On the Analytic Geometry of the Traveling Salesman Problem

Kathleen Bellino
James Madison University

Rekha Narasimhan
Arizona St. University

Introduction

The traveling salesman problem refers to a situation where a salesman visits each of n given cities exactly once and returns to his starting point in such a way that the total distance traveled is minimal. Equivalent problems arise in many industries that deliver goods or services to customers. In this paper, we shall examine the case where a company is interested in building a warehouse or distribution center that will deliver products to a set of customers on a regular basis. Many factors are involved when choosing a building site for a warehouse. For example, the cost of making deliveries from the location site is important. Thus, for a fixed number of clients, the company might want to know the cost of supplying them with a product from several potential warehouse sites. If there are n customers, this requires the solution of a traveling salesman problem to n cities from various starting points. This motivates one to consider the following problem. Let (ai,bi) denote the location of city i in the xy plane, and let $H = (x,y)$ denote a movable point in the plane, which we will refer to as the home point. For each position of H, solve the traveling salesman problem, which starts at H, visits the n cities, and returns to H. Now, color the points in the plane so that two points get the same color if their traveling salesman tours visit the cities in the same order. Clearly, only a finite number of colors will be required for any n. The problem is to determine some interesting properties of the regions defined by these colors.

The Restricted Problem

We will solve this problem for the case n = 3. Although this is the simplest case of interest, it poses some intriguing questions. For instance, it would be of interest to know if the region corresponding to a certain color is ever disconnected, and if so, what is the configuration of the three cities that would cause this? We will also investigate the multicolored points, points for which the traveling salesman problem has more than one solution.

Let A, B, and C be the cities and H represent the warehouse site. We will assume that A, B, and C are non-collinear, and therefore form a triangle. The object is to minimize the total traveling distance from $H(x,y)$ to the three cities and back. We will use the Euclidean distance between points.

There are six possible routes the traveling salesman can take. However, observe that H-A-B-C-H and H-C-B-A-H represent equivalent routes. Therefore, there are only three distinct routes, namely

$H \rightarrow A \rightarrow B \rightarrow C \rightarrow H$
$H \rightarrow A \rightarrow C \rightarrow B \rightarrow H$
$H \rightarrow B \rightarrow A \rightarrow C \rightarrow H$

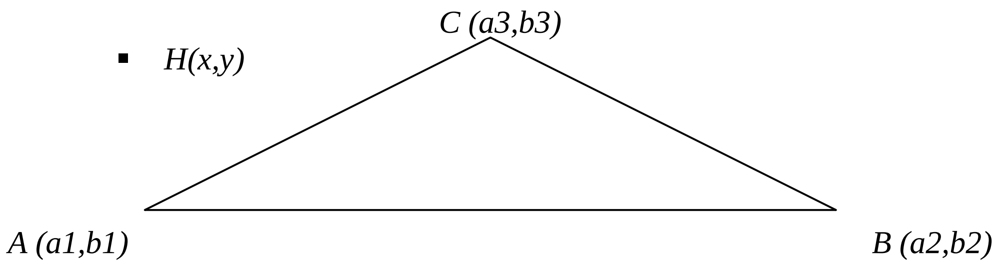


Figure 1
Triangle with vertices (a1,b1), (a2,b2), (a3,b3)

We denote the lengths of the tours over these three routes by d1, d2, and d3, respectively. Using the distance formula, we have

$$
\begin{aligned}
d1 &= \sqrt{(x-a1)^2+(y-b1)^2} + AB + BC + \sqrt{(x-a3)^2+(y-b3)^2} \\
d2 &= \sqrt{(x-a1)^2+(x-b1)^2} + AC + BC + \sqrt{(x-a2)^2+(y-b2)^2} \\
d3 &= \sqrt{(x-a2)^2+(y-b2)^2} + AB + AC + \sqrt{(x-a3)^2+(y-b3)^2}
\end{aligned}
$$

Graphics of Special Cases:

Using Matlab, we will generate the colored regions mentioned in the introduction. The rule for coloring the regions is the following: If d1 is the length of the shortest route from a certain starting point H, a red point is plotted at H. If d2 is the length of the shortest route, a blue point is plotted at H. Similarly if d3 is the length of the shortest route from H, a green point is plotted at H. When this is repeated for all points in the plane, colored regions appear. We will use Matlab graphics to observe interesting phenomena about these regions, then prove mathematically that these phenomena always occur.

The Equilateral Case:

We begin by examining the simplest case where A, B and C form an equilateral triangle. From observing the graphics for this case, it appears that the colored regions intersect along straight lines that meet at a point. We will prove that this is true.

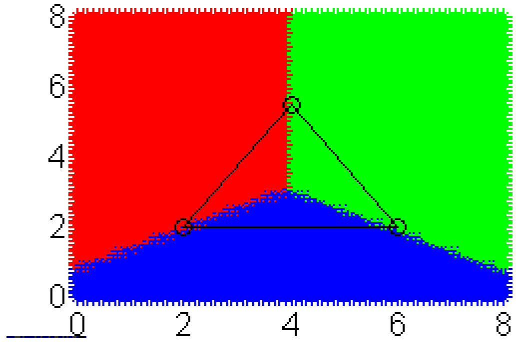


Figure 2
Triangle with vertices *A(2,2), B(6,2), C(4, 2+2$\sqrt{3}$)*

In Figure 2, the boundary between the red and blue regions is defined by the equation d1=d2. We will show that this boundary is a straight line.

Setting d1=d2, we obtain

$$\sqrt{(x-a1)^2+(y-b1)^2}+AB+BC+\sqrt{(x-a3)^2+(y-b3)^2}$$
$$=\sqrt{(x-a1)^2+(y-b1)^2}+AC+BC+\sqrt{(x-a2)^2+(y-b2)^2}$$

Since AB=BC=AC, it follows that

$$\sqrt{(x-a3)^2+(y-b3)^2}=\sqrt{(x-a2)^2+(y-a2)^2} \tag{2.1}$$

Simplifying (1.1), we have

$$2(a3-a2)x+2(b3-b2)y=a3^2-a2^2+b3^2-b2^2 \tag{2.2}$$

Since our triangle is equilateral we know

$$(a1-a2)^2+(b1-b2)^2=(a1-a3)^2+(b1-b3)^2$$

which implies that

$$a3^2-a2^2+b3^2-b2^2=2a1(a3-a2)+2b1(b3-b2)$$

Substituting this into (2.1) we see that the equation of the line separating the red and blue regions is given by

$$(a3-a2)x+(b3-b2)y=a1(a3-a2)+b1(b3-b2)$$

Similar reasoning shows that the green and blue regions intersect along the line

$$(a1 - a3)x + (b1 - b3)y = a2(a1 - a3) + b2(b1 - b3)$$

and the green and red regions intersect along the line

$$(a1 - a2)x + (b1 - b2)y = a3(a1 - a2) + b3(b1 - b2).$$

Using linear algebra and Cramer's rule, it is not hard to verify that these three lines intersect in a unique point.

The Isosceles Case

The next case we shall explore is the isosceles triangle. Consider both the equilateral triangle formed by the points A, B, and C and the colored regions in Fig 2. We want to determine what happens to these regions when we move point C along the line x=4. When the angle at C is acute, our colors define three connected regions, as in the equilateral case. The regions appear to intersect along straight lines, as in Fig 3. By zooming in, however, we can see that two of the intersections are actually curved. We can show this mathematically as follows.

By setting d1 = d2, d2 = d3, d1 = d3 and simplifying, we have the general equations of the intersections for any configuration of the cities.

$$4(AB - AC)^2[(x - a3)^2 + (y - b3)^2] = [(2x - a3 - a2)(a3 - a2) + (2y - b3 - b2)(b3 - b2) - (AB - AC)^2]^2$$
$$4(BC - AB)^2[(x - a1)^2 + (y - b1)^2] = [(2x - a1 - a3)(a1 - a3) + (2y - b1 - b3)(b1 - b3) - (BC - AB)^2]^2$$
$$4(BC - AC)^2[x - a1)^2 + (y - b1)^2] = [(2x - a1 - a2)(a1 - a2) + (2y - b1 - b2)(b1 - b2) - (BC - AC)^2]^2$$

Since AC=BC, we see that the third equation, which describes the boundary between the red and green regions, is the equation of a straight line. The other two equations describe hyperbolas. The graphs of these equations are plotted in Fig. 4 using Maple.

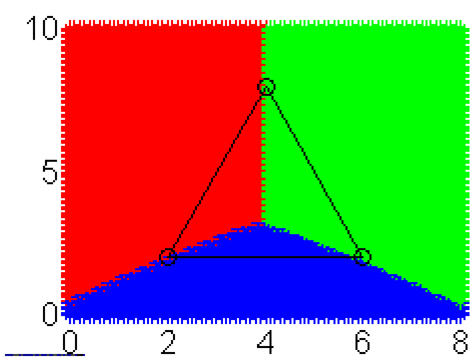


Figure 3
A(2,2), B(6,2), C(4,8)

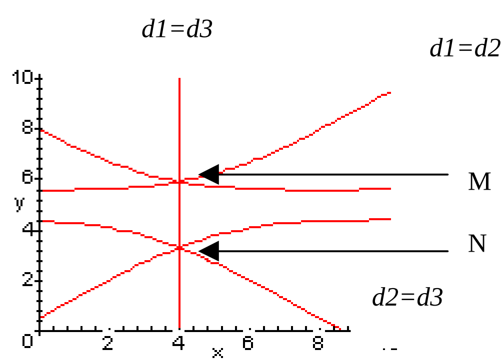


Figure 4

By comparing Fig. 3 with Fig. 4, we see that the red and blue regions are separated by a branch of the first hyperbola, and the blue and green regions are separated by a branch of the second hyperbola.

The Obtuse Case

We would expect a similar result when the angle at C is obtuse. However, this is not the case. Surprisingly, when the angle at C is sufficiently obtuse, the red and green areas split apart, and a second blue region appears between them, as in Fig. 5. In other words, the distance d2 from a point H in the upper blue region is shorter than either d1 or d3. We shall now investigate the boundaries of these regions, along with the exact point at which the split occurs.

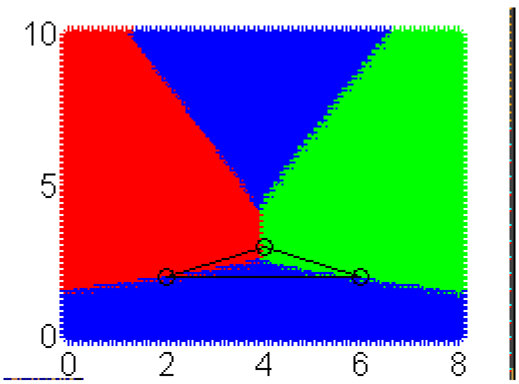


Figure 5
A(2,2), B(6,2), C(4,3)

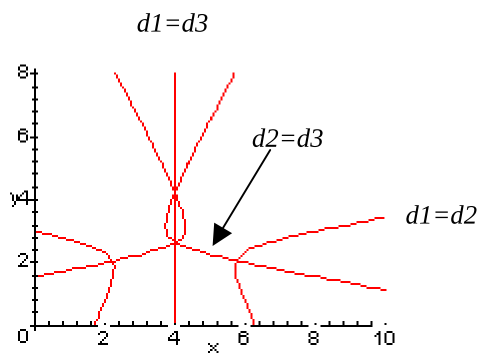


Figure 6

Where Does the Blue Region Disconnect?

We are interested in finding out where the blue region disconnects. Before explaining this occurrence, we recognize a special isosceles triangle that plays an important role in our explanation.

By adjusting the height of point C in Fig. 1, we can cause the hyperbolas in Fig. 4 to undergo some very specific changes. For the relatively high position of C in Fig. 3, the hyperbolas d1=d2 an d2=d3 have two intersections with the line d1=d3. These intersections are indicated by M and N in Fig. 4. As the position of C is lowered, the intersection point N moves off towards $-\infty$. When C is lowered to the point (4, 7/2), the intersection point N disappears at $-\infty$, and we have the situation shown in Fig. 7.

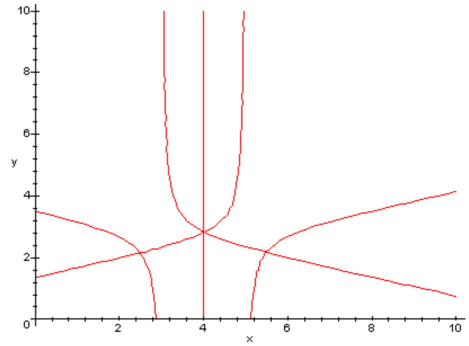


Figure 7

Notice that in this figure, only one branch of each hyperbola intersects the line d1=d3. The hyperbolas of this special case also have a unique property; their asymptotes are parallel to each other and to the intersection x = 4. (Fig. 8,9).

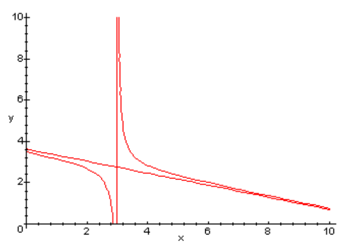


Figure 8

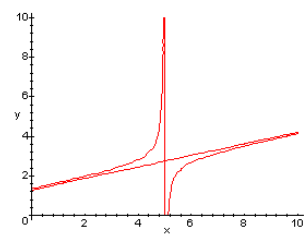


Figure 9

Once point C of the triangle is lowered below (4, 7/2), each of the branches has two points of intersection with d1=d3 again.

When C is lowered below the point (4, 7/2), the two branches of the hyperbola that intersect with d1=d3 divide the plane into four unbounded regions. We are going to show that point (4,7/2) is critical in this regard. In this case, it is instructive to consider the shape of the triangle ABC when C = (4, 7/2); ABC is composed of two triangles that are similar to a 3-4-5 triangle.

Disconnected Regions Theorem

We shall use the 3-4-5 triangles in the following claim.

By taking a point just below the line AB in Fig. 11, it is not hard to show that the optimal tour must be d2, H–B-C-A-H, since this path does not cross itself. Then the color of this point must be blue, since d2 corresponds with blue. We will show that for h large enough, the point, $H = (4p,h)$ must also be blue for $0< \varepsilon <3p$.

Thm 1.1: For any isosceles triangle with base 8p (for any real number p > 0,), if the height is less than 3p, the colored region below the base disconnects.

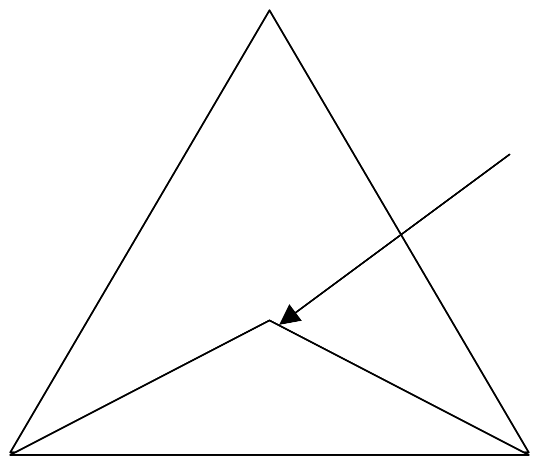


Figure 11

Proof:

We will show that for h sufficiently large, d2<d1 and d2<d3. Moreover, the point (4p,h) must be colored blue (see Fig. 5).

By applying the distance formula to the line segments in Figure 8, we have

$$HB = \sqrt{(4p-8p)^2 + (h-0)^2}$$

$$= \sqrt{16p^2 + h^2}$$

$$AC = \sqrt{(4p-0)^2 + ((3p-\varepsilon)-0)^2}$$

$$= \sqrt{16p^2 + 9p^2 - 6p\varepsilon + \varepsilon^2}$$

$$= \sqrt{25p^2 - 6p\varepsilon + \varepsilon^2}$$

Thus, $HB + AC = \sqrt{16p^2 + h^2} + \sqrt{25p^2 - 6p\varepsilon + \varepsilon^2}$

Also,

$$HC = \sqrt{(4p - 4p)^2 + (h - (3p - \varepsilon))^2}$$

$$= h - 3p + \varepsilon$$

$$BA = 8p$$

Thus, $HC + BA = h - 3p + \varepsilon + 8p = h + \varepsilon + 5p$

Note that $0 < \varepsilon < 3p$.

Then $-6p\varepsilon + \varepsilon^2 < 0$.

So, $\sqrt{25p^2 - 6p\varepsilon + \varepsilon^2} < 5p$

Also, $\sqrt{16p^2 + h^2} = h\sqrt{1 + \frac{16p^2}{h^2}}$, *and* $h + \varepsilon = h\left(1 + \frac{\varepsilon}{h}\right)$

We can choose h large enough so that

$$\sqrt{1 + \frac{16p^2}{h^2}} < 1 + \frac{\varepsilon}{h},$$

or equivalently if

$$1 + \frac{16p^2}{h^2} < 1 + \frac{2\varepsilon}{h} + \frac{\varepsilon^2}{h^2}$$

This inequality holds if we choose h large enough such that

$$\frac{16p^2}{h^2} < \frac{2\varepsilon}{h}$$

$$or\ \frac{16p^2}{h} < 2\varepsilon$$

Thus *HB* + *AC* < *HC* + *BA* if we choose h large enough such that $\frac{16p^2}{h} < 2\varepsilon$

This inequality implies that *HB* + *BC* + *CA* + *AH* < *HC* + *CB* + *BA* + *AH.*

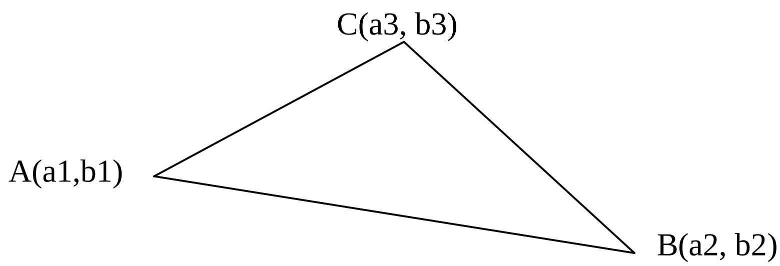
Similar reasoning can be used to show that d2<d3, or

HB + *BC* + *CA* + *AH* < *HB* + *BA* + *AC* + *CH, or d2<d1.* □

If the y value at point C is 3p or 3p+ε , then the inequality d2<d1 or d2<d3 is always false. Thus the height 3p is a critical point at which the colors split.

The Cities and Hyperbolas – A Connection

By experimenting and using Maple, we come across an interesting connection between the cities and hyperbolas. For each hyperbola, two of the cities are the foci, while the third city is a point on the hyperbola. We will show that this observation is always true.



By definition, a hyperbola is the set of all points in the plane, the difference of whose distances from two fixed points (the foci) is a given positive distance (2a). A hyperbola can be uniquely determined by its foci and any point on the hyperbola. Consider the above triangle and suppose there exists a hyperbola containing the point B(a2, b2) with A(a1, b1) and C(a3, b3) as its foci. We will show that this hyperbola is one of the hyperbolas created by the intersections of the distances. Using the definition and an arbitrary point P(x,y) on the hyperbola, we derive the equation of this hyperbola,

$$\left|\sqrt{(x-a1)^2+(y-b1)^2}-\sqrt{(x-a3)^2+(y-b3)^2}\right| = 2a$$

where 2a is the given positive distance.

We can use the point B(a2,b2) on the hyperbola to solve for a.

$$\left|\sqrt{(a2-a1)^2+(b2-b1)^2}-\sqrt{(a2-a3)^2+(b2-a3)^2}\right| = 2a$$

$$|AB-BC| = 2a$$

$$a = \frac{|AB-BC|}{2}$$

Thus, the equation for the hyperbola with foci A,C and pt. B is

$$\left|\sqrt{(x-a1)^2+(y-b1)^2}-\sqrt{(x-a3)^2+(y-b3)^2}\right| = 2\left(\frac{|AB-BC|}{2}\right)$$

$$\left|\sqrt{(x-a1)^2+(y-b1)^2}-\sqrt{(x-a3)^2+(y-b3)^2}\right| = |AB-BC|$$

Case 1:

$$\sqrt{(x-a1)^2+(y-b1)^2}-\sqrt{(x-a3)^2+(y-b3)^2}=AB-BC$$

$$\sqrt{(x-a1)^2+(y-b1)^2}+BC=AB+\sqrt{(x-a3)^2+(y-b3)^2}$$

$$\sqrt{(x-a1)^2+(y-b1)^2}+AC+BC+\sqrt{(x-a2)^2+(y-b2)^2}$$
$$=\sqrt{((x-a2)^2+(y-b2)^2}+AB+AC+\sqrt{(x-a3)^2+(y-b3)^2}$$

d2 = d3.

Case 2:

$$\sqrt{(x-a1)^2+(y-b1)^2}-\sqrt{(x-a3)^2+(y-b3)^2}=BC-AB$$

$$\sqrt{(x-a1)^2+(y-b1)^2}+AB=BC+\sqrt{(x-a3)^2+(y-b3)^2}$$

Note that Case 2 is the other half of the hyperbola that is not present in the intersection of the colored regions.

Conclusion

We have examined the equilateral and isosceles cases in detail and can fully explain the intersections and disconnected colored regions for three fixed cities. To take this problem a step further, one can explore the scalene triangle and its geometry in the traveling salesman problem. Moreover, one can look at the case of four or more fixed cities and explain the different colored regions that appear. In any case, a movable home point adds a new perspective to the traveling salesman problem that is open for more discoveries.

Acknowledgements

We would like to thank our advisor Dr. Earl Barnes for suggesting this problem and guiding us through the research process. According to Dr. Barnes, the problem was first posed to him by Gary Parker of Georgia Institute of Technology. We would also like to thank Dr. Dennis Davenport, coordinator of SUMSRI, for giving us the opportunity to participate in this program at Miami University. Finally, we want to thank our graduate assistant, Thayer Morrill, for reading and editing numerous drafts of this paper.