}\right) - d\left(y_{k_i}, y_{k_{(i-1)}+1}\right) \right]}}$$

(37)

The derivative of Eq. (36) with respect to the codevector y_m allows the characteristic equation of the codevectors to be determined. That derivative is

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta \sum_{R'} [p_r(R) \sum_{i=1}^{M} \left\{ (y_m - y_{m+1})\delta_{k_i-m} + (y_m - y_{m-1})\delta_{k_i-(m-1)} - \left(y_m - y_{k_{(i-1)}+1} \right) \delta_{k_i-m} \right.$$
$$\left. - (y_m - y_{k_i})\delta_{(k_{i-1})-(m-1)} \right\}] = 0$$

Expanding over *i*, the terms are grouped and the y_m's cancel each other. The equation is regrouped to be represented as a sum over *i* once again so that

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta \sum_{R'} \sum_{i=1}^{M} \left[p_r(R) \left\{ \left(y_{k_{(i-1)}+1} - y_{m+1} \right) \delta_{k_i-m} + \left(y_{k_{(i+1)}} - y_{m-1} \right) \delta_{k_i-(m-1)} \right\} \right] = 0$$

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta \left\{ \sum_{R' \neq k_i} \sum_{i=1}^{M} p_r(R|k_i = m)\left(y_{k_{(i-1)}+1} - y_{m+1}\right) \right.$$

$$\left. + \sum_{R' \neq k_i} \sum_{i=1}^{M} p_r(R|k_i = m-1)\left(y_{k_{(i+1)}} - y_{m-1}\right) \right\} = 0$$

Let R_2 be defined as all k_j within the set R excluding k_i, so the equality becomes

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta \left\{ \sum_{R_2'} \sum_{i=1}^{M} p_r(R|k_i = m)\left(y_{k_{(i-1)}+1} - y_{m+1}\right) \right.$$

$$\left. + \sum_{R_2'} \sum_{i=1}^{M} p_r(R|k_i = m-1)\left(y_{k_{(i+1)}} - y_{m-1}\right) \right\} = 0 \tag{38}$$

Using Eq. (37) it can be shown that

$$p_r(k_1, k_2 \ldots k_i \ldots k_M) = p_r(k_2, k_3 \ldots k_i \ldots k_M, k_1) = \cdots = p_r(k_M, k_1, k_2 \ldots k_i \ldots k_{M-1}) \tag{39}$$

An appropriate change of variables is applied to Eq. (38), specifically, substituting k for $k_{(i-1)}$ in the first summation and substituting k for $k_{(i+1)}$ in the second summation. Let R_3 be defined as all k_j within the set R_2 excluding k. Substituting Eq. (39) into this result produces

$$\frac{\partial F_3(Y)}{\partial y_m} = -2\theta M \left\{ \sum_{k=1}^{n} \left[\sum_{R_3'} p_r(m, k, R_3) \right](y_{k+1} - y_{m+1}) \right.$$

$$\left. + \sum_{k=1}^{n} \left[\sum_{R_3'} p_r(R_3, k, m-1) \right](y_k - y_{m-1}) \right\} = 0 \tag{40}$$

Just as in the two salesman case, the instances where y_m appears are removed from the sums in Eq. (40) creating the equation

$$-2\theta M \left\{ \left[\sum_{R_3'} p_r(m, m-1, R_3) \right](y_m - y_{m+1}) + \left[\sum_{R_3'} p_r(R_3, m, m-1) \right](y_m - y_{m-1}) \right\}$$

$$-2\theta M \left\{ \sum_{k \neq m-1}^{n} \left[\sum_{R_3'} p_r(m, k, R_3) \right](y_{k+1} - y_{m+1}) \right.$$

$$\left. + \sum_{k \neq m}^{n} \left[\sum_{R_3'} p_r(R_3, k, m-1) \right](y_k - y_{m-1}) \right\} = 0 \tag{41}$$

The derivative of each component of the free energy is now known, therefore the derivative of Eq. (35) is calculated and by separating the y_m's, the generalized characteristic equations for the codevectors that maximizes free entropy are determined. The characteristic equation for each codevector is

$$y_m = \frac{\begin{array}{l} \theta M\left[\sum_{R_3'} p_r(m, m-1, R_3)\right](y_{m+1} + y_{m-1}) \\ +\sum_{i=1}^{n} p(y_m|x_i)x_i \\ +\theta M\sum_{k \neq m-1}^{n}\left[\sum_{R_3'} p_r(m, k, R_3)\right](y_{k+1} - y_{m+1}) \\ +\theta M\sum_{k \neq m}^{n}\left[\sum_{R_3'} p_r(R_3, k, m-1)\right](y_k - y_{m-1}) \end{array}}{\theta\left\{1 - 2M\left[\sum_{R_3'} p_r(m, m-1, R_3)\right]\right\} + \sum_{i=1}^{n} p(y_m|x_i)} \tag{42}$$

This characteristic equation will solve a returngin mTSP, and it will consider all cases for M of fewer salesmen. If it is desired to consider the single depot multiple salesmen problem the constraint that is applied is that at each partition parameter within R, $y_{ki} = y_{depot}$.

CLOSE ENOUGH TRAVELLING SALESMAN PROBLEM

In the close enough travelling salesman problem the additional radius parameter must be included in the minimization. The formulation here is explained for the non-returning two salesmen TSP for the sake of simplicity. This requires a modification to the distance function from Eq. (12)

$$d_{CE}(x_i, y_j, \rho_i) = (\|x_i - y_j\| - \rho_i)^2 \tag{43}$$

This modified distance function is only used when calculating the distance between nodes and codevectors. Eq. (9) becomes

$$D_1(Y, V) = \sum_{Y'}\sum_{V'} v_{ij} d_{CE}(x_i, y_j, \rho_i) \tag{44}$$

The association probability function needs to be modified slightly to account for the change in the distance function so that

$$p_{CE}(x_i, y_j, \rho_i) = \frac{e^{-\beta d_{CE}(x_i, y_j, \rho_i)}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \left(e^{-\beta d_{CE}(x_i, y_j, \rho_i)}\right)} \tag{45}$$

The free entropy function from Eq. (20) is modified with the adjusted distance function for D_1, so that the free entropy is now

$$F(Y) = -\frac{1}{\beta} \sum_{i=1}^{n} \left\{ log \left[\sum_{j=1}^{n} \left(e^{-\beta d_{CE}(x_i, y_j, \rho_i)}\right) \right] \right\} + \theta \sum_{j=1}^{m} d(y_j, y_{j+1}) - \frac{1}{\beta} log \left[\sum_{k=1}^{n} \left(e^{\beta \theta d(y_k, y_{k+1})}\right) \right] \tag{46}$$

$$Let \ F_1(Y) = -\frac{1}{\beta} \sum_{i=1}^{n} \left\{ log \left[\sum_{j=1}^{n} \left(e^{-\beta d_{CE}(x_i, y_j, \rho_i)}\right) \right] \right\} \tag{47}$$

The derivative of Eq. (47) with respect to the codevector y_m is determined as it is the only unknown component of the derivative of Eq. (46). This derivative is

$$\frac{\partial F_1(Y)}{\partial y_m} = \sum_{i=1}^{n} \sum_{j=1}^{n-1} \left(p_{CE}(x_i, y_j, \rho_i)(y_m - x_i + sgn(x_i - y_m)\rho_i) \right)$$

The derivative of each component of the free energy is now known, therefore the derivative of Eq. (46) is calculated, and by separating the y_m's the generalized characteristic equations for the codevectors that maximize free entropy are determined to be

$$y_m = \frac{\theta y_{m+1}(1 - p_r(m)) + \theta y_{m-1}(1 - p_r(m-1)) + \sum_{i}^{n} p(y_m|x_i)(x_i - sgn(x_i - y_m)\rho_i)}{\theta(2 - p_r(m-1) - p_r(m)) + \sum_{i=1}^{n} p(y_m|x_i)} \tag{48}$$

The inclusion of the new radius parameter for the CETSP has only a minor effect on the characteristic equation compared to Eq.(24), causing the appearance of one extra ρ_i term. This mimics a non-CETSP problem when all ρ_i are set to zero, therefore it is presented as a more general case of any of the previously discussed TSPs.

CONTROLLING LAGRANGE MULTIPLIERS

It is desirable to have a consistent and repeatable method for varying the Lagrange multipliers β and θ which govern the distortion function and the tour length. The improper choice of a scheme to vary β and θ has the potential to result in a final tour that poorly approximates the optimal

solution. For an extreme example, one scheme may choose to first maximize β while maintaining a large θ, followed by holding β constant as θ is decreased.

In this study the β multiplier was considered as the main driver, and the θ multiplier was secondary. As such, the θ parameter was decreased according to an exponential function until a stable tour length was reached, at which point the β multiplier was increased according to an exponential function. This process was repeated until a sufficiently high β value and sufficiently low θ value were both reached, leaving the final solution.

One important consideration is the covariance of θ based on changes in β, as it is desirable to maintain the same tour length L before and after incrementing β. This was addressed by Rose [13] and he presents the formulation of how to modify θ in conjunction with changes in β. This has been generalized to the mTSP case and was applied in the implementation of this algorithm given that at a given β value, θ^* is the optimal value of θ. Therefore

$$\theta' = \theta^* + \Delta\theta^*(\beta) \tag{49}$$

Borrowing the deterministic annealing framework for free energy, it can be shown that

$$\theta^* = -\frac{\partial F^*}{\partial L} \tag{50}$$

In order to maintain a consistent tour length the partial derivative of θ with respect to β is computed to determine how θ changes, and the value can be adjusted accordingly. This creates

$$\frac{\partial\theta^*}{\partial\beta} = -\frac{\partial}{\partial\beta}\frac{\partial F^*}{\partial L} = -\frac{\partial}{\partial L}\frac{\partial F^*}{\partial\beta} \tag{51}$$

It is observed that for this constrained optimization where $h(Y) = L$

$$\frac{\partial F}{\partial y_i} + \theta\frac{\partial h(Y)}{\partial y_i} = 0 \tag{52}$$

As $h(Y)$ is held constant, Eq. (52) yields

$$\frac{\partial F}{\partial y_i} = 0 \tag{53}$$

Expanding Eq. (51) and substituting Eq. (53) generates an equation that no longer contains the optimal free energy, which allows for a solution to be determined

$$\frac{\partial \theta^*}{\partial \beta} = -\frac{\partial}{\partial L}\frac{\partial F(Y^*,\beta)}{\partial \beta} + \sum_k \frac{\partial F(Y^*,\beta)}{\partial y_k}\frac{\partial y_k^*}{\partial \beta} = -\frac{\partial}{\partial L}\frac{\partial F(Y^*,\beta)}{\partial \beta} \tag{54}$$

Substituting the free energy from Eq. (35) and taking the derivative results in

$$\frac{\partial \theta^*}{\partial \beta} = -\frac{\partial}{\partial L}\left\{\frac{1}{\beta^2}\sum_{i=1}^{n}\left\{log\left[\sum_{j=1}^{n} e^{-\beta d(x_i,y_j)}\right]\right\} + \sum_{i=1}^{n}\left\{\sum_{j=1}^{n} d(x_i,y_j)p(y_j|x_i)\right\}\right.$$

$$+ \frac{1}{\beta^2}log\sum_{R'} e^{\beta\theta\left[d\left(y_{k_i},y_{k_i+1}\right)-d\left(y_{k_i},y_{k_{(i-1)}+1}\right)\right]} \tag{55}$$

$$\left. + \sum_{R'} p_r(R)\left[d(y_{k_i},y_{k_i+1}) - d\left(y_{k_i},y_{k_{(i-1)}+1}\right)\right]\right\}$$

$$\frac{\partial \theta^*}{\partial \beta} = -\frac{\partial}{\partial L}\left\{-\frac{F-D_2}{\beta} + \sum_{i=1}^{n}\left\{\sum_{j=1}^{n} d(x_i,y_j)p(y_j|x_i)\right\}\right.$$

$$\left. + \sum_{R'} p_r(R)\left[d(y_{k_i},y_{k_i+1}) - d\left(y_{k_i},y_{k_{(i-1)}+1}\right)\right]\right\} \tag{56}$$

$$\frac{\partial \theta^*}{\partial \beta} = -\frac{\partial}{\partial L}\left\{-\frac{F(Y^*)-D_2(Y^*)}{\beta} + \frac{E(Y^*)}{\beta} + \frac{E_R(Y^*)}{\beta}\right\} \tag{57}$$

Given that

$$E = \sum_{i=1}^{n}\left\{\sum_{j=1}^{n} d(x_i,y_j)p(y_j|x_i)\right\}; \quad E_R = \sum_{R'}\left[d(y_{k_i},y_{k_i+1}) - d\left(y_{k_i},y_{k_{(i-1)}+1}\right)\right]p_r(R) \tag{58}$$

Converting the free energy in Eq. (57) to the optimal free energy once again and substituting into Eq. (51) provides the equation for the new θ' value that should be used when β is adjusted. Note that Eq. (10) is the tour length, so the derivative with respect to tour length is one. This new equation is

$$\theta' = \theta^* - \frac{\partial \beta}{\beta}\left(\frac{E^*(Y) + E_R^*(Y)}{\partial L} + \theta^* - 1\right) \tag{59}$$

The adjustment for parameter θ described in Eq. (59) was employed in this implementation. This does not address the question of the most appropriate way to vary β and θ in conjunction with each other, but it does remove the problem posed by the interdependence of parameters β and θ.

CHAPTER 5

RESULTS AND ANALYSIS

BACKGROUND OF RESULTS

For this work the DA algorithm was implemented in MATLAB to solve first the basic symmetric TSP, and then the more complicated symmetric mTSP. Finally the CETSP was implemented. This section provides an overview of the results of the MATLAB implementations of this heuristic.

At this point the most appropriate evaluations are based on the accuracy of the heuristic approximation of the solution to a given TSP. Evaluations based on computational time are valuable, but will not provide valid data based on the MATLAB implementation of this heuristic. Minimal consideration has been given to runtime optimization for this code, as that is a separate body of work. As a result, the only computation time comparisons being carried out are comparing this code to itself to understand runtime scaling with problem size. This heuristic has been shown to achieve sound results based on tour length, but the additional component of computation time is highly dependent on the implementation of this algorithm in code. As yet, the MATLAB code used for this implementation has not been optimized for minimum computation time, so valid comparisons on the basis on run time are not currently available. For reference, computation time is recorded using the Profiler utility in MATLAB. All code was run on a 2.80 GHz Intel Core i7 Processor with 8 GB RAM, though at no point was more than 3.5GB of RAM in use. All distances are the geometric distance, as defined in Eq. (12).

Tests were performed on both the mTSP heuristic and the CETSP variation. The performance of the mTSP heuristic was evaluated using the two salesman case for the sake of simplicity. Several test cases were evaluated on each algorithm. First a series of random sets of nodes was solved in order to determine an approximate average computation time for comparison based on the number of nodes. The next test was a repeated application of the heuristic to the same random set of nodes in order to quantify the variability of the results. Finally each heuristic was run for a larger test case of 100 nodes from the online database TSPLIB [5].

MUTLI TRAVELLING SALESMAN PROBLEM

The solution shown in Figure 1 demonstrates the successful output of a single run of the mTSP algorithm for a ten node problem. Figure 2 is a solution to a 20 node mTSP and Figure 3 is a solution to a 40 node mTSP. These are each taken from the results of a series of ten trials on random data sets, used to understand the behavior of this heuristic in regard to scaling the number of nodes in the problem.
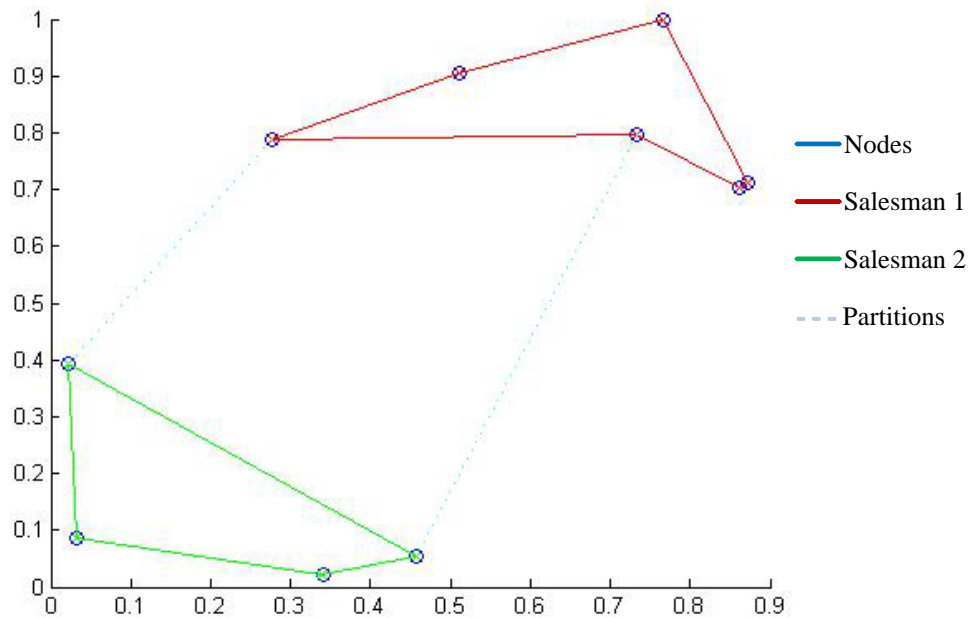

Figure 1: mTSP Solution to 10 node random data set, pseudo-random uniform distribution. Tour length = 0.468, Computation time = 62 seconds
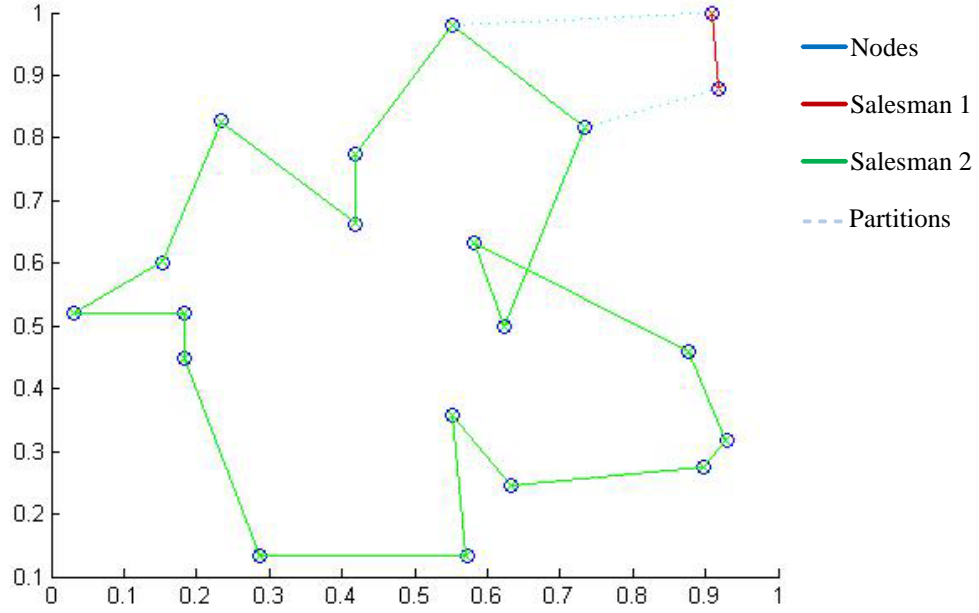
Figure 2: mTSP Solution to 20 node random data set, pseudo-random uniform distribution. Tour length = 1.20, Computation time = 186 seconds
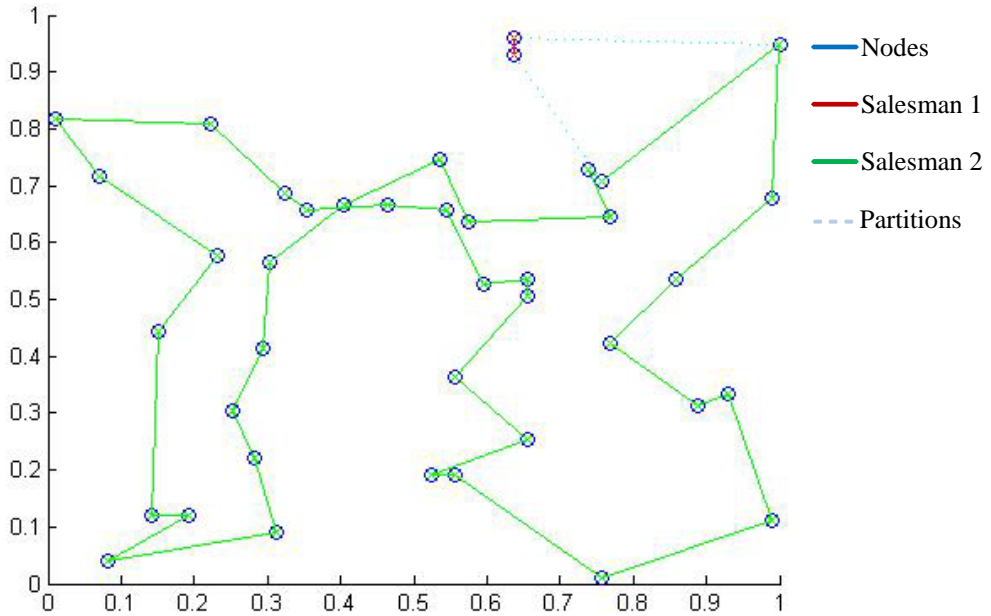

Figure 3: mTSP Solution to 40 node random data set, pseudo-random uniform distribution. Tour length = 0.961, Computation time = 719 seconds

The summaries of three sets of ten random problems for each of 10, 20, and 40 nodes are displayed in Table 1. The critical takeaway from this table is the rate of increase in the mean

computation time with respect to the number of nodes in the problem. The average rate of increase is slightly below the square of the rate of increase for the number of nodes, which is consistent with the hypothesis that this heuristic will operate in approximately $\Theta(n^2)$. Further study is required to confirm this relationship, but the initial results are favorable.

Table 1: mTSP Results – Randomized Data Sets (Uniform Distribution)

Nodes	Representative Figure	Trials	Mean Tour Length	Mean Computation Time
10 random nodes	Figure 1	10	0.665	61.4 seconds
20 random nodes	Figure 2	10	0.751	191.3 seconds
40 random nodes	Figure 3	10	0.858	715 seconds

Table 2: mTSP Results – Repeated Data Set (10 Trials, Uniform Distribution)

Nodes	Mean Tour Length	Standard Deviation	Coef. of Variation	Mean Time	Max Tour Length	Min Tour Length	Max / Min Tour
10 nodes	0.604	$5.41 * 10^{-5}$	$8.95 * 10^{-5}$	60.9 sec	0.6041	0.6039	1.0003
20 nodes	1.15	0.130	0.113	180.5 sec	1.29	0.824	1.57
40 nodes	0.964	0.0609	0.0631	687.8 sec	1.084	0.898	1.206

From this small sample of outputs several things become immediately obvious. First, the solution is demonstrating some degree of unpredictability, which is undesirable. This appears as the coefficient of variation and the ratio of the maximum to minimum tour length. This should not be a function of the MEP, but rather it is likely an artifact of the numeric processes used to solve the system. This will be explored in greater detail in the following section.
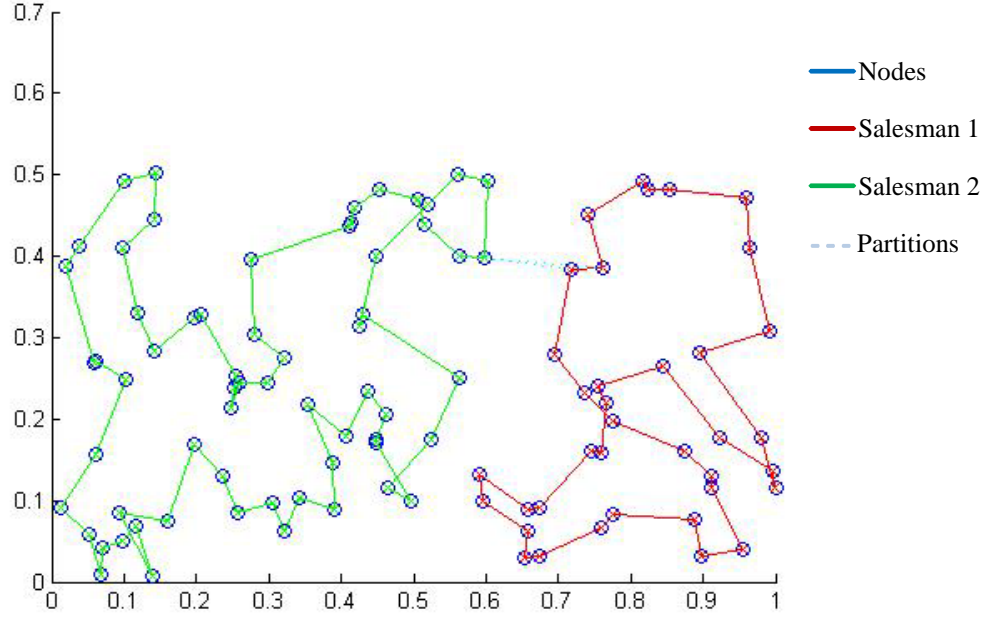
Figure 4: mTSP Solution to kroD100 from TSPLIB [5], Tour length = 185.1, Computation time = 4642 seconds

One application of this heuristic was made on a larger data set. Due to a non-optimal code implementation the runtime starts to become prohibitive with more than 100 data points, as this calculation took nearly 80 minutes. Clearly this is an area with significant opportunities for improvement. The optimization resulted in an acceptable tour, but by no means is it the optimal tour. The optimal TSP has been solved for this data set, but no such optimal solution is available for the mTSP variation, so it isn't possible to determine exactly how much error remains in the system. Despite that, visual observation allows identification of at least two sub-optimal areas, specifically at (0.5, 0.1) and (0.75, 0.2).

Given that the maximum-entropy principle and deterministic annealing were first applied to data clustering [13], it is reasonable to question whether an appropriate alternative heuristic would be to find clusters within the data and then determine the optimal tour among only the subset of nodes in that cluster with a single salesman. Undoubtedly this will work in some cases, especially those similar in composition to Figure 1 where two distinct subsets emerge. On the other hand, there are data sets that are particularly challenging for this proposed clustering approach, but which the MEP framework is capable of successfully solving. The data set to be used in this example is a pair of concentric rings of nodes. The optimal solution as determined by

31

the MEP heuristic is Figure 5. In this case the clustering approach will not be successful as the only cluster identified will be at the origin and when the two salesmen are allocated to the nodes, there is no way to effectively partition the set into two distinct subsets based on the information provided by the clustering solution. The resultant non-optimal solution is shown in Figure 5. On the other hand, the two salesmen from the MEP solution, achieve the optimal tour, assigning one salesman to each ring of nodes as demonstrated in Figure 6.
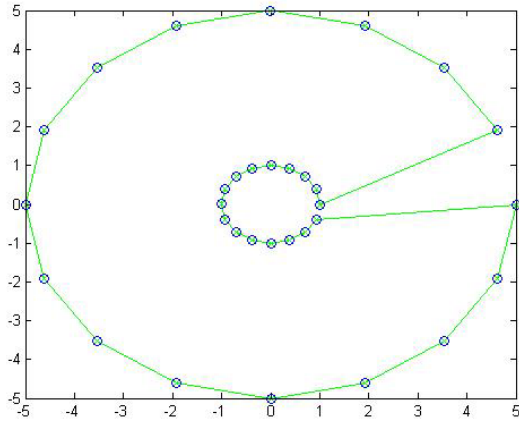

Figure 5: Cluster - TSP non-Optimal Solution to Concentric Rings of Nodes
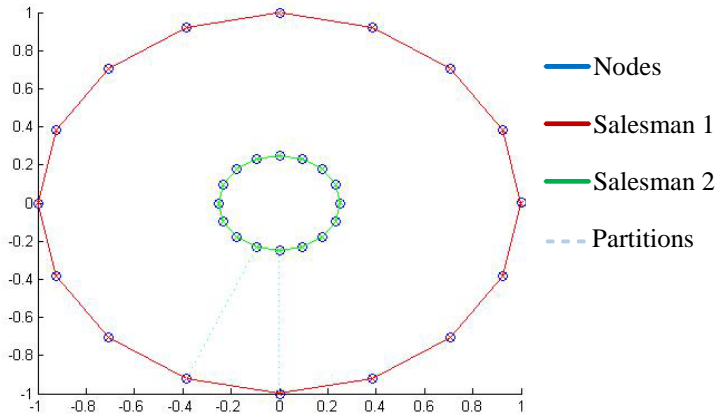
Figure 6: mTSP Optimal Solution to Concentric Rings of Nodes

The concentric ring data set inspires a second data set where clustering also fails to determine the optimal solution while MEP is successful. Figure 7 and Figure 8 demonstrate that clustering followed by TSP yields an inferior solution in cases beyond those where the nodes are uniformly distributed about the mean value such as in Figure 5.
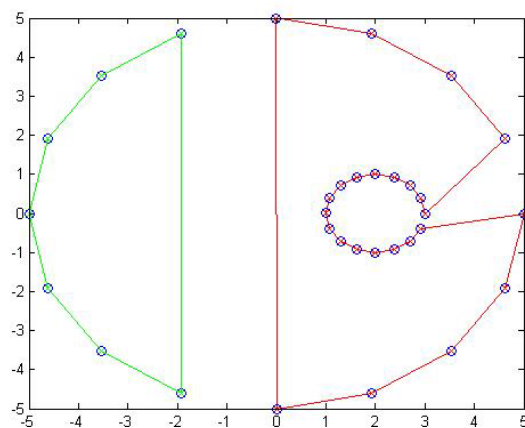

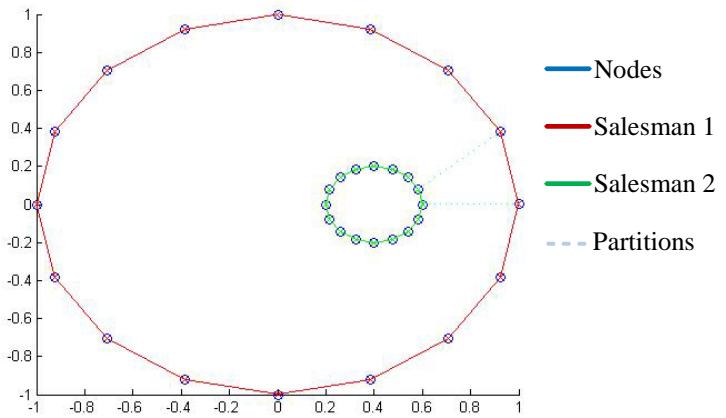Figure 7: Cluster - TSP non-Optimal Solution to Concentric Rings of Nodes

Figure 8: mTSP Optimal Solution to Offset Rings of Nodes

CLOSE ENOUGH TRAVELLING SALESMAN PROBLEM

As described in Eq. (48), the close enough travelling salesman problem requires relatively minor adjustments to the TSP algorithm. These were implemented in MATLAB and this variant on the heuristic was tested on a new data set that includes the necessary radius parameter for each node. It is difficult to determine whether the algorithm arrives at an optimal solution because this is much more difficult to check manually and unlike the standard TSP, there is no database of optimal tours for the CETSP. Data on the results of one CETSP optimization heuristic, mixed-integer nonlinear programming, is available from [21], but there is no optimal tour for comparison, only the best know results to date. Therefore, the tour distance generated by this heuristic could not be compared to a baseline at this time to determine the quality of the heuristic, defined as the normalized difference in tour length $(L_h - L_{opt}) / L_{opt}$ [2].

The solution shown in Figure 9 demonstrates the successful output of a single run of the CETSP algorithm. This solution was determined in 65 seconds. Figure 10 shows an example solution to a randomized 20 point data set, and Figure 11 is an example solution of a randomized 40 point data set. In each of these results it is clearly shown that the generated solution approaches within ρ_i of each node.

Figure 9: CETSP Result for non-returning 2 salesmen problem. 10 nodes; pseudo-random uniform distribution. Tour length = 0.492 units. Computation time = 65 seconds
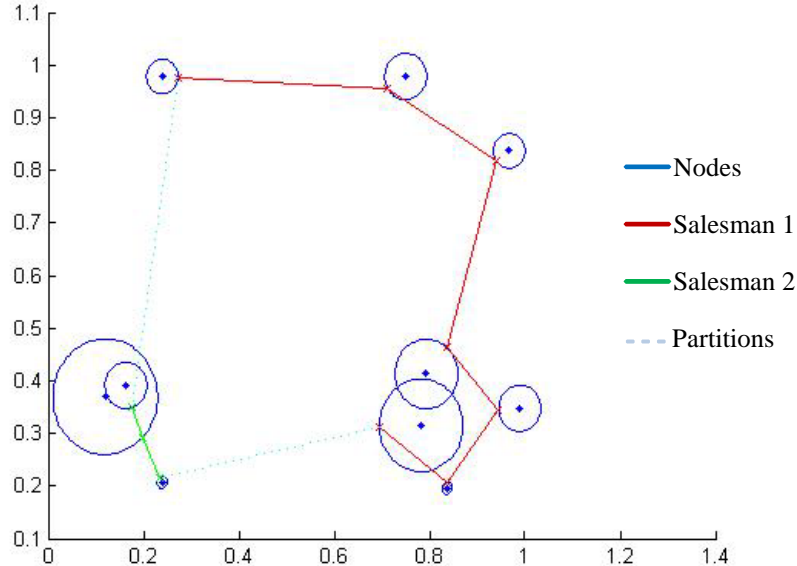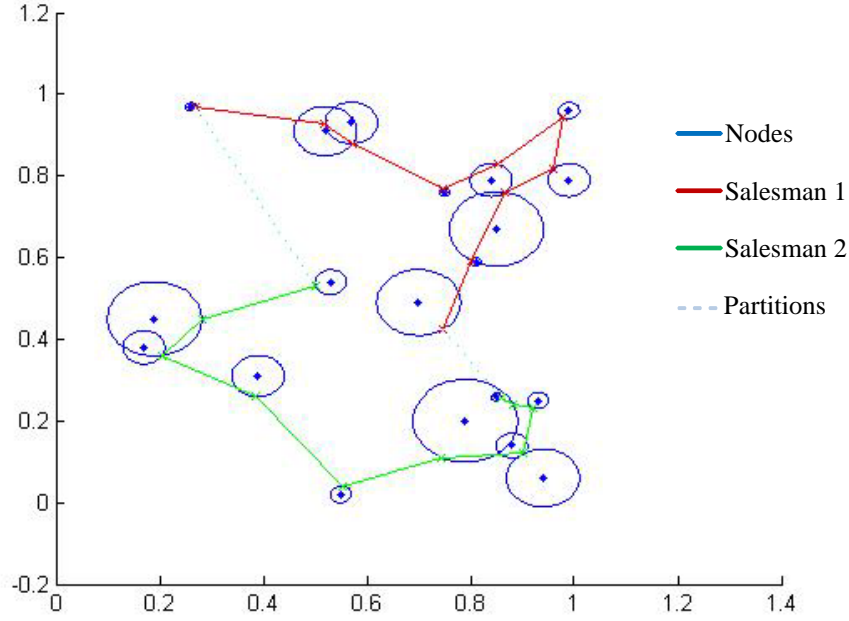
Figure 10: CETSP Result for non-returning 2 salesmen problem. 20 nodes; pseudo-random uniform distribution. Tour length = 0.515 units. Computation time = 185 seconds

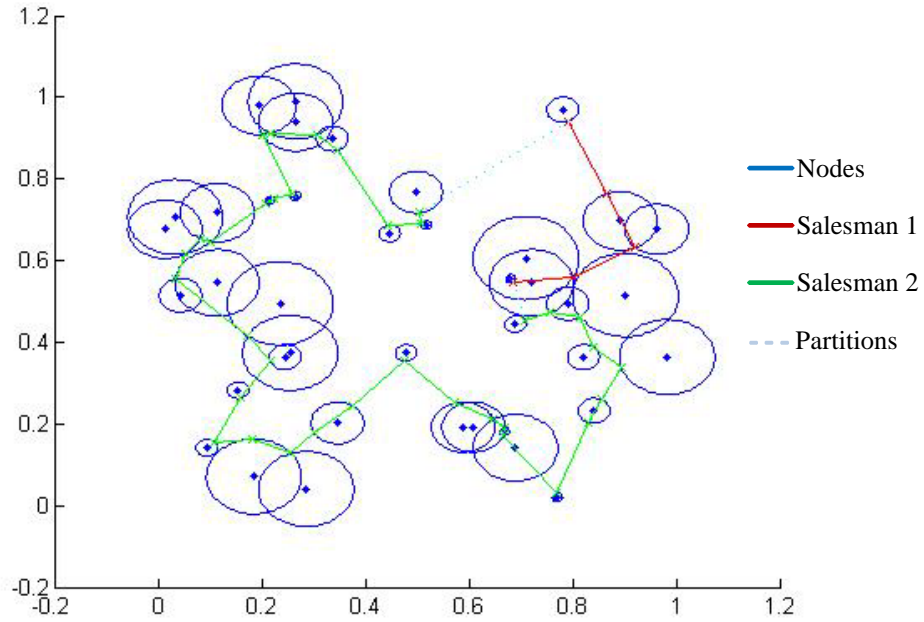Figure 11: CETSP Result for non-returning 2 salesmen problem. 40 nodes; pseudo-random uniform distribution. Tour length = 0.439. Computation time = 352 seconds

The results of ten consecutive runs on randomized node sets for 10, 20, and 40 nodes are summarized in Table 3. A new data set is generated for each run so that the comparison between mean computation time will not be as significantly impacted by the particular set of nodes

generated for a given trial. Although this heuristic is hypothesized to operate in approximately $\Theta(n^2)$ time, it is suggested that for the cases here for the CETSP the increasing overlap between nodes accounts for some of the reduction in the nominal computation time to required arrive at a solution. Although the initial results appear favorable, further research is necessary to characterize the scaling law for the CETSP heuristic.

Table 3: CETSP Trial Results – Randomized Data Sets (Uniform Distribution)

Nodes	Representative Figure	Trials	Mean Tour Length	Mean Computation Time
10 random nodes	Figure 9	10	0.382	67.1 sec
20 random nodes	Figure 10	10	0.375	165.2 sec
40 random nodes	Figure 11	10	0.438	364.2 sec

Repeated trials on the same set of nodes provide greater insight into the consistency of the results of the CETSP heuristic. The data in Table 4 provides the results of repeated solutions on the same data set for the 10 node, 20 node, and 40 node cases. The range between the maximum and minimum tour length is concerning as this shows there can be up to 35% variation between solutions to the same given problem. This is clearly undesirable as it leads to unpredictable outcomes for the heuristic approximation.

Table 4: CETSP Trial Results – Repeated Data Sets (10 Trials each, Uniform Distribution)

Nodes	Mean Tour Length	Standard Deviation	Coef. of Variation	Mean Time	Max Tour Length	Min Tour Length	Max / Min Tour
10 nodes	0.218	0.0302	0.139	71.9 sec	0.265	0.196	1.35
20 nodes	0.495	0.0296	0.0599	169.8 sec	0.533	0.455	1.17
40 nodes	0.398	0.0478	0.120	433.6 sec	0.477	0.344	1.38

This heuristic was compared against one of the 100 node sets tested by Mennell. The test data was the kroD100 data set from TSPLIB [5], with equal radii of 11.697. This was calculated by Mennell to achieve a 0.3 overlap ratio on the data. The output is shown in Figure 12, and the comparison to Mennell's results is shown in Table 5.

It is important to note that for this solution the first salesman visits only a single node near (1.0, 0.4) and the node near (1.0, 0.45) is not reached in this tour. Several areas show potential for minor improvement, such as the region near (0.15, 0.45) or the region near (0.4, 0.2). Working with this large data set makes some of the weaknesses of the current formulation become more evident.
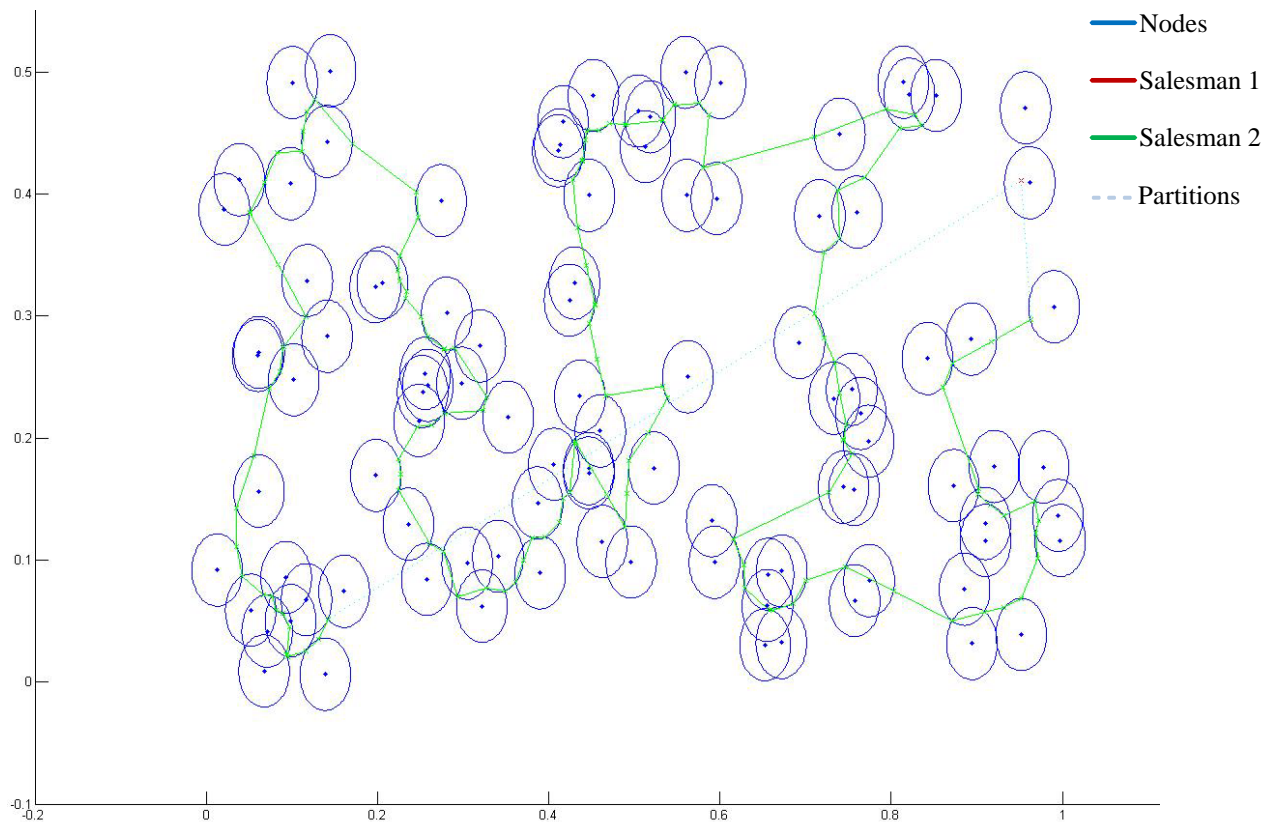

Figure 12: CETSP Solution to kroD100 from TSPLIB [5]

Table 5: Comparison of MEP Heuristic to MINLP for kroK100 Data Set

Solution	Tour Length	Calculation Time	Variation from Best Known
Euclidean Equal (Best Known)	58.54	UNKNOWN	0.00%
MEP Heuristic	64.99	949 seconds	11.02%

Further investigation is warranted in this case. Following optimization of the code implementation of this heuristic, the run time can be more accurately benchmarked against other implementations such as the heuristics that Mennell has proposed, Steiner Zones and mixed integer nonlinear programming [9][21], or the methods that Gulczynski et. al have explored [8]. At that point the extensive performance information included in the paper by Mennell will be very valuable in further validating this framework. In the discussion section a solution is proposed to help reduce the error in the CETSP approximation, potentially allowing it to much more closely approach the best known solution to this problem.

CHAPTER 6

DISCUSSION

FLEXIBILITY

The flexibility of the maximum-entropy principle approach has been clearly demonstrated. The generalized framework developed here for the multiple travelling salesmen problem can be used to approximate solutions to the basic TSP, the mTSP for any number of salesmen, as well as the close enough travelling salesman variant. Of particular relevance to this point is how only slight modifications were required to adapt this framework to the challenging close enough travelling salesman problem. Many conventional algorithms that rely on the strictly defined edges between nodes may not be able to adequately address this type of TSP and so far specialized heuristics have been required in order to solve this problem.

Extension of this framework to variants such as the single depot multiple salesmen TSP or the asymmetric TSP should be relatively straightforward. The single depot multiple salesmen TSP is a constraint which replaces the characteristic equations of several codevectors in Eq. (42). The asymmetric TSP redefines the distance function used to calculate the cost between any two codevectors, and as long as this function is continuously defined it should be able to seamlessly replace the symmetric distance function in Eq. (12).

ADVANTAGES OF MAXIMUM ENTROPY PRINCIPLE

One of the advantages of the DA algorithm as a solution to TSP is that it is easily modified to conform to the many variants of on the basic TSP. The DA for TSP solution also appears to be more computationally efficient than other heuristics currently used in solving TSPs. The current best exact solution works in approximately $\Theta(n^2 2^n)$ time, whereas in theory this algorithm should work in polynomial $\Theta(n^2)$ time, scaling only based on the number of nodes in the system, which is comparable to many other heuristics [6]. This is especially important in large data sets where the computation time can be a major impediment. The time information from Table 2 and Table 4 does not reject this theory as the computation time increases at a rate less than proportional to the square of the rate of increase in the number of nodes. Further study is warranted to verify the

38

relationship between the number of nodes and the computation time, with special consideration towards any impact on the range over which the Lagrange multipliers β and θ should be varied and the acceptable rate at which they should be varied, and the corresponding impact of these changes on computation time.

CHALLENGES OF MAXIMUM ENTROPY PRINCIPLE

The MEP heuristic framework does face the challenge that it is not based directly on the edges defined between pairs of nodes. Extending this framework to encompass the asymmetric TSP may prove difficult for those systems where the cost function is only defined along the edges. On the other hand, asymmetric cost functions that can be modeled as a continuous function such as the jet stream should not require any added complexity. Further exploration in this area is warranted to consider how this framework could be applied to those TSPs with discretely defined asymmetric cost functions.

In some cases the MEP heuristic appears to have reached a local minima rather than the global minimum of the system. For example, by comparing Figure 13 and Figure 14 the problem of non-optimal solutions becomes evident. The situation in Figure 14 is referred to as a 'twisted' solution due to the fact that the removed links at the partition between salesmen cross each other, causing each to have a an artificially high distance to the adjacent codevector. The cause of this 'twisting' in the solution is not well understood, so this is an area to explore to ensure that this phenomena does not negatively impact the quality of results. One potential explanation for the 'twisting' that has been observed could be that the Lagrange multipliers β and θ are adjusted too rapidly, causing the solution to the characteristic equations to approach a local minima instead of the global minimum. If that is the case, slowing the annealing schedule may fix this problem.
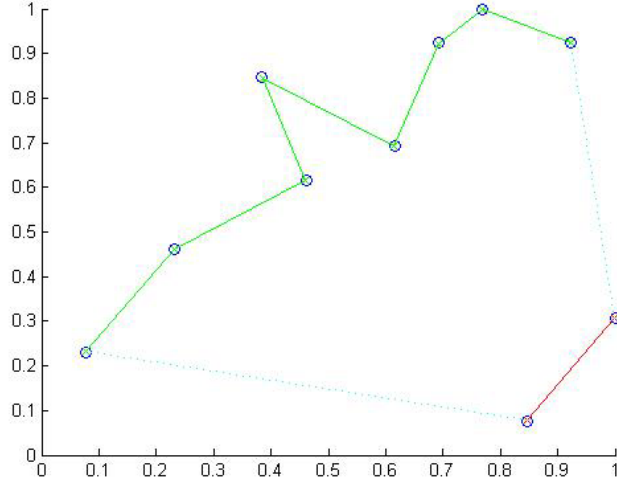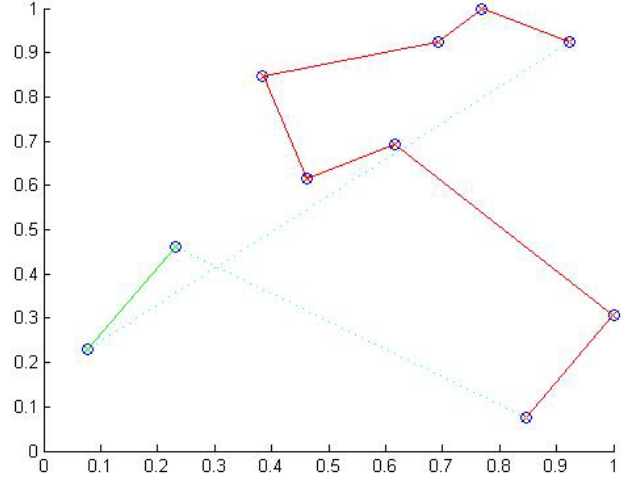
Figure 13: mTSP Optimal Solution

Figure 14: mTSP 'Twisted' non-Optimal Solution

CLOSE ENOUGH TRAVELLING SALEMAN PROBLEM ACCURACY

As the investigation has shown, the current formulation of the CETSP solution does not yet lead to satisfactory approximations of the optimal tour. One significant change is proposed to the framework for the CETSP. Currently the optimization is designed to place codevectors on the borders of the radii around each node. This means there is a penalty for a codevector existing either inside or outside of the circle. According to the problem formulation, there should be no penalty when the codevector exists within the radius of the node. In order to rectify this, the proposed change is that for values of the distance function, Eq. (12) that are less than ρ_i, the derivative with respect to y_m should be set to zero. This negates the penalty incurred for placing a codevector within the radius of a node and should help this heuristic identify more accurate solutions.

NUMERICAL ROUNDING TO ZERO

There are several numerical constraints that may cause problems if not properly addressed when implementing this algorithm. For some codevectors at significant distance from nodes, the association probability can approach zero, and if the numerical solving system rounds this to zero, the codevector will no longer be well treated in the optimization. The association

40

probability for all nodes will become zero, and the codevector will no longer be subject to the minimization of distance from the any nodes. When this occurs the codevectors are referred to as 'zero-weight' to reference the fact that the distance to the nodes now has no weight in the minimization. Zero-weight codevectors are subject solely to the tour length constraint, and therefore will be placed coincident between two non-zero-weight codevectors.

A temporary workaround was developed for this implementation. The 'zero-weight' codevectors are identified and removed. In order to avoid the problem of insufficient codevectors, the remaining codevectors are duplicated and the process repeated until the system converges without any more zero-weight codevectors. In some cases this identification and replacement process may lead to a sub-optimal solution, especially in problems such as the CETSP, so this is not an ideal resolution of this numerical challenge.

PROGRAMMING CONSTRAINTS AND STRATEGIES

One challenge that was encountered during the implementation phase was that occasionally two codevectors will coexist in the exact some position. In this case, the program may not be able to cause them to diverge in this occurs at a saddle point. To counteract this, an extremely small perturbation is applied to each codevector during the iterative solving process in order to prevent sub-optimal solutions.

In the calculation of the association probabilities, as the Lagrange multiplier β increases to large values, the exponentials can become very small. For large enough values of β this may cause these values to round to zero, and if the normalization parameter Z rounds to zero it causes an error. To counteract this the minimum distortion is subtracted from the distance in all of the exponentials so that at least one of the exponentials will calculate as $e^{(-0)} = 1$. This has no effect on the probability function but avoids problems when the program rounds very small numbers. As a complementary adjustment that is also aimed at preventing the program from handling very large or very small values, the data for the nodes is normalized by the maximum value within that set.

COMPUTATION TIME AND SCALABILITY

For this MATLAB implementation, the computation time of the code was not optimized. Significant gains in computation time can be made by replacing nested FOR loops with matrix operations. Additionally, moving this to another software language such as C could also provide operating time improvements.

Currently the Lagrange multipliers β and θ are adjusted according to an exponential function. This is not the most efficient method for adjusting these parameters as many extra iterations are performed. A sizable number of the iterations produce little to no change in the solution to the characteristic equations for the codevectors, while certain iterations produce significant changes. These are referred to as phase transitions where relatively small changes in the Lagrange multipliers cause the characteristic equations to change drastically and reveal a new global minimum. Adaptive Deterministic Annealing (ADA) has been developed to address this phase transition property of the system [22]. ADA first identifies the phase transitions and then selectively adjusts the Lagrange multipliers. Near the phase transitions the incremental changes to the Lagrange multipliers are small, but far from the phase transitions much larger increments can be used. This cuts down on the total number of iterations required to achieve the desired values of β and θ, thereby increasing the speed at which the solution is found. Additionally, if the rate at which the Lagrange multipliers bears partial responsibility for the occurrence of 'twisting,' this may be an effective way to fix that problem without significant time penalties. Applying lessons and methodologies from ADA to this framework will allow for significantly faster computation times, so this is strongly recommended as a future development on this framework.

Another consideration for scalability is the computation of the interaction between distant codevectors and nodes. Due to the nature of the Gibbs distribution, faraway components play a very small role in the specifying a codevector. Working on a similar algorithm based in the MEP, Sharma et. al proposed a scheme to neglect the contributions of faraway components [23] in order to improve the scalability of the algorithm. Additionally, the relationship between the potential error and the reduced computation time is explored.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

As a result of this research it is evident that the maximum-entropy principle presents some useful opportunities as a heuristic for the travelling salesman problem, as well as the many variants. Because this algorithm is independent of the edges between nodes it has more flexibility to address variants such as the close enough travelling salesman problem and the multi travelling salesman problem.

The framework developed here provides the necessary tools for solving the TSP and variants with the maximum-entropy principle. There remain significant opportunities to optimize the code implementation of this framework to achieve more favorable computation times, at which point this algorithm can be run on benchmark mTSP cases and compared against many of the conventional heuristics. Benchmarking will be critical in evaluating the computational advantages and disadvantages of this approach to the mTSP.

The next steps for this heuristic framework should be developing the formulation for further variants on the basic TSP. Because one of the significant advantages this framework provides is the greater flexibility to adapt to variants of the basic TSP, effort should be focused on expanding the number of compatible variants. Further, consideration should be given to the question of whether a single mathematical construct can be successfully implemented to cover the many variants on the TSP, along with the related question of how valuable that kind of framework would be for the field of TSP heuristics.

Finally, hybridization should be considered. Further development will cause the strengths and weaknesses of this heuristic to become more apparent, whether those are based in computation time, flexibility, complexity, accuracy, etc. It is suggested that the strengths of conventional heuristics should be explored to determine whether any of them complement this heuristic such that a hybrid model could be more successful than either of the individual heuristics.

REFERENCES

[1] Bektas, T. "The multiple traveling salesman problem: an overview of formulations and solution procedures." <u>Omega</u> 34.3 (2006): 209-219.

[2] Junger, M., Reinelt, G., Rinaldi, G. "Chapter 4: The Traveling Salesman Problem." <u>Handbooks in Operations Research and Management Science</u>. Vol. 7. Elsevier Science B.V., 1995: 225-330.

[3] Applegate, D.L., et al. "Certification of an Optimal TSP Tour through 85,900 Cities." <u>Operations Research Letters</u> 37.1 (2009): 11-15.

[4] <u>The Traveling Salesman Problem</u>. Cook, W. June 2011. Georgia Tech. Sept. 2011 <http://www.tsp.gatech.edu/index.html>.

[5] <u>TSPLIB.</u> Research Group Discrete Optimization. August 2008. University of Heidelberg. Sept. 2011 <http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/>.

[6] Johnson, D. S., McGeoch, L. A. "Experimental Analysis of Heuristics for the STSP." <u>The Traveling Salesman Problem and Its Variations</u>. Ed. Punnen A. P. Gutin G. Boston: Kluwer Academic Publishers, 2002. 369-443.

[7] Bellmore, M., Hong, S.,. "Transformation of a multisalesmen problem to the standard traveling salesman problem." <u>Journal of the Association of Computer Machinery</u> (1974): 500-504.

[8] Gulczynski, D., Heath, J., Price, C., "Close-Enough Traveling Salesman Problem: A Discussion of Several Heuristics." <u>Perspectives in Operations Research: Papers in Honor of Saul Gass' 80th Birthday</u> (2006): 271-283.

[9] Mennell, W.K. "Heuristics for solving three routing problems: Close-Enough Traveling Salesman Problem, Close-Enough Vehicle Routing Problem, Sequence-Dependent Team Orienteering Problem." <u>Phd Thesis</u>. College Park: University of Maryland, 2009.

[10]   Cirasella, J., et al. "The Asymmetric Traveling Salesman Problem: Algorithms, Instance Generators, and Tests." Algorithm Engineering and Experimentation, Third International Workshop, <u>ALENEX 2001.</u> Berlin: Springer, 2001. 32-59.

[11]   Yadlapalli, S., Malik, W.A., Darbha, S., Pachter, M., "A Lagrangian-based algorithm for a multiple depot, multiple travelling salesmen problem." <u>Proceedings of the American Control Conference</u>. 2007: 4027-4032.

[12]   Jaynes, E. T. "Information Theory and Statistical Mechanics." <u>The Physical Review</u> 106.4 (1957): 620-630.

[13]   Rose, K., "Deterministic Annealing, Clustering, and Optimization." <u>PhD Thesis</u>. Pasadena, CA: California Institute of Technology, 1991.

[14]   Tang, L., Liu, J., Rong, <u>A</u>., Yang, Z. "A Multiple Traveling Salesman Problem Model for Hot Rolling Scheduling in Shanghai Baoshan Iron & Steel Complex." <u>European Journal of Operational Research</u> (2000): 267-282.

[15]   Gorenstein, S. "Printing Press Scheduling for Multi-Edition Periodicals." <u>Management Science</u> (1970): 373-383.

[16]   Jaynes, E.T. "Chapter 11: Discrete Prior Probabilities - The Entropy Principle." <u>Probability Theory - The Logic of Science</u>. Cambridge University Press, 2003. 1101-1125.

[17]   Rose, K., "Compression, Classification, Regression and Related Optimization Problems." <u>Proceedings of the IEEE</u> 86.11 (1998): 2210-2239.

[18]   Applegate, D.L., et al. "Chained Lin-Kernighan for large traveling salesman problems." <u>INFORMS Journal on Computing</u> 15.1 (2003): 82-92.

[19]   Junjie, P., Dingwei, W. "An Any Colony Optimization Algorithm for Multiple Travelling Salesman Problem." <u>Proceedings of the First International Conference on Innovative Computing, Information and Control</u>. IEEE, 2006.

[20]   Song, C., Lee, K., Lee, W.D. "Extended Simulated Annealing for Augmented TSP and Multisalesmen TSP." Proceedings of the IEEE (2003): 2340-2343.

[21]   Mennell, W.K., Ciric, A.R. "Mixed-Integer Nonlinear Programming Models for the Close-Enough Travelling Salesman Problem." Report. 2009.

[22]   Chang, W., et al. "Adaptive Determininstic Annealing for two Applications: Competing SVR of Switching Dynamics and Travelling Salesman Problems." Proceedings of the 9th International Conference on Neural Information Processing. 2002: 920-924.

[23]   Sharma, P., Salapaka, S., Beck, C. "A Scalable Approach to Combinatorial Library Design for Drug Discovery." Journal of Chemical Information and Modeling (2008): 27-41.