The Traveling Salesman at a Home Near You:
Investigating the Traveling Salesman Problem with Varying Home Points

Natalie Collar	Manchester College
Ashlee Edwards	Virginia Wesleyan College
Jaime Jeke	University of Pittsburgh
Shana Johnson	Lynchburg College
Kayse Lee	Bethel University
Cathalee Soergel	Miami University

Advisor: Dr. Lakeshia Legette
Graduate Assistant: Brett Jefferson

Abstract

Have you ever wondered how MapQuest® works? It has the capability of finding the shortest route from one location to another. Suppose you want to visit n cities. In which order would you visit these cities to travel the shortest distance? We explore the Traveling Salesman Problem with both three and six cities on a two-dimensional plane. In both cases, we consider a varying Home point, **H**. It turns out that the plane is divided into distinct Home point regions such that for each region there is a unique minimal circuit to travel. We find that for three cities, some regions become disconnected, or *split*. This split suggests that it is not always optimal to travel the most intuitive circuit. We find all the city arrangements where this phenomenon occurs. Furthermore, we find that a particular arrangement of six cities guarantees only six distinct minimal circuits of the 720 possible circuits.

1. Introduction

The Traveling Salesman Problem (TSP) is a well-known combinatorial optimization problem requiring a salesman to start at a given city, called *Home*, travel to a number of other cities exactly once, and then return Home. The goal of the TSP is to minimize the distance of the total *circuit*, which is the cycle the salesman takes when he starts at Home, follows a particular route through all the other cities, and then returns Home.

There are many variations of the TSP. Some variations consider visiting each city more than once, minimizing cost or time, or using two or more modes of transportation. Furthermore, many variations consider the use of *asymmetrical distances*. That is, the distance from point **A** to point **B** is not necessarily the same as the distance from point **B** to point **A**. In this paper, we first explore the salesman traveling to three cities and then move to the six city case. In each, the Home point is a varying location.

Each city the salesman visits is represented by a fixed point in the Cartesian plane and is labeled with a distinct letter. The salesman's Home, which must be distinct from each of the cities, is denoted $\mathbf{H} = (x, y)$. Previous Summer Undergraduate Mathematical Sciences Research Institute (SUMSRI) groups considered solutions to the TSP when Home was moved about the Cartesian Plane. They considered both the three city and four city cases. For specific arrangements of these cities, they determined any Home points for which all circuits have equal distance. We extend their research by further investigating the regions of Home points with various solutions to the TSP. Specifically, we focus on studying how the regions change with different arrangements of the fixed cities.

We calculate the distance between two cities using Euclidean distance, denoted $d(City_1, City_2)$. If $City_1 = (x_1, y_1)$ and $City_2 = (x_2, y_2)$, then $d(City_1, City_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. We only consider *symmetrical* distances. That is, the distance from city **A** to city **B** is the same as the distance from city **B** to city **A**, written as $d(\mathbf{AB}) = d(\mathbf{BA})$.

We use a method of coloring the plane by which each circuit is represented with a unique color. At each location for the Home point, the shortest circuit that begins and ends at that point is determined. That point is then colored with the color that corresponds to the circuit with the shortest distance.

2. The Three City TSP

We begin by supposing the traveling salesman must visit three cities, with locations $\mathbf{A} = (a_1, a_2)$, $\mathbf{B} = (b_1, b_2)$, and $\mathbf{C} = (c_1, c_2)$. Since any three non-collinear points in the Cartesian plane form a triangle, we consider equilateral, isosceles, and scalene triangles formed by the three cities. For each arrangement of these cities, there are 3! possible circuits for the salesman to travel. One possible circuit is from Home to city **A** to city **B** to city **C** and back to Home. This circuit is denoted **HABCH**. The six possible circuits that the salesman can travel that start and end at a particular Home point are:

$$\begin{array}{cc} \mathbf{HABCH} & \mathbf{HCBAH} \\ \mathbf{HACBH} & \mathbf{HBCAH} \\ \mathbf{HBACH} & \mathbf{HCABH}. \end{array}$$

However, recall that we are assuming all distances are symmetric. This means that the circuits on the right are the same as the corresponding circuits on the left. We define the unique circuits on the left as Circuit 1, Circuit 2, and Circuit 3, and we assign the colors red, blue, and green to these circuits, respectively.

Circuit 1	**HABCH**	Red
Circuit 2	**HACBH**	Blue
Circuit 3	**HBACH**	Green

Table 2.1: Circuits for the Three City Case

Throughout this paper, we let d_1, d_2, and d_3 represent the distances of the corresponding circuits. Since $\mathbf{H} = (x, y)$ is the only varying point, d_1, d_2, and d_3 are functions of $\mathbf{H}$. Now we consider when the three cities are arranged in an equilateral triangle so that $d(\mathbf{AB}) = d(\mathbf{AC}) = d(\mathbf{BC})$.

2.1 The Equilateral Case

Since the distances between vertices are equal, we consider the equilateral arrangement as the base case for three cities. With the use of MATLAB, we find the circuits with minimal distances for each Home point and color the Home points according to Table 2.1. We observe that distinct colored regions of red, blue, and green appear. An example of the coloring we see in the equilateral case is shown in Figure 2.1.

Example 2.1. *Let the vertices be defined as* $\boldsymbol{A} = (3, 3)$*,* $\boldsymbol{B} = (7, 3)$ *and* $\boldsymbol{C} = (5, 2\sqrt{3} + 3)$*. After using MATLAB to color the Home points, we obtain the results in Figure 2.1.*

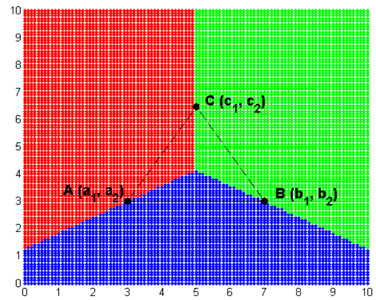

Figure 2.1: Three Cities Equilateral Case

Notice that $\mathbf{H} = (1, 8)$ lies in the red region; therefore, it is best for the salesman to travel along Circuit 1. That is, **HABCH**. The same is true for $\mathbf{H} = (3, 10)$. However, this does not imply that the distance of each circuit from the two Home points is the same. Now we explore colored regions.

Theorem 2.2. *For any self-intersecting circuit formed from non-cogeodesic vertices in the Cartesian plane, there exists a non self-intersecting circuit which has shorter distance.*

Proof. See [3].

□

Lemma 2.3. *Let* ***A****,* ***B****, and* ***C*** *be three fixed vertices in the Cartesian plane, and color the Home points red, blue, and green respective to the minimum circuits. Then each colored region is unbounded.*

Proof. For any line $\mathbf{L}$ through the vertex $\mathbf{B}$ and intersecting the edge $\overline{AC}$, we can form a ray $\mathbf{R}$. The beginning vertex of $\mathbf{R}$ is the point on $\overline{AC}$ and $\mathbf{R}$ continues along $\mathbf{L}$ in the opposite direction of $\mathbf{B}$.

For any Home point on $\mathbf{R}$, Circuit 2 and Circuit 3 are self-intersecting circuits. The reason is that both of these circuits require the salesman to travel from $\mathbf{H}$ to $\mathbf{B}$, which is a segment of $\mathbf{L}$. In addition, these circuits include the edge $\overline{AC}$. By definition of $\mathbf{L}$, these two segments must intersect. Therefore, Circuit 1 is the only possible non self-intersecting circuit.

By Theorem 2.2, we know that there exists a non self-intersecting circuit that has shorter distance than Circuits 2 and 3. Since Circuit 1 is the only other circuit and it is non self-intersecting, Circuit 1 must be the minimal circuit for any Home point on $\mathbf{R}$. Thus all Home points on $\mathbf{R}$ are colored red, making the red region unbounded. A similar argument follows for the other two circuits and their corresponding regions.

□

Now that we know unbounded colored regions always appear, we consider how these colored regions are divided in the equilateral case. The boundaries between the colors represent Home points for which multiple circuits are minimal. For example, the boundary between red and blue regions represents the Home points for which $d_1 = d_2 < d_3$. For each point along this boundary, the distance of Circuit 1 is the same as the distance of Circuit 2, and this distance is less than that of Circuit 3.

We can determine the equation of the boundary curve, $d_1 = d_2$, by noting that

$$d_1 = d(\mathbf{HA}) + d(\mathbf{AB}) + d(\mathbf{BC}) + d(\mathbf{CH})$$

and

$$d_2 = d(\mathbf{HA}) + d(\mathbf{AC}) + d(\mathbf{CB}) + d(\mathbf{BH}).$$

Therefore,

$$d_1 = d_2 \equiv d(\mathbf{HA}) + d(\mathbf{AB}) + d(\mathbf{BC}) + d(\mathbf{CH}) = d(\mathbf{HA}) + d(\mathbf{AC}) + d(\mathbf{CB}) + d(\mathbf{BH}).$$

Because this arrangement forms an equilateral triangle, we can simplify the equation to

$$\sqrt{(x-c_1)^2+(y-c_2)^2}=\sqrt{(x-b_1)^2+(y-b_2)^2},$$

for any arbitrary Home point, $\mathbf{H}=(x,y)$. After some algebraic manipulation, the following is the equation of the $d_1=d_2$ curve

$$-(c_1-b_1)x-(c_2-b_2)y=(b_1-c_1)a_1+(b_2-c_2)a_2.$$

We use similar reasoning to determine the curves

$$d_2=d_3:-(a_1-c_1)x-(a_2-c_2)y=(c_1-a_1)b_1+(c_2-a_2)b_2$$

and

$$d_1=d_3:-(a_1-b_1)x-(a_2-b_2)y=(b_1-a_1)c_1+(b_2-a_2)c_2.$$

Clearly all three of these equations are linear, as seen in Figure 2.2. Graphically, the line $d_1=d_2$ represents the division of the red and blue regions. However, this division does not appear in the green region, because d_3 is less than both d_1 and d_2. Another important observation is that there is a Home point for which all three circuits have the same distance. That is, $d_1=d_2=d_3$. This Home point has coordinates $\mathbf{H}=(x,y)$, where

$$x=\frac{c_1c_2(a_1-b_1)+b_1b_2(c_1-a_1)+a_1a_2(b_1-c_1)+c_2(b_2^2-a_2^2)+b_2(a_2^2-c_2^2)+a_2(c_2^2-b_2^2)}{c_1(b_2-a_2)+b_1(a_2-c_2)+a_1(c_2-b_2)}$$

and

$$y=\frac{c_1c_2(a_2-b_2)+b_1b_2(c_2-a_2)+a_1a_2(b_2-c_2)+c_1(b_1^2-a_1^2)+b_1(a_1^2-c_1^2)+a_1(c_1^2-b_1^2)}{-c_1(b_2-a_2)-b_1(a_2-c_2)-a_1(c_2-b_2)}.$$

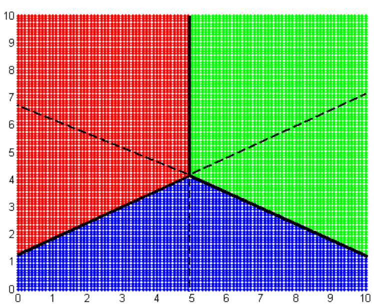

Figure 2.2: Boundary Lines for Equilateral Triangle

We now consider the TSP when the three cities are not arranged in an equilateral triangle.

2.2 The Isosceles/Scalene Case

We fix cities $\mathbf{A}=(0,0)$ and $\mathbf{B}=(b_1,0)$, where b_1 is a positive real number, along a horizontal line and move point $\mathbf{C}=(c_1,c_2)$ about the Cartesian plane to investigate solutions of the TSP when the three cities no longer form an equilateral triangle. Notice when $c_1=\frac{b_1}{2}$, an isosceles triangle is formed. For all other values of c_1, the triangle is scalene. Furthermore, when $c_2<0$, the triangle and its colored regions are a reflection over the x-axis. Therefore, we restrict $c_2>0$. As point $\mathbf{C}$ changes, we again expect to see colored regions, and we analyze the boundary curves.

Positions for City C

As we begin to investigate the scalene and isosceles arrangements of the three cities, recall that $\mathbf{A}=(0,0)$ and $\mathbf{B}=(b_1,0)$ are fixed positions. We consider three positions for c_1.

Scalene Case 1	$\frac{b_1}{2} < c_1 \leq b_1$
Scalene Case 2	$c_1 > b_1$
Isosceles Case	$c_1 = \frac{b_1}{2}$

Table 2.2: Possible Cases for City **C**

Figure 2.3 is a graphical representation of Table 2.2.

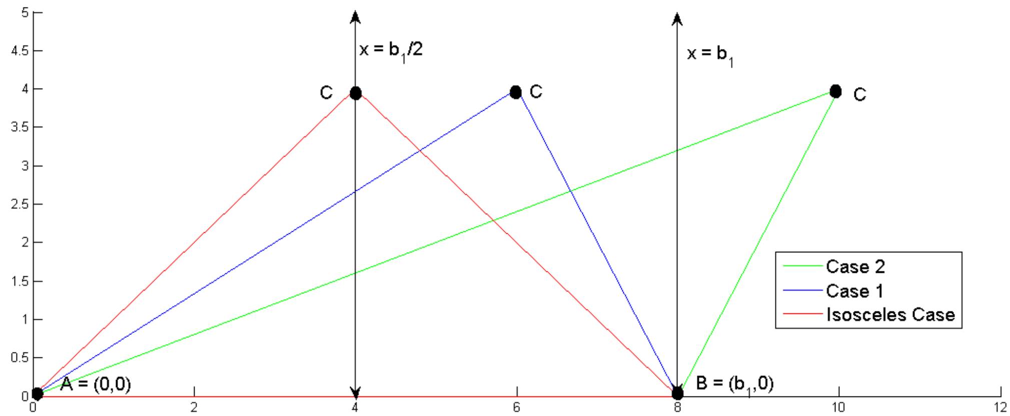

Figure 2.3: Graphical view of three cases for City **C**

With the line $x = \frac{b_1}{2}$ acting as an axis of symmetry, it is only necessary to consider the three cases listed in Table 2.2.

Hyperbola Properties

Before proceeding, we review the relevant properties of hyperbolas, as we use them in our analysis of boundary curves.

By definition, a hyperbola is a conic section defined as the locus of all points P in the plane, the difference of whose distances $d(F_1, P)$ and $d(F_2, P)$ from two fixed points (the foci F_1 and F_2) is equal to $\pm l$, where l is a constant real number. We let the center of our hyperbola be the fixed point (h, k) and let the horizontal and vertical distances from the center to one focal point be denoted as a and b, respectively. An example is illustrated in Figure 2.4.

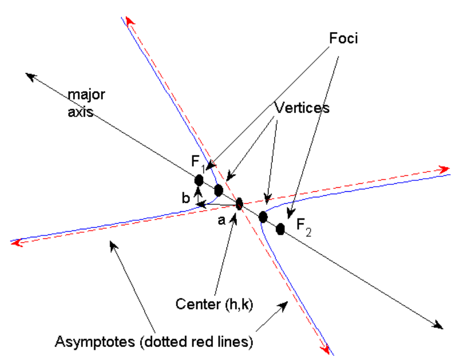

Figure 2.4: Properties of a Hyperbola

Finding the Hyperbolic Equation

We define coordinates for the foci as $F_1 = (f_1, g_1)$ and $F_2 = (f_2, g_2)$ and choose F_1 to be the focal point with the greater y-coordinate. If the foci of and one point on the hyperbola are known, we can find an explicit formula for this hyperbola. We start by finding the center (h, k) and values a and b by solving the system

$$f_1 = h + a,$$

$$f_2 = h - a,$$

$$g_1 = k + b,$$

and

$$g_2 = k - b.$$

By the definition of foci, the points $P = (x, y)$ on the hyperbola are given by the equation

$$\sqrt{(x-(h+a))^2 + (y-(k+b))^2} - \sqrt{(x-(h-a))^2 + (y-(k-b))^2} = \pm l.$$

Note, if we have a specific point $P = (x_0, y_0)$ on the hyperbola, we can solve for l in the equation above. In general, however, the above equation can be further simplified to:

$$-4a(x-h) - 4b(y-k) - l^2 = \pm 2l\sqrt{((x-h)+a)^2 + ((y-k)+b)^2}.$$

Finally,

$$(16a^2 - 4l^2)(x-h)^2 + (16b^2 - 4l^2)(y-k)^2 + 32ab(x-h)(y-k) = -l^4 + 4l^2a^2 + 4l^2b^2,$$

which is in the general form for a hyperbola.

Properties of the Curves

As before, we determine the boundary curves that appear to separate the colored regions. We define the equations

$$d_1 = d_2 \equiv d(\mathbf{HA})+d(\mathbf{AB})+d(\mathbf{BC})+d(\mathbf{HC}) = d(\mathbf{HA})+d(\mathbf{AC})+d(\mathbf{BC})+d(\mathbf{HB}) \equiv d(\mathbf{HC})+d(\mathbf{AB}) = d(\mathbf{AC})+d(\mathbf{HB}),$$

$$d_1 = d_3 \equiv d(\mathbf{HA}) + d(\mathbf{BC}) = d(\mathbf{AC}) + d(\mathbf{HB}),$$

and

$$d_2 = d_3 \equiv d(\mathbf{HA}) + d(\mathbf{BC}) = d(\mathbf{AB}) + d(\mathbf{HC}).$$

An example of these curves for a scalene triangle is graphed in Figure 2.5.

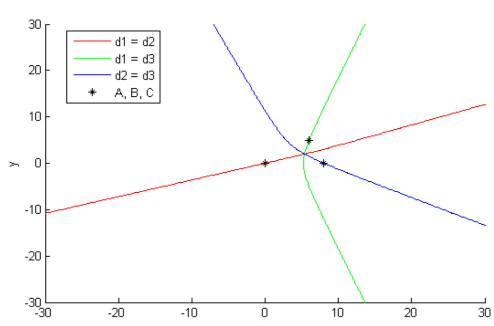

Figure 2.5: Boundary Curves for a Scalene Triangle

We begin by identifying properties of these curves. The following results are based on the research from SUMSRI 1999 and 2000.

Theorem 2.4. *Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be vertices in the Cartesian plane that form a scalene triangle. Let $\mathbf{H}$ be an arbitrary Home point. Each of the equations $d_1 = d_2$, $d_1 = d_3$, and $d_2 = d_3$ represents half of a hyperbola. In each case, two of the vertices of the triangle are foci of the hyperbola, and the third vertex is a point on the hyperbola.*

Proof. Rewrite the given equations as

$$d_1 = d_2 \equiv d(\mathbf{HC}) - d(\mathbf{HB}) = d(\mathbf{AC}) - d(\mathbf{AB}),$$

$$d_1 = d_3 \equiv d(\mathbf{HA}) - d(\mathbf{HB}) = d(\mathbf{AC}) - d(\mathbf{BC}),$$

and

$$d_2 = d_3 \equiv d(\mathbf{HA}) - d(\mathbf{HC}) = d(\mathbf{AB}) - d(\mathbf{BC}).$$

Consider the equation $d(\mathbf{HC}) - d(\mathbf{HB}) = d(\mathbf{AC}) - d(\mathbf{AB})$, which is equivalent to $d_1 = d_2$. For a given triangle, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are fixed. Consequently the distances, $d(\mathbf{AB})$, $d(\mathbf{AC})$, and $d(\mathbf{BC})$, are constants. Therefore, the right side of the equation, $d(\mathbf{AC}) - d(\mathbf{AB})$, is a constant value, which we define as l. Since $d(\mathbf{AC}) \neq d(\mathbf{AB})$, l is not zero. Hence this equation is the set of points $\mathbf{H}$ for which the difference of the distance from $\mathbf{H}$ to fixed point $\mathbf{C}$ and the distance from $\mathbf{H}$ to fixed point $\mathbf{B}$ is equal to l, which is either positive or negative. Therefore, the equation represents half of a hyperbola, where $\mathbf{C}$ and $\mathbf{B}$ are the foci.

Similar reasoning applies for the equation $d_1 = d_3$, which has foci $\mathbf{A}$ and $\mathbf{B}$ and for the equation $d_2 = d_3$, which has foci $\mathbf{A}$ and $\mathbf{C}$.

To see that $\mathbf{A}$ is a point on $d_1 = d_2$, recall that the constant value of l is $d(\mathbf{AC}) - d(\mathbf{AB})$. So for any point $\mathbf{P}$ to be on the hyperbola, the difference of the distances from $\mathbf{P}$ to each of the foci must be equal to l. So for point $\mathbf{P} = \mathbf{A}$, this difference is $d(\mathbf{AC}) - d(\mathbf{AB})$, which is equal to l. Therefore, $\mathbf{A}$ is a point on the hyperbola. Similarly, we can see that $\mathbf{C}$ is a point on $d_1 = d_3$ and $\mathbf{B}$ is a point on $d_2 = d_3$.

□

Corollary 2.5. *Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be vertices in the Cartesian plane that form an isosceles triangle, with $d(\mathbf{AC}) = d(\mathbf{BC})$, and let $\mathbf{H}$ be an arbitrary Home point. Then at least one boundary curve degenerates to a line. In the equilateral case, all three of the equations degenerate to those of lines.*

Proof. To show the former statement, recall that $d_1 = d_2$ and $d_2 = d_3$ are half hyperbolas. Consider the equation

$$d_1 = d_3 \equiv d(\mathbf{HA}) - d(\mathbf{HB}) = d(\mathbf{AC}) - d(\mathbf{BC}).$$

Since $d(\mathbf{AC}) = d(\mathbf{BC})$, the constant value l given by the right side of the equation is equal to 0. Therefore, $d(\mathbf{HA}) = d(\mathbf{HB})$ is linear.

To show the latter statement, note that for an equilateral triangle, all three sides have equal lengths. So all three equations have constants $l = 0$. Thus all the hyperbolas degenerate to lines.

□

Now that we have equations for the curves, we again look for a point where $d_1 = d_2 = d_3$. Note that if $d_2 = d_1$ and $d_1 = d_3$ intersect, then at that intersection point, $d_2 = d_3$ by transitivity. Thus, if two of the hyperbolas intersect, the third also intersects at that point.

Proposition 2.6. *Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be fixed vertices in the Cartesian plane, and let $\mathbf{H}$ be an arbitrary Home point. Then there exists at least one Home point $\mathbf{H}$ inside the triangle such that $d_1 = d_2 = d_3$.*

Proof. Consider $d_1 = d_2$ and $d_2 = d_3$. Since, by Theorem 2.4, $d_1 = d_2$ has foci $\mathbf{B}$ and $\mathbf{C}$, the hyperbola has a vertex on the line segment $\overline{BC}$. We also know this hyperbola passes through vertex $\mathbf{A}$. Similarly, since $d_2 = d_3$ has foci $\mathbf{A}$ and $\mathbf{C}$, this second hyperbola must have a vertex on the segment $\overline{AC}$ and through the vertex $\mathbf{B}$. A hyperbola connecting its vertex on $\overline{AC}$ to $\mathbf{B}$ must intersect a hyperbola connecting its vertex on $\overline{BC}$ to $\mathbf{A}$. Therefore, these two hyperbolas always intersect and by transitivity, the third hyperbola intersects at the same point. Furthermore, since each of the three hyperbolas has a vertex on a different side of the triangle formed by $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, the point $\mathbf{H}$ is located within the triangle, as proven in [2].

□

Now that we have studied the curves and where they intersect, we next investigate why the colored regions appear as they do.

Splitting Phenomenon

For some positions of **C**, we find that there are no longer three colored regions. Instead, one color appears in two disconnected regions. We call this phenomenon a *split*. There exist different locations for **C** for which each of the colored regions has a split. An example of each is given in Figure 2.6.

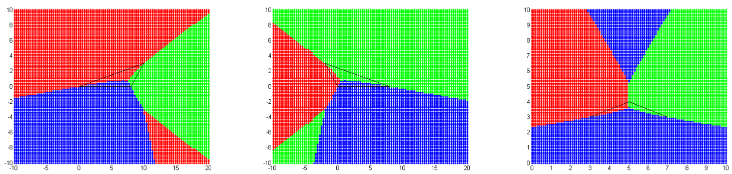

Figure 2.6: Red split(left), Green split(center), and Blue split(right)

Since the boundary curves are fundamental to the analysis of colored regions, we examine these curves in more detail. Based on our observations, we find that a second distinct point of intersection sometimes exists outside the triangle. Now we investigate this second intersection point to see if it relates to the splitting phenomenon.

Definition 2.7. Let **A**, **B**, and **C** be three fixed vertices in the Cartesian plane, and color the Home points red, blue, and green respective to the minimum circuits. Assume that there is a second point of intersection, $d_1 = d_2 = d_3$, for the boundary curves. The *steady center curve* is the curve that maintains its center position beyond the second point of intersection.

Theorem 2.8. *Let* ***A****,* ***B****, and* ***C*** *be fixed vertices in the Cartesian plane that form a scalene or isosceles triangle, and let* ***H*** *be an arbitrary Home point. If there exist two distinct points of intersection among all three boundary curves, then there is a split in one of the colored regions.*

Proof. Recall that by Proposition 2.6, the hyperbolas intersect at a point within the triangle. Suppose a second point of intersection exists. Follow the three curves from the first intersection point toward the second. Since the hyperbolas intersected inside the triangle, they cannot lie tangent to each other at the second intersection point. Thus, a steady center curve exists.

Suppose we have any intersection point of the three curves, $d_1 = d_2$, $d_1 = d_3$, and $d_2 = d_3$. Without loss of generality, we can assume that above and below the intersection point, $d_1 = d_3$ is located between the other two curves. In addition, we can assume that $d_1 < d_2$ to the left of the curve $d_1 = d_2$. Under these assumptions, there are only two possible arrangements for the curves. One possible arrangement of the curves at this intersection point is shown in Figure 2.7.

Note that at any point along the boundary curve $d_1 = d_2$, two options are possible:
1) $d_1 = d_2 < d_3$ meaning that the boundary curve divides the red and blue regions at this point
2) $d_1 = d_2 > d_3$ meaning that the point is colored green and either side of the boundary curve is colored green.
This is true for any boundary curve. We use this to determine the colored regions which appear in Figure 2.7.

Based on the assumed inequality stated above, we know that d_1 cannot be minimum to the right of $d_1 = d_2$. Therefore, neither of the two regions below the intersection and between $d_1 = d_2$ and $d_2 = d_3$ can be colored red. Since these regions are immediately left and right of $d_1 = d_3$, they must both be colored blue. Using the same reasoning, the regions to the left of $d_1 = d_2$ and $d_1 = d_3$ are colored red. In addition, the regions to the right of $d_2 = d_3$ and $d_1 = d_3$ are colored green. This is illustrated in Figure 2.7.

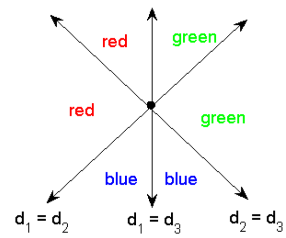

Figure 2.7: Illustration of an Intersection Point Coloring

If we have two points of intersection, then one of the intersection points has colored regions as pictured in Figure 2.7. The other has colored regions in Figure 2.8, which is a reflection of Figure 2.7 and the only other possible arrangement of the curves with $d_1 = d_3$ in the center.

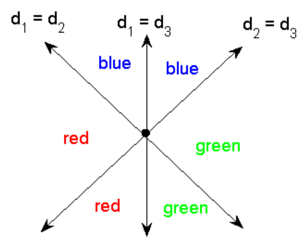

Figure 2.8: Illustration of Reflected Intersection Point Coloring

Therefore, these colorings from Figures 2.7 and 2.8 can be arranged in the two ways shown in Figure 2.9.

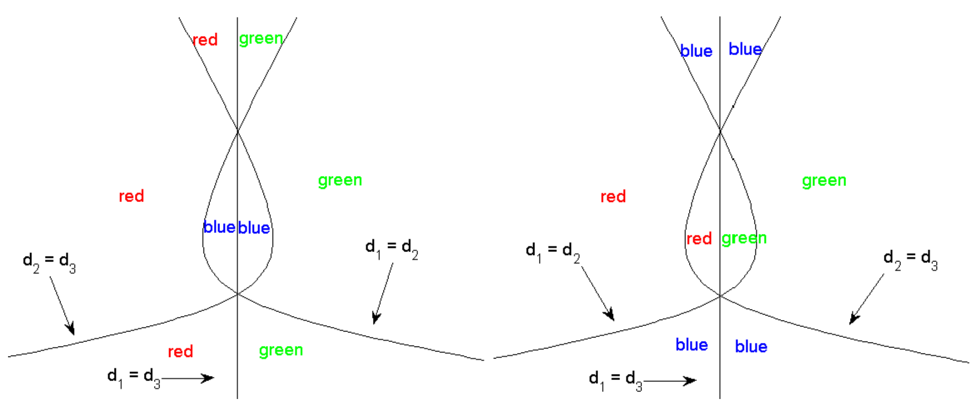

Figure 2.9: Two options for colorings if two intersection points exist

Since the figure on the left has a bounded blue region, this arrangement of intersection points is not possible by Lemma 2.3. Thus, two points of intersection must be arranged to give the coloring shown on the right, with the steady center curve $d_1 = d_3$, and leading to a blue split. The same is true for any arrangement of the three curves and their associated colorings.

□

Corollary 2.9. *Let $i, j, k \in \{1, 2, 3\}$ such that $i \neq j \neq k$. If $d_j = d_k$ is a steady center curve, then there will be a split in the colored region corresponding to Circuit i.*

Proof. Follows immediately from Theorem 2.8.

□

Now that we know a second intersection causes a split, we study where a second intersection can occur.

Theorem 2.10. *Let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ be fixed vertices in the Cartesian plane that form a scalene or isosceles triangle. Also let $\boldsymbol{H}$ be an arbitrary Home point. Let Region 1 be defined as the region below $\overleftrightarrow{AB}$ and to the right of $\overleftrightarrow{BC}$. Let Region 2 be defined as the region to the left of $\overleftrightarrow{AC}$ and to the right of $\overleftrightarrow{BC}$. Let Region 3 be defined as the region below $\overleftrightarrow{AB}$ and to the left of $\overleftrightarrow{AC}$. A second point of intersection among the three curves can only occur in one of the three regions listed above. Furthermore, an intersection in Regions 1, 2, or 3 will lead to a red, blue, or green split, respectively.*

Proof. Lemma 2.3 defines a subspace of $\mathbb{R}_2$ for which only one circuit is non self-intersecting. Outside of the triangle, the complement of this subspace is made up of Regions 1, 2, and 3. Therefore, these regions contain all of the Home points outside of the triangle for which it is possible to have multiple non self-intersecting circuits, and thus where the minimal circuit may not be unique. So a second intersection point can only occur in Regions 1, 2, or 3.

By Theorem 2.4, we know that the curve $d_2 = d_3$ goes through point **B** and that the curves $d_1 = d_3$ and $d_1 = d_2$ have vertices on $\overline{AB}$ and $\overline{BC}$, respectively. Thus, in Region 1, $d_2 = d_3$ is in between the other two curves. If a second intersection point exists in this region, then $d_2 = d_3$ will be the steady center curve, causing a red split by Corollary 2.9. Similarly, an intersection in Region 2 will lead to a blue split, and an intersection in Region 3 will lead to a green split.

□

Previous SUMSRI research groups developed and proved a theorem in [1] which gives the positions of **C** in an isosceles triangle for which there is a blue split.

Theorem 2.11 (Bellino and Narasimhan)**.** *Let $p \in \mathbb{R}$ and $p > 0$. Let $\boldsymbol{A} = (0,0)$, $\boldsymbol{B} = (8p,0)$, and $\boldsymbol{C} = (4p, c_2)$ be the three vertices forming an isosceles triangle. If $c_2 < 3p$, a blue split will occur.*

Proof. Assume $c_2 = 3p - \epsilon$ with $0 < \epsilon < 3p$. Let the x-coordinate of a Home point **H** be $x = 4p$.
Since we know that the first intersection occurs inside the triangle, we find a point above **C** that is colored blue, then a blue split occurs. We want to show that there exists a y-coordinate, $y > 3p$ for **H** such that $d_2 < d_3$ and $d_2 < d_1$.

$d_2 < d_1 \Leftrightarrow d(\mathbf{HB}) + d(\mathbf{BC}) + d(\mathbf{CA}) + d(\mathbf{AH}) < d(\mathbf{HC}) + d(\mathbf{CB}) + d(\mathbf{BA}) + d(\mathbf{AH}) \Leftrightarrow d(\mathbf{HB}) + d(\mathbf{CA}) < d(\mathbf{HC}) + d(\mathbf{BA})$.

By applying the distance formula to the line segments, we have

$$d(\mathbf{HB}) = \sqrt{(4p - 8p)^2 + (y - 0)^2} = \sqrt{16p^2 + y^2}$$

and

$$d(\mathbf{AC}) = \sqrt{(4p - 0)^2 + ((3p - \epsilon) - 0)^2} = \sqrt{25p^2 - 6p\epsilon + \epsilon^2}.$$

Therefore, $d(\mathbf{HB}) + d(\mathbf{AC}) = \sqrt{16p^2 + y^2} + \sqrt{25p^2 - 6p\epsilon + \epsilon^2}$.

Similarly, we find $d(\mathbf{HC}) + d(\mathbf{BA}) = y - 3p + \epsilon + 8p = y + 5p + \epsilon$.

In order to prove $d(\mathbf{HB}) + d(\mathbf{CA}) < d(\mathbf{HC}) + d(\mathbf{BA})$, first compare $\sqrt{25p^2 - 6p\epsilon + \epsilon^2}$ to $5p$. We want to show

$$\sqrt{25p^2 - 6p\epsilon + \epsilon^2} < 5p.$$

Note that $0 < \epsilon < 3p$. Thus, $\epsilon < 3p < 6p$, implying $\epsilon^2 - 6p\epsilon < 0$. Therefore,

$$\sqrt{25p^2 + (-6p\epsilon + \epsilon^2)} < \sqrt{25p^2} = 5p.$$

Next, we compare $d(\mathbf{HB}) = \sqrt{16p^2 + y^2}$ to $(y + \epsilon)$. Note that the following inequalities are equivalent:

$$\sqrt{16p^2 + y^2} < (y + \epsilon).$$

$$y\sqrt{1+\frac{16p^2}{y^2}} < y(1+\frac{\epsilon}{y})$$

$$\sqrt{1+\frac{16p^2}{y^2}} < (1+\frac{\epsilon}{y})$$

$$1+\frac{16p^2}{y^2} < (1+\frac{2\epsilon}{y}+\frac{\epsilon^2}{y^2})$$

$$\frac{16p^2}{y} < 2\epsilon$$

Since ϵ is a fixed value, y can be made large enough so that this inequality is true. We have thus proven that $\sqrt{16p^2+y^2}+\sqrt{25p^2-6p\epsilon+\epsilon^2} < y+5p+\epsilon$, for y sufficiently large, and

$$d(\mathbf{HB})+d(\mathbf{BC})+d(\mathbf{CA})+d(\mathbf{AH}) < d(\mathbf{HC})+d(\mathbf{CB})+d(\mathbf{BA})+d(\mathbf{AH}),$$

meaning $d_2 < d_1$.

Similarly, for $d_2 < d_3$, we need to show that

$$d(\mathbf{HB})+d(\mathbf{BC})+d(\mathbf{CA})+d(\mathbf{AH}) < d(\mathbf{HB})+d(\mathbf{BA})+d(\mathbf{AC})+d(\mathbf{CH}),$$

which can be simplified to

$$d(\mathbf{BC})+d(\mathbf{AH}) < d(\mathbf{BA})+d(\mathbf{CH}).$$

The same reasoning can be used to show $d_2 < d_3$.

□

We further this result by considering any arrangement of the three cities. We investigate additional points **C** for which a blue split occurs, and then we consider points **C** for which a red or green split occurs. We can locate the points **C** where a split occurs by finding when there are two distinct intersection points.

Theorem 2.12 (Split Theorem)**.** *Let $\boldsymbol{A}=(0,0)$, $\boldsymbol{B}=(b_1,0)$, and $\boldsymbol{C}=(c_1,c_2)$ be fixed vertices in the Cartesian plane that form a scalene or isosceles triangle. Also let $\boldsymbol{H}=(x,y)$ be an arbitrary Home point. For a given value of b_1, then for $\boldsymbol{C}$ above the x-axis, a blue split occurs if and only if $\boldsymbol{C}$ is also below the curve defined by*

$$\frac{c_2(-\sqrt{c_1^2+c_2^2}+b_1)+\sqrt{2b_1}\sqrt{-c_1+(c_1^2+c_2^2)^{1/2}}(b_1-c_1)}{(b_1-c_1)(\sqrt{c_1^2+c_2^2}-b_1)+c_2\sqrt{2b_1}\sqrt{-c_1+(c_1^2+c_2^2)^{1/2}}} =$$

$$\frac{c_2(-\sqrt{(b_1-c_1)^2+c_2^2}+b_1)+c_1\sqrt{2b_1}\sqrt{c_1-b_1+((b_1-c_1)^2+c_2^2)^{1/2}}}{-c_1(\sqrt{(b_1-c_1)^2+c_2^2}-b_1)-c_2\sqrt{2b_1}\sqrt{c_1-b_1+((b_1-c_1)^2+c_2^2)^{1/2}}}.$$

For $\boldsymbol{C}$ above the x-axis, a red split occurs if and only if $\boldsymbol{C}$ is also to the right of the curve defined by

$$\frac{\sqrt{2c_1b_1-2c_2^2-2c_1^2+2(c_1^2(c_1-b_1)^2+c_2^2(c_1-b_1)^2+c_1^2c_2^2+c_2^4)^{1/2}}}{\sqrt{(c_1-b_1)^2+c_2^2}-\sqrt{c_1^2+c_2^2}} =$$

$$\frac{c_2(-\sqrt{(b_1-c_1)^2+c_2^2}+b_1)+c_1\sqrt{2b_1}\sqrt{c_1-b_1+((b_1-c_1)^2+c_2^2)^{1/2}}}{-c_1(\sqrt{(b_1-c_1)^2+c_2^2}-b_1)-c_2\sqrt{2b_1}\sqrt{c_1-b_1+((b_1-c_1)^2+c_2^2)^{1/2}}}.$$

For $\boldsymbol{C}$ above the x-axis, a green split occurs if and only if $\boldsymbol{C}$ is also to the left of the curve defined by

$$\frac{\sqrt{2(b_1-c_1)b_1-2c_2^2-2(b_1-c_1)^2+2(c_1^2(b_1-c_1)^2+c_2^2c_1^2+c_2^2(b_1-c_1)^2+c_2^4)^{1/2}}}{\sqrt{c_1^2+c_2^2}-\sqrt{(b_1-c_1)^2+c_2^2}} =$$

$$\frac{c_2(-\sqrt{(b_1-c_1)^2+c_2^2}+b_1)+c_1\sqrt{2b_1}\sqrt{c_1-b_1+((b_1-c_1)^2+c_2^2)^{1/2}}}{-c_1(-\sqrt{(b_1-c_1)^2+c_2^2}+b_1)+c_2\sqrt{2b_1}\sqrt{c_1-b_1+((b_1-c_1)^2+c_2^2)^{1/2}}}.$$

Proof. Recall the regions defined in Theorem 2.10. We know that if a second point of intersection exists in Region 2, a blue split occurs. We consider the hyperbolas $d_1 = d_2$ and $d_2 = d_3$ to find locations of $\mathbf{C}$ for which a second intersection point occurs in Region 2. Finding this intersection point can be done by considering the asymptotes of $d_1 = d_2$ and $d_2 = d_3$ and looking for their intersection points.

We split the hyperbola $d_1 = d_2$ into two portions, each lying on one side of the vertex. Let the portion containing $\mathbf{A}$ be denoted as $\mathcal{P}_1$ and the other portion be denoted as $\mathcal{P}_2$. We do the same for $d_2 = d_3$, where the portion $\mathcal{Q}_1$ contains $\mathbf{B}$, and $\mathcal{Q}_2$ lies on the other side of the vertex. If a second point of intersection exists in Region 2, $\mathcal{P}_2$ and $\mathcal{Q}_2$ must intersect. This is illustrated in Figure 2.10.

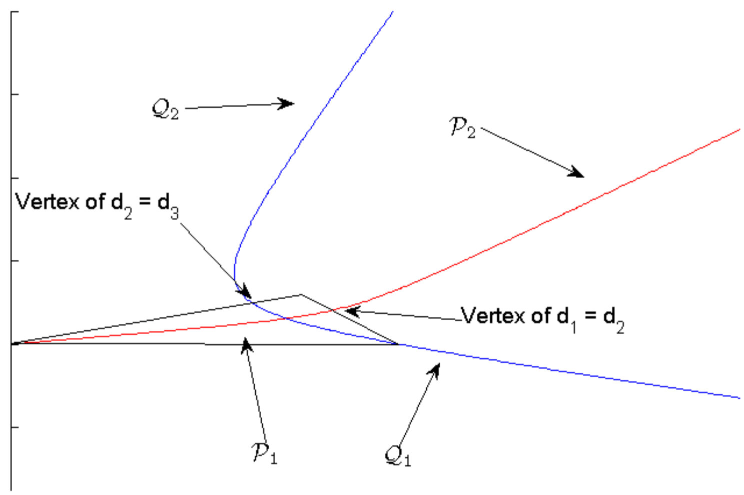

Figure 2.10: Location of second intersection point in Region 2

We use the Box Method, which is a standard way of finding asymptotes for a hyperbola. Since we are only interested in the asymptotes which $\mathcal{P}_2$ and $\mathcal{Q}_2$ approach, we draw these asymptotes using the appropriate corners of the box. For each of the hyperbolas, we solve for a, b, l, h, and k using the method described in Section 2.2. These asymptotes have the equation

$$y - k = \frac{-bl/2 + |a|\sqrt{c^2 - l^2/4}}{-al/2 - |a|b\sqrt{c^2 - l^2/4}/a}(x - h),$$

where $c^2 = a^2 + b^2$.

Recall that b_1 is fixed. Consequently, we find the slope of the $d_1 = d_2$ asymptote as a function of c_1 and c_2 to be

$$\frac{c_2(-\sqrt{c_1^2 + c_2^2} + b_1) + \sqrt{2b_1}\sqrt{-c_1 + (c_1^2 + c_2^2)^{1/2}}(b_1 - c_1)}{(b_1 - c_1)(\sqrt{c_1^2 + c_2^2} - b_1) + c_2\sqrt{2b_1}\sqrt{-c_1 + (c_1^2 + c_2^2)^{1/2}}}.$$

Similarly, we find the slope of the $d_2 = d_3$ asymptote to be

$$\frac{c_2(-\sqrt{(b_1 - c_1)^2 + c_2^2} + b_1) + c_1\sqrt{2b_1}\sqrt{c_1 - b_1 + ((b_1 - c_1)^2 + c_2^2)^{1/2}}}{-c_1(\sqrt{(b_1 - c_1)^2 + c_2^2} - b_1) - c_2\sqrt{2b_1}\sqrt{c_1 - b_1 + ((b_1 - c_1)^2 + c_2^2)^{1/2}}}.$$

Now that we know the slopes of the asymptotes, we can determine where $\mathcal{P}_2$ and $\mathcal{Q}_2$ intersect. We do this by considering the portions of the asymptotes which $\mathcal{P}_2$ and $\mathcal{Q}_2$ approach. If these portions of the asymptotes intersect, so will the corresponding portions of the hyperbolas. Hence this method will give us locations of $\mathbf{C}$ which cause a blue split.

Now we define a ray R_P to be the portion of the asymptote for $d_1 = d_2$ which begins at the center of the hyperbola (h, k) and runs along $\mathcal{P}_2$. Thus, R_P has the same slope as the asymptote for $d_1 = d_2$, but begins at the center and points in the direction of $\mathcal{P}_2$. We define R_Q to be the ray which represents the portion of the asymptote for $d_2 = d_3$ which begins at center of this hyperbola and points in the direction of $\mathcal{Q}_2$.

From the derived h and k for each hyperbola found above, we know the center of the $d_1 = d_2$ hyperbola is $(\frac{c_1+b_1}{2}, \frac{c_2}{2})$. Likewise, the center of $d_2 = d_3$ is $(\frac{c_1}{2}, \frac{c_2}{2})$. Since l is defined as the difference $d(\mathbf{H}F_1) - d(\mathbf{H}F_2)$, then for $d_1 = d_2$, a positive $l = \sqrt{c_1^2 + c_2^2} - b_1$ indicates that the half hyperbola is concave around focus **B**. Similarly, for $d_2 = d_3$, if $l = \sqrt{(b_1 - c_1)^2 + c_2^2} - b_1$ is positive, then the half hyperbola is concave around focus **A**. For both of these hyperbolas, if l is negative then the half hyperbola is concave around focus **C**. For each **C**, we can calculate the slope of the rays. In addition, we find the direction of the ray from the center based on the sign of l, by considering which focus the hyperbola is concave around.

Fix c_1 between 0 and b_1. Now we consider the rays R_P and R_Q as c_2 moves from 0 to infinity. First consider R_P as c_2 approaches 0. Let m be the slope of R_P, given above. We find that $\lim_{c_2 \to 0} m = 0$. Similarly, for $d_1 = d_2$, $\lim_{c_2 \to 0} l = c_1 - b_1$, which is negative. So as c_2 approaches 0, R_P points from its center $(\frac{c_1+b_1}{2}, \frac{c_2}{2})$ horizontally in the negative x direction. We do the same for $d_2 = d_3$ and find that as $c_2 \to 0$, R_Q points from its center $(\frac{c_1}{2}, \frac{c_2}{2})$ horizontally in the positive x direction. As c_2 approaches infinity, we find R_P and R_Q to point vertically from their respective centers in the negative y direction, as shown in Figure 2.11.

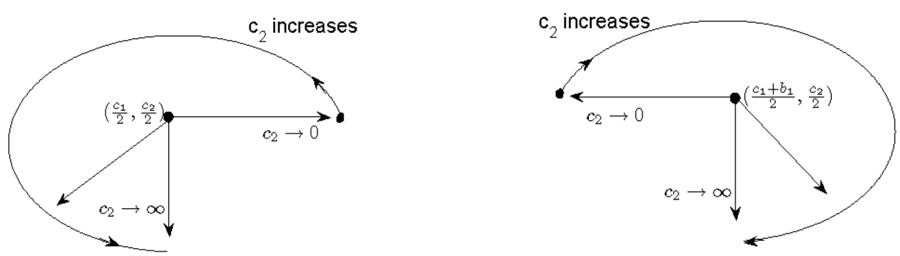

Figure 2.11: R_Q on left and R_P on right

Further, we find the direction of the rays for some c_2 between 0 and infinity. Since c_1 and b_1 are arbitrary, we can use $b_1 = 8$, $c_1 = 6$, and $c_2 = 16$. We find that R_P has slope $-.5135$ and $l = 9.09$, causing the ray to point from the center in the positive x direction and negative y direction. In addition, R_Q has slope .138 and $l = 8.12$, causing the ray to point from the center in both the negative x and y directions. Since the slopes of the asymptotes and equations for l are all continuous with respect to c_2, the rays rotate continuously about their centers. Based on the equations for slopes, we know that the limits of the rays are the bounds of c_2. In addition, we know the directions of the rays for some c_2 between the bounds. Thus, we know that the rays must change as shown in Figure 2.11 as c_2 increases.

The rate at which the rays approach their limit depends on c_1; however, since the rays must change continuously as c_2 increases from 0, and since they cannot pass the limit as c_2 approaches ∞, we can examine where the rays intersect. Because the center of R_Q is to the left of the center of R_P, the rays will intersect for any $c_2 > 0$ until some value of c_2 for which the rays are parallel. This is the first defined value of c_2 for which the slopes of the asymptotes are equal, thus it is the lower curve shown in Figure 2.12. After this point, the rays no longer intersect, since R_Q has moved to the right of R_P. The slopes will again be equal at some greater c_2 value, shown by the above curve in Figure 2.12. However, along this curve, the rays point in opposite directions. Therefore, for **C** below the lower curve and above the x-axis, a second point of intersection occurs in Region 2, causing a blue split. Based on the analysis of R_P and R_Q for **C** above this curve, a blue split cannot occur elsewhere for Scalene Case 1.

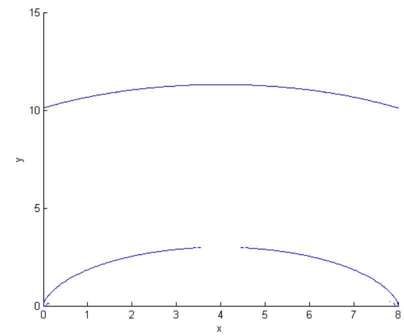

Figure 2.12: Example of equation where slopes of asymptotes are equal, given in theorem, for $\mathrm{b}_1 = 8$

Note that in the isosceles case, the rays change at the same rate, since $c_1 = \frac{b_1}{2}$. In this case, the rays become parallel, and no longer intersect when they both are vertical. Here, their slopes are undefined, which explains the gap in the lower curve of Figure 2.12. It is still below this open point, however, that a blue split occurs, confirming Theorem 2.11.

Furthermore, if we fix c_2 and consider the rays as c_1 increases from b_1 to infinity, illustrating Scalene Case 2, we find that the rays can never intersect in Region 2. Thus, for any **C** above the x-axis, a blue split occurs if and only if **C** is also below the lower curve in Figure 2.12. Note that based on these results, we know that a second intersection cannot occur within the triangle, and so splits only occur for second intersections in Regions 1, 2, or 3.

In a similar manner, we find the equations given in the theorem for the locations of **C** which have an intersection in Regions 1 or 3, causing red or green splits respectively. Figure 2.13 illustrates the locations of **C** for each of these colored splits.

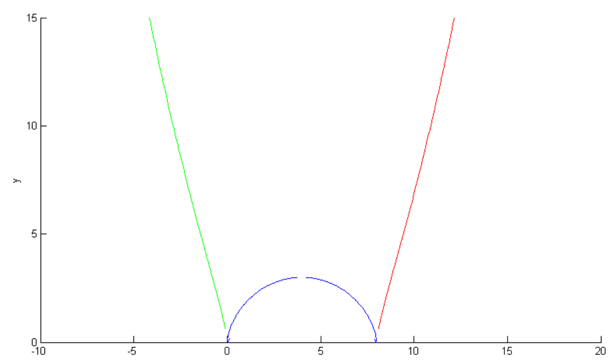

Figure 2.13: Red splits occur right of red curve, green splits occur left of green curve, and blue splits occur below blue curve

$\square$

At first glance, the simple synopsis of the TSP might underestimate the complexity of its solutions. One tends to assume the salesman can travel the shortest distance by traveling to the closest city each time. However, with the result of the splits, we find this intuitive idea is not always accurate. Within the realm of our splitting phenomenon, we have found arrangements of three cities for which the salesman does not travel to the closest city first, but still minimizes his distance.

For example, consider a blue split. Imagine the salesman's home is a point in the upper blue region. Since the Home point is colored blue, the optimal circuit is **HACBH**. Clearly the distance from **H** to **C** is shorter than **H** to **A**, but the salesman starts his tour by traveling to **A**. After visiting **C**, the closest city to Home, he ends with city **B**, which is not the closest to his home.

The results of the splits not only provide us with a way to analyze colored regions, but they also provide us with the important realization that traveling to the closest city from any location does not always result in the shortest circuit.

After the satisfying results of the Split Theorem, we move on to the $n = 6$ case.

3. The Six City TSP

We extend our knowledge of the $n = 3$ case to a situation where the traveling salesman visits six cities exactly once, starting and ending at his Home. Let **A**, **B**, **C**, **D**, **E**, and **F** be non-collinear fixed vertices in the Cartesian plane. Let **H** be an arbitrary Home point. At first glance, it appears that there are $6! = 720$ possible circuits to examine. However, as in the three city case, we use symmetric distances to reduce the number of circuits to $\frac{6!}{2} = 360$.

To further reduce the number of circuits to consider, we recall Theorem 2.2. It states that for any circuit that is self-intersecting, there exists a non self-intersecting circuit which is shorter. Thus we will not consider self-intersecting circuits when looking for the minimum distance circuit.

3.1 Regular Hexagon

Just as the equilateral triangle is the simplest arrangement of vertices in the $n = 3$ case, the regular hexagon provides us with a basic foundation for six cities. The Home point **H** can be located either inside or outside the regular hexagon. We consider **H** to be inside the regular hexagon if **H** lies on a line segment. When analyzing the circuits for cities arranged in a hexagonal manner, the *beginning vertex* is the first vertex the traveling salesman visits after leaving Home. Likewise, the *ending vertex* is the last vertex visited by the traveling salesman before returning Home. A *walk* refers to the portion of the circuit excluding the Home point.

First we analyze the solutions for the TSP when the Home point **H** is located inside the regular hexagon.

Theorem 3.1. *Let **A**, **B**, **C**, **D**, **E**, and **F** be fixed vertices in the Cartesian plane which form a regular hexagon, and let **H** be an arbitrary Home point located inside the hexagon. There are exactly six non self-intersecting circuits that begin and end with **H**. Furthermore, all of these circuits have beginning and ending vertices which are adjacent and contain the walk where, from each vertex, the salesman travels to the next adjacent vertex.*

Proof. Suppose toward a contradiction that there is a non self-intersecting circuit that has beginning and ending vertices which are not adjacent. These vertices can have one vertex between them or have two vertices between them. Assume they have one vertex between them. Without loss of generality, let **F** and **D** be two non-adjacent beginning and ending vertices on the regular hexagon. Since **H** is inside the hexagon, $\overline{\mathbf{FH}}$ and $\overline{\mathbf{HD}}$ are two connected edges. The joined edges divide the hexagon into two distinct regions.

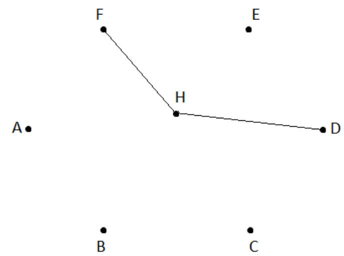

Figure 3.1: Beginning and Ending Vertices

Since **F** and **D** are non-adjacent, there exists at least one vertex contained in each region that is not one of the beginning or ending vertices. That is, at least one vertex is on either side of the joined edges $\overline{\mathbf{FH}}$ and $\overline{\mathbf{HD}}$.

We know vertex **E** is the only vertex in one of the regions of the divided hexagon. Since the vertices **A**, **B**, and **C** are in the other region, we cannot travel from one of those vertices to **E** without intersecting $\overline{\mathbf{HD}}$ or $\overline{\mathbf{FH}}$. This is because every regular hexagon is convex, meaning every line segment between two vertices remains inside or on the boundary of the polygon. Therefore, a non self-intersecting circuit with beginning and ending vertices containing one vertex between them does not exist.

If the non-adjacent vertices have two vertices between them, the hexagon is still divided into two distinct regions. As above, an edge cannot connect a vertex in one region to a vertex in the other region without intersecting the edges that divide the two regions. Therefore, a non self-intersecting circuit must have adjacent beginning and ending vertices. There are six distinct pairs of beginning and ending vertices which are adjacent. They are

A and **B**,
B and **C**,
C and **D**,
D and **E**,
E and **F**,

and

F and **A**.

We know that for **H** inside the hexagon, the six pairs of beginning and ending vertices above are the only possibilities for a non self-intersecting circuit. Further, we need to show that for each pair above, there exists only one non self-intersecting circuit with this pair of beginning and ending vertices.

Consider a circuit with beginning and ending vertices which are adjacent. Within the walk between these vertices, if the salesman travels from some vertex to a non-adjacent vertex, then the walk divides the hexagon into two regions on either side of the segment connecting the two non-adjacent vertices. Stated earlier in the proof, this must result in a self-intersecting circuit. It follows that the salesman must travel the regular hexagon via adjacent vertices, in order to have a non self-intersecting circuit.

□

Theorem 3.2. *Let **A**, **B**, **C**, **D**, **E**, and **F** be fixed vertices in the Cartesian plane which form a regular hexagon. For each of the six possible non self-intersecting circuits there exists a Home point **H** inside the regular hexagon such that that circuit is minimal.*

Proof. The six non self-intersecting circuits have adjacent beginning and ending vertices, namely, **A** and **B**, **B** and **C**, **C** and **D**, **D** and **E**, **E** and **F**, or **F** and **A**. For each of these pairs of beginning and ending vertices, there is one circuit which is non self-intersecting. We want to show that each of these six circuits is minimal for at least one Home point **H** inside the hexagon.

Consider **A** and **B** to be the pair of adjacent beginning and ending vertices. Now choose **H** to be the midpoint of the edge $\overline{\mathbf{AB}}$. Since our vertices are arranged in a regular hexagon, $d(\mathbf{HA}) = d(\mathbf{BH})$ is the minimum distance between **H** and each of the vertices.

We know from Theorem 3.1 that the only non self-intersecting circuit with beginning vertex **A** and ending vertex **B**, contains the walk between **A** and **B** where, from each vertex, the salesman travels to the next adjacent vertex around the hexagon. This is the circuit **HAFEDCBH**.

To prove this circuit is the shortest, we want to show that the distance of **HAFEDCBH** is less than or equal to the distance of the five other possible non self-intersecting circuits.

The five other circuits are

HABCDEFH,
HBAFEDCH,
HDEFABCH,
HEDCBAFH,

and

HDCBAFEH.

Since each circuit is traveled via adjacent vertices, we know the distance traveled along the walk between vertices is equal in all six circuits. It is, therefore, equivalent to prove $d(\mathbf{HA}) + d(\mathbf{BH})$ is less than or equal to $d(\mathbf{HA}) + d(\mathbf{FH})$, $d(\mathbf{HC}) + d(\mathbf{BH})$, $d(\mathbf{HC}) + d(\mathbf{DH})$, $d(\mathbf{HD}) + d(\mathbf{EH})$, and $d(\mathbf{HE}) + d(\mathbf{FH})$. Recall that we originally chose **H** such that **A** and **B** have minimal distance to **H**. Therefore, the corresponding circuit also has minimal distance. Since adjacent vertices **A** and **B** were arbitrary, using the method described above gives **H** so that the circuit is minimal.

□

Since we know the result of a Home point located inside the regular hexagon, we proceed to discuss Home points lying outside the hexagon. In order to do this, we recall the idea of a walk and identify the types of possible walks for the regular hexagon:

(1)-The beginning and ending vertices are adjacent.
(2)-The beginning and ending vertices have one vertex between them.
(3)-The beginning and ending vertices have two vertices between them.

Note that walks with beginning and ending vertices with three vertices between them are the same as type (2), and those with four vertices between them are the same as type (1). It is also important to remember that intersecting walks are not being considered.

To create an accurate representation of distances, we label the lengths in Figure 3.2 in increasing order, s, m, z. The length s connects two adjacent vertices; m connects vertices with one vertex between them, and z connects vertices with two vertices between them.

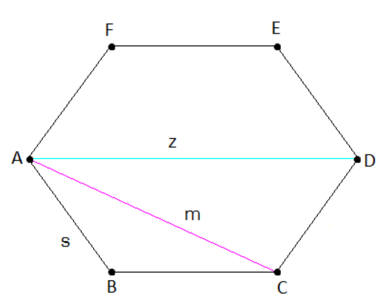

Figure 3.2: Lengths of segments in a regular hexagon

Consider the geometry of a regular hexagon in Figure 3.3.

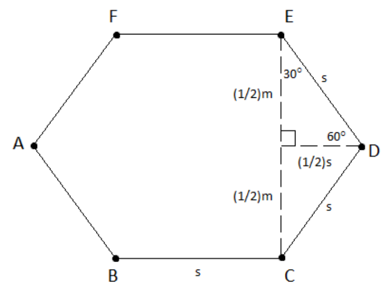

Figure 3.3: Geometric Structure

In order to compare walks of type (2) and (3) to those of type (1), we write the lengths m and z, in terms of s.

Using the basic trigonometry shown in Figure 3.3, we know $m = \sqrt{3}s$ and $z = 2s$. Given these lengths, we can find the shortest walk among each type. The distance from the beginning and ending vertices to the Home point **H** is a fixed distance for all types, so we focus on the distance traveled during a given walk inside the hexagon.

Recall that among type (1) walks, there is only one possible way to complete such a walk, as stated in Theorem 3.1. Arbitrarily choosing beginning and ending vertices, we illustrate this walk in Figure 3.4.

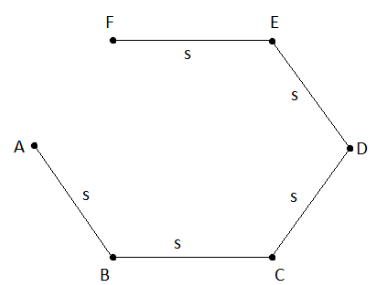

Figure 3.4: Adjacent Walk

Proposition 3.3. *Let* $\boldsymbol{A}$*,* $\boldsymbol{B}$*,* $\boldsymbol{C}$*,* $\boldsymbol{D}$*,* $\boldsymbol{E}$*, and* $\boldsymbol{F}$ *be fixed vertices in the Cartesian plane which form a regular hexagon. Define a type (2) walk as a non self-intersecting walk where beginning and ending vertices are spaced with one vertex between them. Among type (2) walks, the minimum distance is* $4s + \sqrt{3}s$*, where* s *is the side length of the hexagon.*

Proof. First let us determine the possible walks with beginning and ending vertices that have one vertex between them and do not include any intersecting edges. If we arbitrarily choose beginning and ending vertices to be **F** and **D**, then we arrive at the following four non self-intersecting walks of type (2):

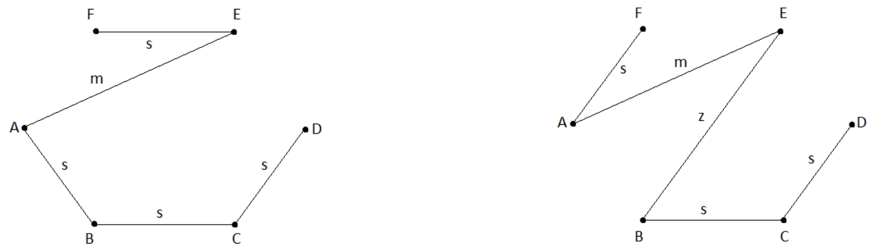

Figure 3.5: Walks I_2 and II_2

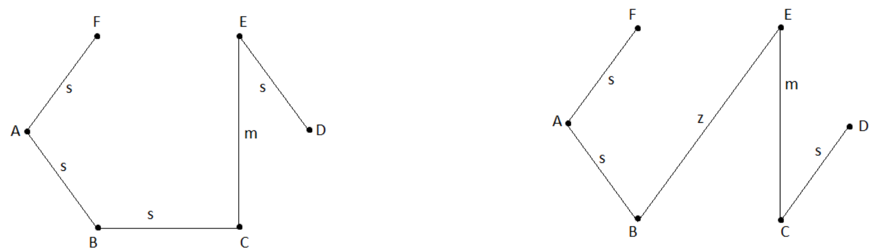

Figure 3.6: Walks III_2 and IV_2

From Figures 3.5 and 3.6, we note that walks I_2 and II_2 have the same distances as walks III_2 and IV_2, respectively. Therefore, it suffices to compare the distance of walk I_2 to walk II_2.

Expressing the the total distance of walks I_2 and II_2 using the notation described earlier, we arrive at the following equations:

$$d(I_2) = 4s + m$$
$$d(II_2) = 3s + z + m.$$

Since each circuit can be expressed in terms of s, determining the shortest distance becomes clear. We see that

$$d(I_2) = 4s + \sqrt{3}s$$

and

$$d(II_2) = 5s + \sqrt{3}s.$$

Hence $d(I_2) < d(II_2)$. Thus, of all type (2) walks, we consider $4s + \sqrt{3}s$ to be the minimum distance. Further, we can use I_2 to represent the shortest type (2) walk.

□

Proposition 3.4. *Let* $\boldsymbol{A}$*,* $\boldsymbol{B}$*,* $\boldsymbol{C}$*,* $\boldsymbol{D}$*,* $\boldsymbol{E}$*, and* $\boldsymbol{F}$ *be fixed vertices in the Cartesian plane which form a regular hexagon. Define a type (3) walk as a non self-intersecting walk where beginning and ending vertices are spaced with two vertices between them. Among type (3) walks, the minimum distance is* $6s$*, where* s *is the side length of the hexagon.*

Proof. We begin by determining the possible walks containing beginning and ending vertices that have two vertices between them and do not include intersecting edges. We arbitrarily choose **F** and **C** as beginning and ending vertices. There are six unique non self-intersecting walks, however three of the walks its are of the same distance as the other three.

Therefore we need only consider the three circuits below:

Figure 3.7: Walks I_3 and II_3

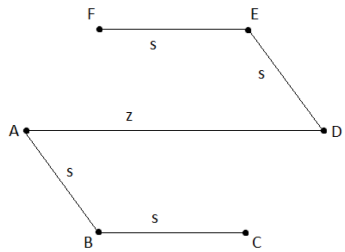

Figure 3.8: Walk III_3

We express the total distance of walks I_3, II_3, and III_3 in terms of s as:

$$d(I_3) = 3s + 2\sqrt{3}s,$$
$$d(II_3) = 4s + 2\sqrt{3}s,$$

and

$$d(III_3) = 6s.$$

Hence, $d(III_3) < d(I_3) < d(II_3)$.

Thus, of all type (3) walks, we find $6s$ to be the minimum distance. Further, we can use III_3 to represent the shortest type (3) walk.

□

We establish the following theorem regarding the coloring of Home points when the cities form a regular hexagon.

Theorem 3.5. *Let* $\boldsymbol{A}$*,* $\boldsymbol{B}$*,* $\boldsymbol{C}$*,* $\boldsymbol{D}$*,* $\boldsymbol{E}$*, and* $\boldsymbol{F}$ *be fixed vertices in the Cartesian plane which form a regular hexagon. Let* $\boldsymbol{H}$ *be an arbitrary Home point. When Home points in the plane are colored according to the circuit which gives minimum distance, exactly six unique colors are present. Furthermore, these minimal circuits contain the six possible type (1) walks.*

Proof. First we consider a Home point inside the regular hexagon. By Theorem 3.2, there exist six type (1) walks that are represented by six colors.

Now consider a Home point outside the hexagon. We need to show that there exists a circuit containing a type (1) walk, hereafter denoted **T**, which is minimal. We prove this in two parts.

Part 1 We need to show that there exists a circuit containing walk **T** that is shorter than a circuit containing walk I_2. Without loss of generality, let **F** and **D** be the beginning and ending vertices for the walk in I_2 defined above. Therefore, **HFEABCDH** defines the complete circuit containing I_2. With this choice of beginning and ending vertices **H** is restricted to being above $\overleftarrow{\mathbf{FE}}$ and to the right of $\overleftarrow{\mathbf{ED}}$ so to create a non self-intersecting circuit. We call this region **P**.
The proof of Part 1 continues by considering **P** to be divided by $\overleftrightarrow{\mathbf{BE}}$ into two regions.

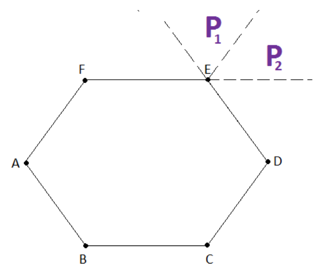

Figure 3.9: Appropriate Region

Part 1a Consider Home points located in region P_1. Then it follows that **H** is located to the left of $\overleftrightarrow{\mathbf{BE}}$ and to the right of $\overleftrightarrow{\mathbf{DE}}$. Therefore, $d(\mathbf{HF}) < d(\mathbf{HD})$. For Home points within this region, we choose **F** as an ending vertex for **T**. Note that along $\overleftrightarrow{\mathbf{CF}}$, $d(\mathbf{HA}) = d(\mathbf{HE})$. Therefore, since **H** has been restricted to the region to the right of $\overleftrightarrow{\mathbf{CF}}$, $d(\mathbf{HE}) < d(\mathbf{HA})$. Thus, we choose **E** as the remaining endpoint for **T**.

We need to show that if a salesman were to travel along the circuit containing I_2, he will always travel a longer distance than if he were to travel along the circuit containing **T**. Since the walks of these two circuits already have specified distances in terms of s, the distances of the complete circuits containing I_2 and **T** simplify to

$$d(\mathbf{HF}) + 4s + m + d(\mathbf{HD})$$

and

$$d(\mathbf{HF}) + 5s + d(\mathbf{HE})$$

respectively.

We want to show

$$\text{distance of circuit containing } I_2 > \text{distance of circuit containing } \mathbf{T}$$

which can be written as

$$d(\mathbf{HF}) + 4s + m + d(\mathbf{HD}) > d(\mathbf{HF}) + 5s + d(\mathbf{HE}).$$

This can be further simplified to

$$\sqrt{3}s + d(\mathbf{HD}) > s + d(\mathbf{HE}),$$

since we know $m = \sqrt{3}s$ by our previous trigonometric work.

Since we know any Home point along the perpendicular bisector of $\overline{\mathbf{DE}}$ is such that **HD**=**HE**, a Home point above this line will be closer to **E**. As Figure 3.9 shows, we limit **H** to being above $\overleftrightarrow{\mathbf{EF}}$. Moreover, **H** must certainly be above the perpendicular bisector of $\overline{\mathbf{DE}}$ since in P_1, $d(\mathbf{HF}) < d(\mathbf{HD})$. This proves the inequality,

$$d(\mathbf{HD}) > d(\mathbf{HE}).$$

Therefore, it follows that

$$\sqrt{3}s + d(\mathbf{HD}) > s + d(\mathbf{HE}).$$

Therefore,

$$\text{distance of circuit containing } I_2 > \text{distance of circuit containing } \mathbf{T},$$

which shows a circuit containing a walk of $\mathbf{T}$ has a shorter distance than a circuit containing a walk of type (2) in the region the portion of $\mathbf{P}$ where $d(\mathbf{HD}) > d(\mathbf{HF})$.

Part 1b Consider Home points located in region P_2. Then it follows that $\mathbf{H}$ is located to the right of $\overleftrightarrow{\mathbf{BE}}$ and above $\overleftrightarrow{\mathbf{FE}}$. Therefore, $d(\mathbf{HD}) < d(\mathbf{HF})$. For Home points in this region we choose $\mathbf{D}$ and $\mathbf{E}$ as beginning and ending vertices for $\mathbf{T}$. We need to show that if a salesman were to travel along the circuit containing I_2, he will always travel a longer distance than if he were to travel along the circuit containing $\mathbf{T}$. Since the walks of these two circuits already have specified distances in terms of s, the distances of the complete circuits containing I_2 and $\mathbf{T}$ simplify to:

$$d(\mathbf{HF}) + 4s + m + d(\mathbf{HD})$$

and

$$d(\mathbf{HD}) + 5s + d(\mathbf{HE})$$

respectively.

We want to show

$$\text{distance of circuit containing } I_2 > \text{distance of circuit containing } \mathbf{T}$$

which can be written as

$$d(\mathbf{HF}) + 4s + m + d(\mathbf{HD}) > d(\mathbf{HD}) + 5s + d(\mathbf{HE}).$$

This can be further simplified to

$$\sqrt{3}s + d(\mathbf{HF}) > s + d(\mathbf{HE})$$

since we know $m = \sqrt{3}s$ by our previous trigonometric work.

Since we know any Home point along the perpendicular bisector of $\overline{\mathbf{FE}}$ is such that $d(\mathbf{HF}) = d(\mathbf{HE})$, a Home point to the right of this line will be closer to $\mathbf{E}$. Figure 3.9 shows, we limit $\mathbf{H}$ to being above $\overleftrightarrow{\mathbf{FE}}$ since $d(\mathbf{HD}) < d(\mathbf{HF})$. Moreover, $\mathbf{H}$ must certainly be to the right of the perpendicular bisector of $\overline{\mathbf{FE}}$. In a regular hexagon, it follows that if $\mathbf{H}$ lies in a region above $\overleftrightarrow{\mathbf{FE}}$, then it also lies to the right of the perpendicular bisector of $\overline{\mathbf{FE}}$, proving the inequality,

$$d(\mathbf{HF}) > d(\mathbf{HE}).$$

Since $d(\mathbf{HF}) > d(\mathbf{HE})$ always holds for the region P_2, and $\sqrt{3}s > s$ it follows that

$$\sqrt{3}s + d(\mathbf{HF}) > s + d(\mathbf{HD}).$$

Therefore,

$$\text{distance of circuit containing } I_2 > \text{distance of circuit containing } \mathbf{T},$$

which shows that there exists a circuit containing a walk $\mathbf{T}$ which has a shorter distance than a circuit containing a walk of type (2) in the region P_2.

Part 2 Next, we consider a Home point outside the hexagon with circuits containing walks $\mathbf{T}$ and type (3). We want to show that there exists a circuit containing a walk $\mathbf{T}$ such that the total distance of such a circuit is smaller than a circuit containing a type (3) walk.

We arbitrarily choose $\mathbf{F}$ and $\mathbf{C}$ to be our beginning and ending vertices in III_3. Therefore, $\mathbf{HFEDABCH}$ is the complete circuit containing walk III_3.

Part 2a Assume $\mathbf{F}$ is the beginning vertex of $\mathbf{T}$. If $d(\mathbf{HE}) < d(\mathbf{HA})$, choose $\mathbf{E}$ as an ending vertex for $\mathbf{T}$, since $\mathbf{E}$ is closer to $\mathbf{H}$. With this choice of beginning and ending vertices $\mathbf{H}$ is restricted to lie in regions $\mathbf{P}_3$

and $\mathbf{P}_4$ so that the resulting circuit with a type (3) walk has no intersecting edges. See Figure 3.10.

We restrict $\mathbf{H}$ to lie in the region right of $\overleftrightarrow{\mathbf{CD}}$ and above $\overleftrightarrow{\mathbf{FE}}$. The left figure below defines such a viable region, $\mathbf{P}_3$. For this case, we can choose circuit with $\mathbf{T}$ for which there are no intersecting edges.

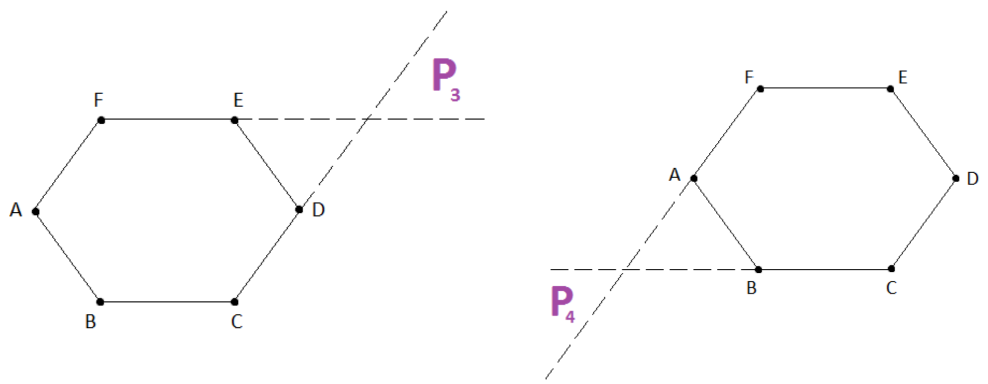

Figure 3.10: Appropriate Regions

Since the walks of these two circuits have distances in terms of s, the distances of the complete circuits simplify to:

$$d(\mathbf{HF}) + 5s + d(\mathbf{EH}) \text{ and } d(\mathbf{HF}) + 6s + d(\mathbf{CH}).$$

We want to show:

$$d(\mathbf{HF}) + 5s + d(\mathbf{EH}) < d(\mathbf{HF}) + 6s + d(\mathbf{CH}).$$

We further simplify the inequality to:

$$d(\mathbf{EH}) < s + d(\mathbf{CH}).$$

Since we know that any Home point along $\overleftrightarrow{\mathbf{AD}}$ is such that $d(\mathbf{EH}) = d(\mathbf{CH})$, a Home point above this line is closer to $\mathbf{E}$. As the region P_3 in Figure 3.10 shows, we limit $\mathbf{H}$ to being above $\overleftrightarrow{\mathbf{EF}}$. Thus, $\mathbf{H}$ is certainly above $\overleftrightarrow{\mathbf{AD}}$, proving the inequality

$$d(\mathbf{EH}) < d(\mathbf{CH}).$$

It follows that

$$d(\mathbf{EH}) < s + d(\mathbf{CH}),$$

since adding s to $d(\mathbf{CH})$ makes $d(\mathbf{CH})$ larger.

Part 2b In the alternative event that $d(\mathbf{HA}) < d(\mathbf{HE})$, choose $\mathbf{A}$ as an ending vertex for $\mathbf{T}$. Note that $\mathbf{F}$ is still the beginning vertex for $\mathbf{T}$. Again, we must define a region $\mathbf{P}_4$ in which $\mathbf{H}$ can lie so that there are not intersecting edges in walk III_3, and so we can choose $\mathbf{T}$ so that it does not contain intersecting edges. Therefore, $\mathbf{H}$ must lie below $\overleftrightarrow{\mathbf{BC}}$ and to the left of $\overleftrightarrow{\mathbf{AF}}$, as shown in the region P_4 in Figure 3.10.

Following a similar method from above, we compare the distances of the circuits, wanting to show:

$$d(\mathbf{HF}) + 5s + d(\mathbf{AH}) < d(\mathbf{HF}) + 6s + d(\mathbf{CH}).$$

Simplifying leaves us with

$$d(\mathbf{AH}) < s + d(\mathbf{CH}).$$

Again, we know that any Home point along $\overleftrightarrow{BE}$ is such that $d(\mathbf{AH}) = d(\mathbf{CH})$. Therefore, any Home point to the left of this line is closer to $\mathbf{A}$. In a regular hexagon, it follows that if $\mathbf{H}$ lies in a region to the left of $\overleftrightarrow{\mathbf{AF}}$, then it also lies left of $\overleftrightarrow{\mathbf{BE}}$, proving the inequality,

$$d(\mathbf{AH}) < d(\mathbf{CH}).$$

If $d(\mathbf{AH}) < d(\mathbf{CH})$, then it certainly follows that

$$d(\mathbf{AH}) < s + d(\mathbf{CH}).$$

$\square$

We have shown that circuits containing walks of type (1) are the shortest circuits for Home points lying inside or outside the regular hexagon. Therefore, a circuit containing a walk of type (1) always produces the optimal circuit, regardless of the location of the Home point. For cities arranged in a regular hexagon, only the six colors representing the six circuits with type (1) walks appear. An illustration is given in Figure 3.11.

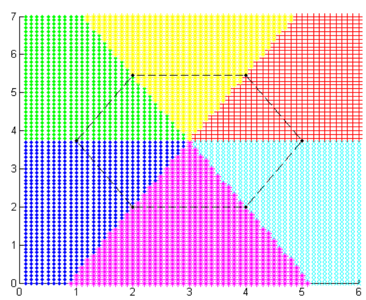

Figure 3.11: Example of results of Theorem 3.5

4. Future Work

Using the information we have found thus far, and considering all of the possible variations of the TSP, we hope to further explore the problem in the future. Options for future research include investigating the TSP with asymmetrical distances, various numbers of cities, multiple modes of transportation, or moving vertices. Extending on our six city case, we could investigate the TSP when the six cities are not arranged in a regular hexagon. All of these variations could be considered with varying Home points, as we have studied in this paper.

Acknowledgments

We would like to thank National Security Agency(NSA) and The National Science Foundation(NSF) for making SUMSRI possible. We greatly appreciate the direction and motivation from Dr. Lakeshia Legette and Brett Jefferson while completing our research. We thank Miami University, Dr. Patrick Dowling, Dr. Reza Akhtar, Dr. Tom Farmer, the remaining faculty, staff and facilities. Finally, we extend many thanks to the 2010 SUMSRI participants for their friendship, advice, and encouragement.

References

[1] K. Bellino and R. Narasimhan, *On the Analytic Geometry of the Traveling Salesman Problem*, SUMSRI Journal, 1999.

[2] J. Herbert, *The Traveling Saleswomen Meets Sperner*, SUMSRI Journal, 2000.

[3] L. V. Quintas and F. Supnick, *On some properties of shortest Hamiltonian circuits*, American Mathematical Monthly **72**(1965), 977-980.

[4] D. Sanders, *Metric Spaces in which Minimal Circuits Cannot Self-Intersect*, American Mathematical Society, **56**(1976), 383-387.