from *SIAM News*, Volume 36, Number 7, September 2003

A Tour of TSP Results Makes for Optimal Community Lecture

By James Case

Opening the 2003 I.E. Block Community Lecture, "The Traveling Salesman Problem and Optimization on a Grand Scale," Bill Cook of Georgia Tech pointed out that actual traveling salesmen have rarely given much thought to routing problems—unfriendly dogs and unpleasant accommodations rank higher on their lists of concerns. As evidence, he cited a variety of 19th- and 20th-century guides and manuals for the business traveler, few of which devoted more than a passing remark to routing issues.

Among the first to recognize the problem's potential import was Austrian mathematician Karl Menger, who called it "the messenger problem" in 1930. Hassler Whitney referred to it as "the 48 states problem" in 1931 and 1932, and Merrill Flood ran across it in 1937 while routing New Jersey school buses. Practitioners P.C. Mahalanobis in Bengal and R.J. Jessen in Iowa encountered a purer form of the problem during the 1940s, while seeking to determine the least time-consuming order in which a single team of soil testers (as opposed to multiple school buses) might visit (along with their bulky equipment) each of an assigned list of testing sites. Routing considerations were particularly important to them because soil testing itself is not very time consuming, meaning that testers spend most of their time on the road. Flood later worked extensively on the TSP—mainly at RAND in the aftermath of World War II—but confessed to A.W. Tucker near the end of his life that he didn't know who first bestowed the "peppier name" on what he himself always thought of as "Whitney's problem."

The TSP continues to be of interest for a variety of reasons: It is among the easiest of the NP-complete problems to state, it has important applications of its own, it often forms an identifiable part of other practical problems, and it serves as a convenient test problem for general methods of discrete optimization.

The TSP is most easily stated in terms of the complete graph on N vertices. Given (for each $1 \leq i \neq j \leq N$) the cost c_{ij} of traversing the edge ij between vertices i and j, one can compute the cost of traversing any path Γ composed of adjacent edges by adding up the costs c_{ij} for which $ij \in \Gamma$. Missing or non-navigable edges ij can be accommodated by writing $c_{ij} = \infty$. It then makes sense to ask—as the TSP does—for the lowest-cost path Γ that visits each vertex exactly once before returning to its starting point. Such a path will naturally consist of exactly N edges.

A TSP is called symmetric if $c_{ij} = c_{ji}$ for each ij, and geometric if the c_{ij}'s are in constant proportion to the Euclidean distances $\rho(i,j)$ between points (vertices) i and j in the plane. In that case—though not in general—the cost matrix will satisfy the triangle inequality $c_{ij} + c_{jk} \geq c_{ik}$, which simplifies the TSP slightly. Because the edges in an optimal tour Γ^* can never cross in a geometric TSP,* every optimal tour constitutes a simple closed curve in the plane, and by virtue of the Jordan curve theorem must surround a bounded subset thereof. One can therefore represent Γ^* by shading the bounded set, as is done for the optimal tour of 15,112 German cities shown in Figure 1.

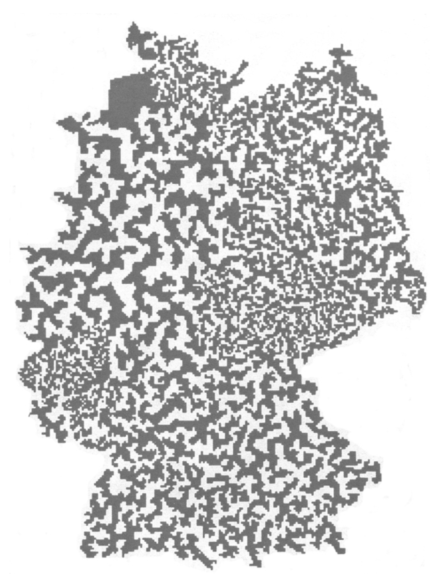

Figure 1. *Jordan curve representation of the optimal tour of 15,112 German cities.*

An obvious application of TSP software is to the coin-collection problem faced by a pay telephone owner. It is natural for anyone owning a substantial number of such phones to seek an optimal (shortest-time) tour of their several locations. One-way streets and traffic patterns that change with the time of day require that standard TSP solvers be modified for the purpose. Another application was found by a baseball fan, who designed an optimal tour of Major League Baseball's 30 active ballparks. I am not aware that the plan has been mo-dified for the 2003 season, during which the Montreal franchise played a number of its home games in Puerto Rico.

More important applications arise in genome sequencing, genetic engineering, VLSI fabrication, and space science, to name but a few. As an example of the latter, Cook mentioned that a team of engineers from Hernandez Engineering in Houston and Brigham Young University have experimented with the use of TSP software to optimize the sequence of celestial objects to be imaged by two satellites in the proposed NASA Starlight space interferometer program. Their goal was to minimize the fuel consumed by the satellites (interferometers) as they train in unison on each in turn of the designated celestial objects selected for observation on a given date. The "cities" in this application are the celestial objects to be observed, while c_{ij} is the quantity of fuel required to retarget a satellite from object i to object j. No modification of standard TSP software was required for this project.

Other applications require substantial modification. A team consisting of Cook, David Applegate, Sanjeeb Dash, and Andre Rohe, for instance, recently received a significant cash prize for demonstrating that the winning solution in the 1996 Whizzkids

*If edges ik and jl did cross, they would constitute the diagonals of a quadrilateral $ijkl$, and could profitably be replaced by a pair of opposite sides.

competition was in fact optimal. The problem consists of finding the best collection of routes for four newsboys to deliver papers to 120 customers.

An important milestone in TSP history was reached in 1954, when George Dantzig, Ray Fulkerson, and Selmer Johnson described a method for solving TSPs, and demonstrated its power by solving a 49-city example. This was an impressive feat at the time, given the computing power then available. Considering only the symmetric TSP, the researchers strung the c_{ij}'s out into a (column) vector c of dimension $N(N - 1)/2$, and considered its inner product $c'x$ with an indicator vector x of equal length. They set the component of x corresponding to an edge ij equal to 1 if they wanted to include that edge in the current tour, and to 0 otherwise. This associated every subset of the edge set with a vertex of the $N(N - 1)/2$-dimensional hypercube H, and allowed them to observe that the set S of allowable tours corresponds to the intersection of $v(H)$—the vertex set of H—with a certain linear manifold M in the Euclidean space containing H. It also allowed them to rephrase the TSP as follows:

$$\text{Minimize } c'x \text{ subject to } x \in S. \tag{1}$$

Indeed, they could replace the discrete feasible set S in the specification (1) with $K = co(v(H) \cap M)$—where $co(W)$ denotes the convex hull of W—because the minimum value of a linear functional over a compact convex feasible set is always attained at an extreme point of the latter.

They then relaxed the constraints on x to allow its components to assume values in the interior of the interval [0,1], as well as at its endpoints, to arrive at the problem

$$\text{Minimize } c'x \text{ subject to } Ax \leq b, \tag{2}$$

where $Ax \leq b$ is a system of linear equalities and inequalities satisfied by all $x \in S$, and observed that the solution of (2)—which they were able to obtain with the help of the simplex method—furnishes a lower bound on the solution of (1). Indeed, (1) differs from (2) only in specifying a more exclusive set of feasible vectors x. Finally, the researchers observed that if $\hat{x} \notin K$ is a solution of (2), then $\hat{x}$ can be separated from K by a hyperplane $a'x = b$ and the inequality $a'x \leq b$ can be appended to the system $Ax \leq b$ to yield a more exclusive version of (2), furnishing a necessarily tighter lower bound on the solution of (1). And solution of the modified version of (2) gives a still tighter lower bound on the solution of (1), and so on. The hope is therefore that, by iterating the foregoing procedure, researchers will eventually encounter a version of (2) whose solution does belong to K, and therefore solves (1) as well as (2).

The problem for which Dantzig, Fulkerson, and Johnson carried all this out involved cities located in each of the lower 48 states, plus Washington, DC, with distances obtained from a road atlas. They simplified the problem somewhat by removing Baltimore, Wilmington, Philadelphia, Newark, New Jersey, New York, Hartford, Connecticut, and Providence, Rhode Island, from their map, solving the reduced problem, and then inferring a solution to the original problem from the fact that the missing cities all lie along the shortest path from Washington to Boston.

The three researchers thanked I. Glicksberg for suggesting the method by which they constructed the hyperplanes $a'x = b$ needed to generate the requisite sequence of type (2) problems. In his community lecture, Cook pointed out that any integer linear program can be attacked in more or less the same way—given an adequate method for generating separating hyperplanes—and remarked on the "breathtaking elegance" with which Ralph Gomory's *cutting-plane algorithms* satisfy this need.

Cutting-plane methods, although they are effective for problems of small to moderate size and furnish demonstrably optimal solutions on termination, seldom terminate within an acceptable time when applied to large problems. As a result, ongoing research tends to focus on approximate methods capable of solving remarkably large TSPs to within a fraction of 1%. To demonstrate the effectiveness of such methods, Cook exhibited a tour of 24,978 cities in Sweden that was 855,597 km long, and observed that the software that generated it furnished the additional information that no tour can be shorter than 855,528 km. The TSP is therefore solved to within an "optimality gap" of 0.008%. Methods that do this consist of two parts: an algorithm for generating low-cost tours, and another for producing lower bounds on the cost of any tour.

The method Cook described for generating low-cost tours is an iterative one, beginning with an arbitrary tour and improving it through a succession of "k-swaps." A k-swap is performed by removing k edges from the current tour and replacing them in all possible ways to find the cheapest alternative. It is followed by another such swap, and then another, until either the optimality gap is acceptably small or the allotted time has expired. Such methods were first explored in 1973 by S. Lin and B.W. Kernighan, who employed only 2-swaps, and improved in 2000 by K. Helsgaun, who advanced to 5-swaps. Such advancement is complicated by the fact that the number of possible k-swaps grows rapidly with k, being 1 when $k = 2$, 148 when $k = 5$, and 2,998,656 when $k = 9$. Cook and his collaborators found the Sweden tour using 9-swaps.

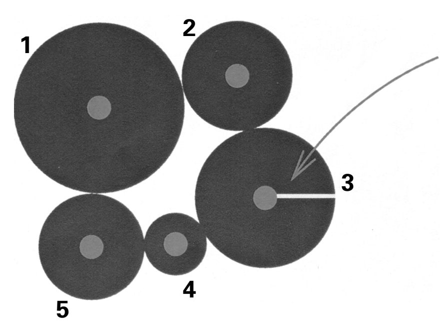

Figure 2. *No tour of the cities* 1,2,3,4,5 *can be shorter than* $2r_1 + 2r_2 + 2r_3 + 2r_4 + 2r_5$ *cm in length.*

The problem of finding acceptable lower bounds on the cost of any tour is most easily addressed for geometric TSPs. If, for instance, the points 1,2,3,4,5 can be surrounded by disjoint discs of radius r_1, r_2, r_3, r_4, r_5 cm, respectively, as indicated in Figure 2, then no tour of 1,2,3,4,5 can be shorter than $2r_1 + 2r_2 + 2r_3 + 2r_4 + 2r_5$ cm, because it is impossible to get from i to j without traveling $r_i + r_j$ cm. Moreover, the radii of the largest disjoint discs surrounding the points 1,2,3,4,5 in Figure 2 can be found by solving the following LP:

$$\text{Maximize} \quad 2r_1 + 2r_2 + 2r_3 + 2r_4 + 2r_5$$

subject to

$$r_i + r_j \leq \rho(i,j) \quad \forall\, 1 < i \neq j < 5$$

$$\text{and } r_i \geq 0 \ \forall\ i.$$

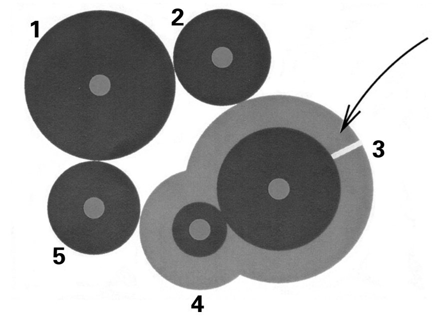

Figure 3. *The portion of any tour that enters and exits the area enclosed by the indicated moat, and that visits en route all the cities within, must be at least* $2(r_1 + r_2 + r_3 + r_4 + r_5) + 2w$ *cm in length.*

This idea can be extended to groups of cities that seem likely (if not certain) to be visited in succession by surrounding them with "moats" of width w, as indicated in Figure 3. Every tour of those five cities has length at least $2(r_1 + r_2 + r_3 + r_4 + r_5) + 2w$, and the optimal values of r_1, r_2, r_3, r_4, r_5, and w can again be determined by solving an appropriate LP.

It was by combining the foregoing insights with a number of more (pedestrian and) familiar ones that Cook and his collaborators obtained their quite excellent lower bound for tours of Sweden. Such pictures are most informative when colorized, and Cook combined a host of colored pictures with a sampler of exciting results to create a community lecture that both informed and entertained.

Readers interested in exploring the topic further are encouraged to begin with Bill Cook's Web site: www.math.princeton.edu/tsp.

James Case writes from Baltimore, Maryland.