CONTOUR PLOTTING OF RESPONSE SURFACES

By

EDGAR LOUIS BUTLER

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1963

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1965

Submitted to the faculty of the Graduate
College of the Oklahoma State University
in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY
May, 1967

CONTOUR PLOTTING OF RESPONSE SURFACES

Thesis Approved:

Carl E. Marshall
Thesis Adviser

David L. Weeks

Robert D. Morrison

Paul A. McCollum

D. D. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to thank Dr. Carl E. Marshall, chairman of my advisory committee, for the valuable contribution of his time and knowledge in the preparation of this thesis. Indebtedness is also acknowledged to the members of my advisory committee: Dr. Robert D. Morrison, Dr. David L. Weeks, and Prof. Paul A. McCollum. My appreciation is extended to Dr. Dale D. Grosvenor for his invaluable guidance and assistance in connection with this study.

I am grateful to Dr. J. Leroy Folks for his suggestion of the thesis problem and to Mr. Jim Stewart for sharing with me his experiences in the field of contour estimation.

Finishing a thesis in absentia requires the encouragement and understanding of one's employer. I am deeply grateful, therefore, to Prof. Ralph Lee and Dr. Charles Antle of the University of Missouri at Rolla. Dr. Antle has provided vital assistance with his criticism of this manuscript.

I am indebted to the computer centers at Oklahoma State University and University of Missouri at Rolla for the use of their computing equipment.

My deepest thanks go to my wife for her patience and impatience while I have been completing this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure | Page

CHAPTER I

INTRODUCTION

Graphical analysis of experimental data is an important tool in many fields of study. In particular, the investigation of response surfaces in geology, meteorology, mineralogy, oceanography, and other related fields often involves one aspect of graphical analysis, namely, contour mapping. The problem of estimating response surface contours for the purpose of plotting contour maps is the subject of this thesis.

Contour plotting in a three dimensional situation is simply a way of reducing an unwieldy three dimensional graph to the more familiar two dimensional graph. Consider a function of two variables, $z=f(x,y)$. For a given value of z, say z_0, define $P_{z_0}=\{(x,y)\mid f(x,y)=z_0\}$. The locus of the set of points P_{z_0} is called a contour of z_0. A graph showing contours for various z values is called a contour plot of the function $f(x,y)$. The reader may consult Cochran and Cox (1) for a further discussion of contouring.

Given a known function, one should, theoretically, always be able to construct a contour plot. For functions such as the quadric surfaces, contours of equal response are easily found by analytic methods. Even if the function is not as well behaved as the quadric surfaces, it can be evaluated at enough points to construct contours. Unfortunately, in experimental situations where a contour plot is desired, the function in question is rarely known except at sample data points. In fact, determining estimated contours of the function may be the major purpose of the experiment. To do this, one must have an algorithm for estimating certain values of the function between sample data points.

Before the advent of high-speed computing equipment, the construction of surface contour maps from experimental data was generally done by hand. That is, the researcher plotted the experimental data points on a graph, and, with the assistance of a French curve or some similar device, drew the contours where he thought they should lie. The time required to complete computations prohibited, for the most part, the use of numerical methods for estimating contours.

Since computers have become widely available there have been two basic approaches to the problem of contour estimation. The first of these approaches is the least squares polynomial fit. In this process the researcher selects a model to fit to the experimental data. This model is usually a polynomial in x and y of order two or higher. The coefficients of the polynomial are obtained by finding the set of coefficients which minimizes the sum of squares of deviations between the polynomial model and the observed response points. Contours of equal response may be plotted once the set of polynomial coefficients is known.

The least squares fitting technique has a disadvantage because irregular surfaces may require the use of high order polynomials to achieve a good fit to the data. It is well known that high order polynomial surfaces may fluctuate widely over small areas of the x-y plane. Thus, the result may be that, between data points, the polynomial representation of the surface behaves in a manner not characteristic of the true response surface (5). Since surface fitting techniques have been thoroughly discussed in the technical literature, no further description will be made in this study. For the reader interested in these techniques, Krumbein (7) is suggested as a reference.

The technique of nonlinear precise-data-fit is the second approach to the problem of contour estimation. This approach attempts to overcome the difficulties which arise in least squares fitting as a result of using high order polynomials. Basically, the precise-data-fit techniques fit a series of surfaces over small areas of the x-y plane, and then combine these surfaces to form a highly nonlinear representation of the response surface over the area of data collection. Working on small sections of the area of data collection allows the use of low order polynomial models for each section. The representation which results fits the data precisely and is not allowed to behave in an unrestrained manner between data points. Two estimation procedures which make the precise-data-fit approach to contour mapping will be described in Chapter II. From the discussion of these methods it shall become evident that a new procedure is needed to produce accurate contour estimates for randomly spaced data points.

In Chapter III a method is developed for obtaining contour estimates of a single valued continuous function of two variables with no restrictions on the spacing of data points in the x-y plane. A detailed description of the computer algorithm for this method is included along with a discussion of the mathematical principles upon which the technique is based.

Because of the tremendous ability of computers to process large amounts of data in short periods of time, too many researchers seem willing to assume that the answers the computer prints out are *the* answers to the problem. However, in a computer program which processes experimental data there are three sources of errors: round-off errors which are inherent in the computer, experimental errors in the data, and errors in the computing algorithm. A thorough program produces not only results but some analysis of these errors. Chapter IV is a discussion of the nature of these errors for the contour estimation problem and provides procedures for analyzing them.

Chapter V presents results obtained by using the estimation method of Chapter III. Also included in this chapter are results of the error analysis and a comparison of all the methods of contour estimation which have been included in this study.

CHAPTER II

TWO EXISTING PLOTTING METHODS

To indicate the usual approach to contour estimation in precise-data-fit techniques and to consider the need for a new procedure, two methods will be discussed here. The discussion will be general in nature because, for each method, there are details which may differ according to the application. The first method will be called the grid method and the second will be known as the quadrilateral method.

These two methods, as well as the procedure to be described in Chapter III, have in common a technique known as triangulation. Let A be a specified area in the Euclidean plane. In the case of the contour estimation procedures, A is the area of data collection. Triangulation of A, according to Moise (8), is simply the process of subdividing A into a set of triangles $\{A_i\}$ such that $A_i \cap A_j = \phi$ for all i and j when $i \neq j$, and $\bigcup_{i=1}^{r} A_i = A$. For the contour estimation methods a further restriction imposed upon the triangulation of A is that each triangle vertex be a point at which some estimate of the response variable is known. This triangulation lends itself to a simple mathematical procedure for estimating contours, which will be described later.

All the methods discussed in this study differ from one another in the manner of achieving triangulation. The grid method uses the sample data to find estimates of the response surface at certain points on a predefined triangular system. The method developed in Chapter III fits a triangular system to the sample data points themselves. The quadrilateral method uses a technique which lies somewhere between these two extremes.

Triangulation in the Grid Method

The first step in the grid method is the selection of a rectangular grid to be superimposed on the area of data collection in the x-y plane. The researcher must determine the size and positioning of this grid. The next step is to obtain an estimate of the response variable at the mesh points of the grid. In order to estimate the functional value at a mesh point there must be nearby a sufficient number of values which adequately surround the point. The estimate is a weighted average of the sample responses at these nearby points. There is no specific rule which dictates the number of sample points to be averaged or the weighting factor to be used. As a simple example consider the situation illustrated in Figure 1. One function for estimating the response at the mesh point shown could be

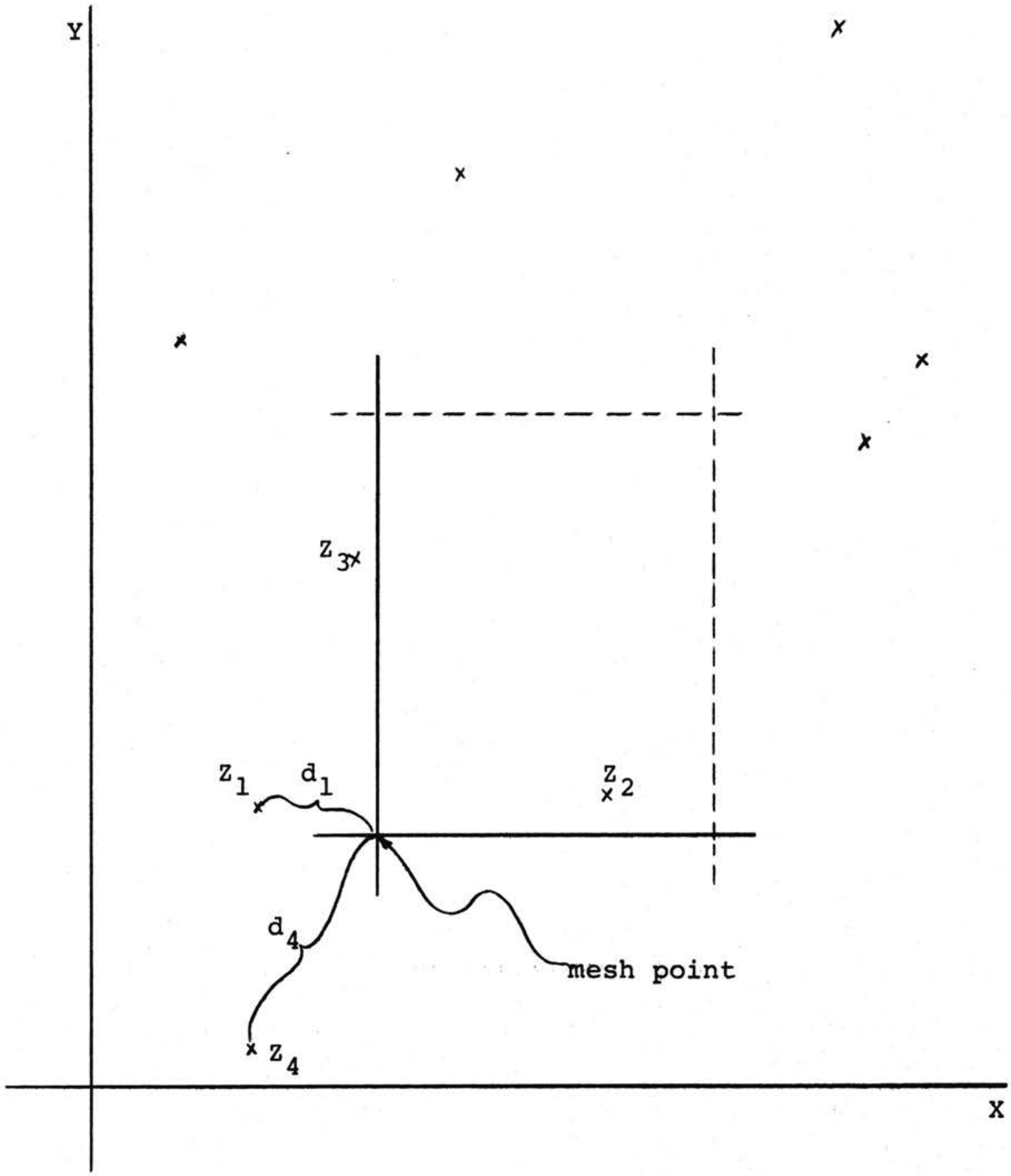


Figure 1. Estimation of Mesh Point Response in Grid Method

$$Z_m = \frac{\sum_{i=1}^{4} Z_i/d_i}{\sum_{i=1}^{4} 1/d_i} \, , \qquad [2.1]$$

where the Z_i are the observed responses at the four data points nearest the mesh point and d_i represents the distance between the (x,y) coordinates of the ith point and the mesh point. To reduce extrapolation it would be desirable to place some restriction on the magnitude of the d_i.

Once responses at these mesh points are estimated, the mesh values may be used in an interpolation surface to estimate responses on a smaller grid before proceeding to triangulation. If this is done, a nonlinear interpolation method such as Newton's Interpolation Formula or LaGrange's Interpolation Method (9) is generally used.

When the grid size is sufficiently small, the triangulation is effected by dividing into triangles the rectangles formed by the grid. One simple rule for division of the rectangles would be to construct the diagonal which runs from the lower left corner to the upper right corner of the rectangle. In practice a more complicated rule may be used to divide the rectangle into more than two triangles. International Business Machines has programmed one version of the grid method for use on an IBM 1620 digital computer. The User's Manual (10) for this program explains in detail the procedures used in this version for estimating mesh values, reducing grid size, and triangulating the rectangles.

Triangulation in the Quadrilateral Method

The quadrilateral method differs from the grid method in that it attempts to fit the triangles to the data points rather than the data points to the triangles. This method trades the simple mathematics of the grid method for a more general figure and a more direct attack.

A grid size is chosen by the researcher such that there is approximately one and only one datum point inside each rectangle. When two or more points lie within a rectangle, some weighted average of the points is used to reduce the number of points to one point. One could, for example, find the simple arithmetic average of all the data point responses within a rectangle and consider this to be an estimate of the response at the x-y centroid of the points. When a rectangle is void of data and there are a significant number of surrounding points, another weighted average is used to supply an estimated response for that rectangle. An estimation function similar to the one mentioned in the grid method would be applicable here. An optimum size of grid is one which leaves the fewest number of rectangles which are either void of data or contain more than one data point.

Having established the grid, all possible sets of four rectangles having a common vertex are chosen, and, for each set, the four data points within a set form a quadrilateral. Triangulation is completed by constructing the shorter diagonal, thereby forming two triangles within each quadrilateral. Although the author knows of no published version of the quadrilateral method, oral conversation with Mr. Jim Stewart of the Pan American Petroleum Company indicates that it has been used by that company.

From Triangulation to Contour Estimates

Once triangulation of the area of data collection has been accomplished, both the grid method and the quadrilateral method use the same procedure for obtaining contour estimates. For each triangle, linear interpolation between the vertices is done to determine the points where a given contour enters and leaves the triangle. In general there will be two such points, and these are joined by a straight line segment which is taken as the estimate of the path of the contour through the triangle.

There are three special cases which arise when one or more vertices of a triangle have a response equal to the value of the contour. If only one vertex has a response equal to the value of the contour and this response is the maximum or minimum response of all three vertices, then this one point becomes the estimate of the contour in the triangle. If the response at the vertex is not the maximum or minimum response on the triangle, the contour enters the triangle at the vertex and must leave the triangle at some point on the side opposite the vertex. In the case where two vertices have responses equal to the contour value, the side of the triangle joining these vertices is taken as the contour estimate. The case where all three vertices of a triangle have a response equal to the contour value presents several options, one of which is to regard the triangle boundary itself as the contour estimate.

When the contour has been estimated in every triangle by a line segment, these line segments form a continuous broken line as the final contour estimate. The procedure above is repeated until estimates of all desired contours have been constructed.

Smoothing Procedures

For aesthetic reasons it may be desirable to smooth the continuous line segments which result from the above estimation procedure. In the grid method a closer approximation to a smooth curve can be achieved by reducing the size of the grid system. As the grid size decreases, the line segments comprising a contour become shorter so that a smooth curve is more closely approximated. Another process which can be applied to any method is to fit some continuous curve through a series of vertices of the contour estimate. Since contours, in general, are not single valued functions in the x-y plane, an equation for the smoothing curve could not be written in functional notation. At best, it might be possible to express the equation of the curve in parametric form. Because of this, the mathematics could become very cumbersome. Hand smoothing may prove to be the most economical method. However, experience will show that no two people smooth a curve the same way, and, therefore, the original broken line estimate should not be discarded.

Conclusions

In the previous sections of this chapter the grid and quadrilateral methods have been described without regard to the consequences of using these procedures. That is, no attempt has been made to justify or criticize any portion of these methods. The author feels there are some points in both procedures which need to be discussed because of their effects on the resulting contour estimates.

Grid Size and Placement

Both the grid and the quadrilateral methods require the researcher to determine the size and location of the grid system to be used. Since these two factors will influence the contour estimates, the researcher may be faced with the temptation to experiment with different grid sizes and various grid placements until the resulting contours are as he theorizes them to be. In this case, the researcher's experiments with the grid may influence the contour estimates more than the data he collected.

Aside from disposing with the grid system altogether, one solution to the above problem would be to make the determination of grid size and location an integral part of the estimation procedure itself. That is, some criterion should be developed which specifies an appropriate grid system for producing accurate estimates. The condition for the quadrilateral method of using the system which has the fewest number of grids to be modified is such a requirement. However, because of the large number of possible grid systems, it may be difficult to meet this requirement in practice. Further restrictions, based upon experience, would be in order here. Of course, in the quadrilateral method, data collected on a fairly regular system will not require modification of any grids.

Estimation of Mesh Points and Grid Responses

The next step in the grid method of estimating response values at mesh points of the grid system is particularly unpalatable to the statistician since it consists in a sense of throwing away information. Responses are estimated at points where no data was collected and the original data points are discarded in further computations. In addition to the experimental error already inherent in the data, this procedure of estimating mesh responses introduces interpolation error in the data. This makes optimization of the grid system for the quadrilateral method desirable and shows one advantage of the quadrilateral method over the grid method.

Grid Reduction

The grid method allows the researcher the option of reducing grid size before obtaining final contour estimates. This option should be exercised with caution because, as the IBM User's Manual (10), p. 10 states, "The derived grid is not more accurate." There can be no more information in the derived points on a smaller grid than in the original data, so attempts to improve accuracy by reducing the grid size are futile. One asks then, what is the purpose of reducing grid size? As mentioned before, a smaller grid size will tend to produce smoother contours than a larger one, so that if one desires above all a smooth contour, grid size reduction will help achieve this goal. Also, there is a process whereby two contour maps may be compared, which may require the derivation of a smaller grid.

Improving Accuracy

One procedure which might be used to increase accuracy in the contour estimates in nonlinear interpolation. However, this requires a priori knowledge about the nature of the true response surface. There are so many unknown factors in nonlinear estimation that one may actually be defeating his purpose by using this method to obtain accuracy. Also, there is the temptation to search through many models until one is found which pleases the researcher or supports his theory without regard to finding the true response model. These are some of the reasons why linear interpolation has been used to obtain contour estimates from the triangulation.

There is at least one step which can be taken to improve the accuracy of estimates obtained by linear interpolation. According to Conte (3), accuracy is dependent upon the interval of interpolation, smaller intervals yielding more accurate estimates. If data are collected so that the distance between points is small, and if some attempt is made to minimize the lengths of the sides of the triangles to be used, linear interpolation can provide accurate estimates. This is the reason the shorter diagonal of the quadrilateral is used in triangulation for the quadrilateral method.

Advantage of Linear Interpolation

Using linear interpolation on the triangles derived in the grid and quadrilateral methods has some definite advantages. The first is that two contours of different responses will not cross, which is necessary in describing a single valued continuous response surface such as is considered in this study. If two contours were to cross, it would indicate a point where the surface has two values and is, therefore, undefined or discontinuous. To prove that the contour estimates will not cross, it will suffice to consider a triangle with responses z_1, z_2, and z_3 at the vertices p_1, p_2, and p_3, such that $z_1<z_2<z_3$. Suppose that the estimated contours for responses c_1 and c_2 are such that $z_1<c_1<c_2<z_3$. By the previously discussed algorithm these two contours must pass through the triangle. Let c_{11} and c_{12} be the points where the contour estimate of c_1 enters and leaves the triangle. Likewise, let c_{21} and c_{22} be the points of entry and exit of the contour estimate of c_2 as

in Figure 2. Let

$|ab|$ denote the distance in the x-y plane from point a to point b.

Then $|c_{21}p_3| < |c_{11}p_3|$ along the path from p_1 to p_3 ,

and $|c_{22}p_3| < |c_{12}p_3|$ along the path from p_1 to p_2 to p_3. This means that the entire line segment from c_{21} to c_{22} is nearer p_3 than the line segment from c_{11} to c_{12}, and, thus, the segments cannot cross. Since the entire area is covered with triangles, the above proof for one general triangle is sufficient to prove that no contours cross in the entire data area.

One problem in the grid method is that contour estimates may be drawn on the wrong side of the data points. That is, a sample point with a response of 47 will not always lie where it is expected to lie, namely between the 40 and 50 contours. It is reasonable to expect a sample point to lie between the two contours which bracket the value of the observed response since the expected value of experimental error is assumed to be zero. In the quadrilateral method, if no grid values have to be estimated, the contours will always be drawn on the proper side of data points. This is a second important feature of linear interpolation from the triangles when the triangles have data points as vertices.

A Criticism of Smoothing Procedures

The final step of smoothing contours is a difficult

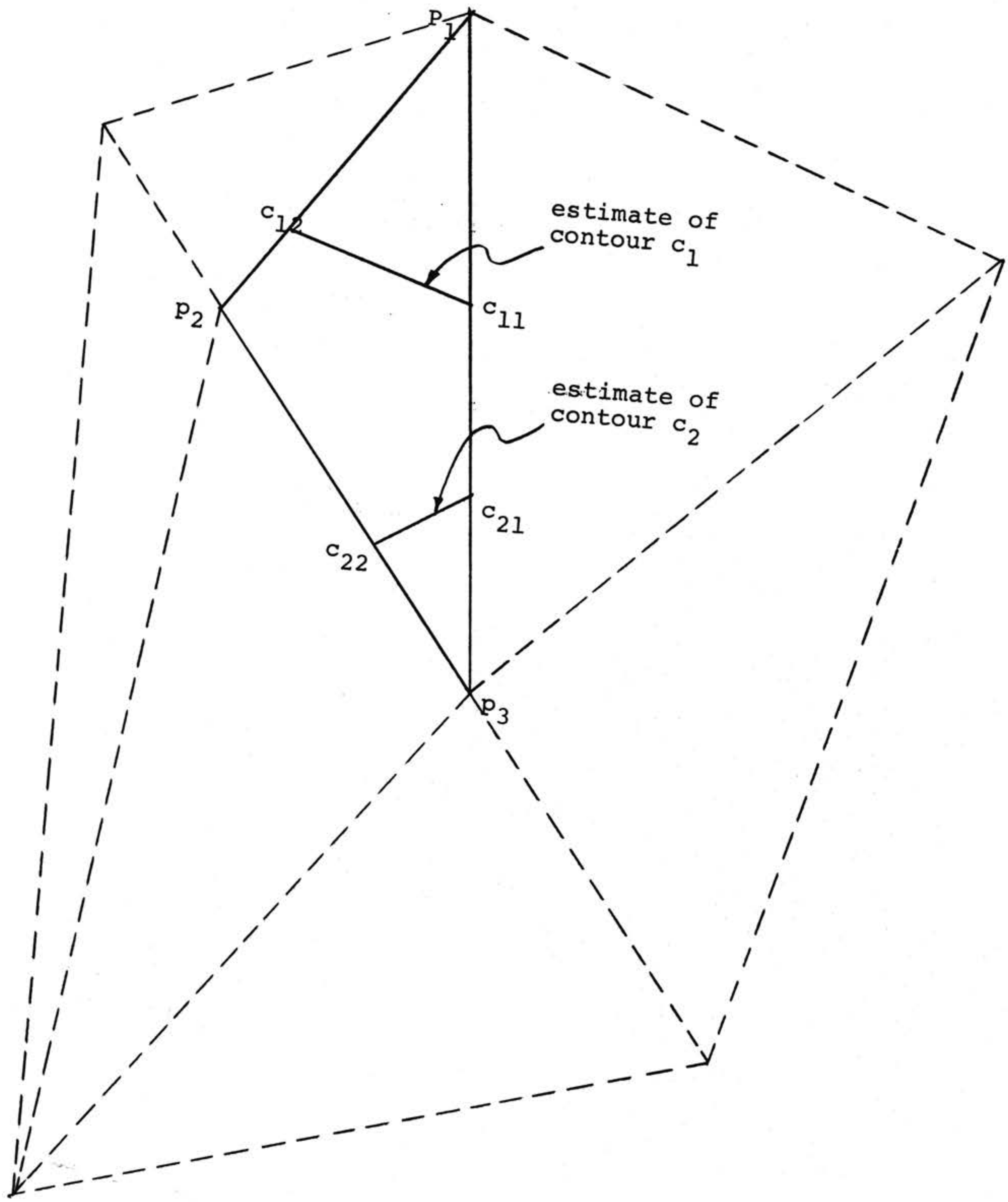
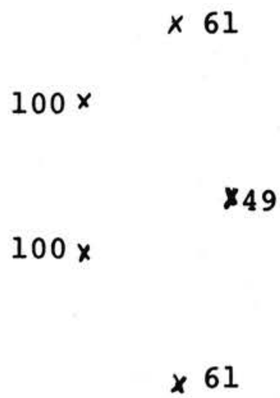
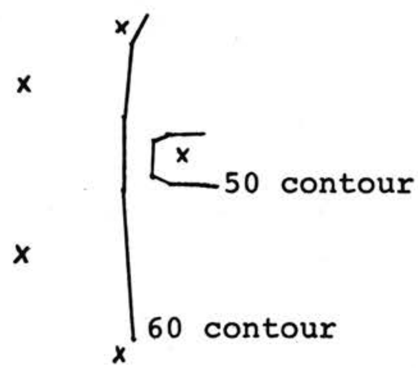


Figure 2. Contour Estimates Obtained by Linear Interpolation

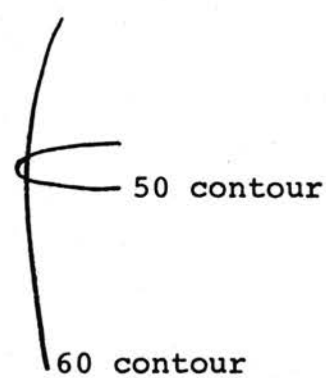
task. Under certain smoothing techniques, two contours may cross as in Figure 3, which would be contrary to the definition of a single valued continuous function. Since smoothing is a final step, hand smoothing might be acceptable were it not that each person will smooth the contours differently. Therefore, the broken line segments should not be destroyed or completely disregarded. As a result of discussions with researchers, the author feels that the only reason estimates need to be smoothed is that researchers are accustomed to seeing the smooth contours which were made by hand before computers made contouring by applied mathematical methods practical. Consequently, attempts should be made to convince the researcher that broken line contours are at least as accurate and have as much meaning as smoothed contours.



Data Points and Response Values



Broken Line Estimates for Contours 50 and 60



Results of Smoothing Procedure

Figure 3. An Example of Contour Smoothing, Invoking a Cross

CHAPTER III

DEVELOPING A NEW TECHNIQUE

The following description is of a contour estimation technique which attempts to improve upon the flexibility of the quadrilateral method of the preceding chapter. The quadrilateral method will yield accurate contour estimates with very little discarded information if the data are collected in at least a semiregular system. However, if the data points are irregularly or randomly spaced, the accuracy of estimates obtained by the quadrilateral method becomes questionable since information will be discarded and responses estimated at points where no data is taken. The method of this chapter does not use a grid system and needs only observed responses at data points; i.e., regularly spaced data points are not required.

As a preliminary step to the procedure, the data points are read into the computer and stored as three vectors: $X=\{x_i\}$, $Y=\{y_i\}$, and $Z=\{z_i\}$, $i=1,2,\ldots,N$, where (x_i,y_i,z_i) is the Cartesian representation of the ith datum point. Some sorting of the elements in the X and Y vectors is done to determine if there is any replication. Replication occurs when an x-y combination appears in the data more than once. For ease of computing error terms, it is necessary that the experiment be balanced which means that there is equal replication at all data points. This study shall be restricted to balanced experiments. If there is replication, the responses at each point are averaged to obtain the element of the Z vector which corresponds to that point. After this is accomplished, the number of elements in the vectors X, Y, and Z is reduced to n, where rn=N, and r equals the number of replications.

The three major steps in generating the contour estimates are: (1) finding a set, $\{P_k\}$, of convex polygons such that $P_k \subset P_j$ for $j<k$, (2) joining these polygons by straight lines to complete triangulation, and (3) interpolating within each triangle to find estimates of all contours which pass through the triangle.

Constructing Convex Polygons

The first step of finding the set of convex polygons is merely a convenience which facilitates the formation of triangles over the area of data. The polygons order the data in a fashion which allows a quick search to find appropriate triangles, thus avoiding a time-consuming continual search through all the data points.

According to Karlin (6), the convex hull of a set X is defined as the smallest convex set which contains X. The first polygon to be constructed is the convex hull of the set of data points. The reason for constructing the convex hull of the data is that the convex hull defines the maximum area of estimation without extrapolation.

Using all data points as vertices of the triangles, it may be shown that all triangulations of the convex hull of the data result in the same number of triangles. Suppose that the convex hull has m data points as vertices. If there are n data points, (n-m) points will lie within P. From plane geometry, the sum of the interior angles of P is (m-2)180 degrees. When drawing triangles between the n points, each of the (n-m) interior points of P will represent 360 degrees. The total (n-m)360+(m-2)180 represents the number of degrees to be divided among the triangles. Each triangle will use 180 degrees so that there are exactly $\frac{(n-m)360+(m-2)180}{180} = 2n-m-2$ triangles which may be formed.

The method of constructing the convex hull consists first of a search through the vectors X and Y for the minimum and maximum x_i and y_i. By assuming lines drawn parallel to the x and y axes through these points, a rectangle is formed which contains all the data points. Let the sides of this rectangle be labeled X_L, X_U, Y_L, and Y_U as in Figure 4. During the search the minimum and maximum x_i on Y_L and Y_U are found along with the minimum and maximum y_i on X_L and X_U. This produces eight points, some of which may be duplicates. An octagon is formed by joining these points, beginning at the maximum y_i on X_L and continuing counterclockwise around the rectangle. This octagon may be degenerate if some of the eight vertices are duplicates. The data points which will make up the convex hull are found on

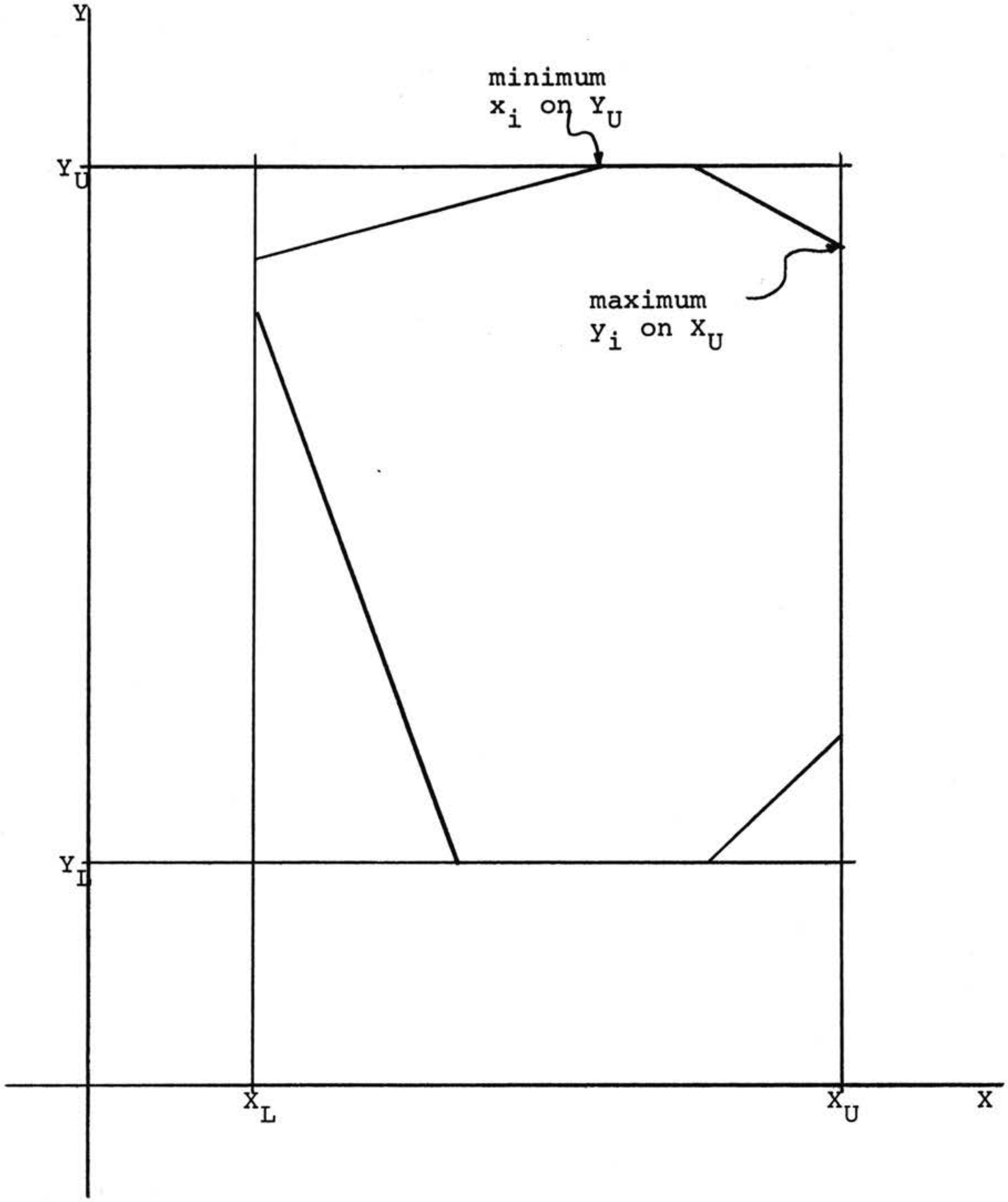


Figure 4. Construction of Convex Hull of Set of Data Points

or outside of this octagon, but on or within the rectangle.

To increase the speed of the search for vertices of the polygon, some sorting of the points in the area between the octagon and the rectangle is done. The next problem is to determine which of these points are vertices of the convex hull. Consider a polygon of p sides and form vectors of the sides with the vectors directed counterclockwise around the polygon. It is easily shown that the polygon is convex if and only if the cross product of the vector pointing toward a vertex with the vector pointing away from the vertex is nonnegative for all p vertices.

Suppose that one side of the convex polygon is known. The above idea is used to find points which may qualify as being the next vertex of the polygon. Of course, there may be many points which form vectors such that their cross product as defined above is nonnegative. This means that another criterion must be used to decide which point is, in fact, the vertex of the desired polygon. The point in question is the one which forms the largest interior angle with the preceding vertex. These criteria are used repetitively, working counterclockwise around the polygon, until all vertices are established.

If the data points on the boundary of the convex hull are disregarded, another convex hull may be generated in the same way for the remaining data points. Using this procedure repetitively, the series of convex polygons is obtained.

In the computer the elements in the vectors X, Y, and Z are reordered to match the order of the vertices of each polygon. The ordering begins with the point having the maximum y value on X_L and proceeds counterclockwise around the polygon. No distinction is made in X, Y, or Z to indicate when a new polygon is started. Therefore, another short vector is defined to indicate the starting points of each polygon, and the number of polygons generated is noted. Care must be taken here in that the innermost polygon may contain only one or two points. If this is the case, the polygon will not have positive area, but in general terms it is still a convex polygon.

Triangulation

The second major step is to find connecting lines to form triangles between the polygons. The criterion used for constructing the triangles is the fact that the errors of linear interpolation are dependent upon the interval of interpolation. Clearly, triangles with short sides should produce more accurate contour estimates than triangles with longer sides. Therefore, for every choice of lines to construct triangles, the shortest line will be used.

The first line may be selected from the set of all lines joining vertices of the first two polygons, excluding lines which pass through the inner polygon. To save time, any convenient point on the inner polygon is chosen, and the shortest line between it and any point on the outer polygon which is wholly between the two is chosen as a starting place. Then, moving counterclockwise around the polygons, consider the next point of the inner polygon and the next two points on the outermost polygon. Using the points at the ends of the original line segment and these new points, three triangles are drawn as in Figure 5. The first triangle has as vertices the two original points and the next point on the outermost polygon. The vertices of the second triangle are the two original points and the next point on the inner polygon. The third triangle has vertices at the original point on the outermost polygon and the next two points on this same polygon. It is noted that each of the three triangles has one side S, say, which is not a side of either polygon. The three triangles are compared, and the one for which the side S is shortest is selected as the triangle to be used. When the choice between these three triangles is made, at least one of the other two triangles is excluded as a possibility. In addition, if the side S is such that the triangle includes area other than the area between the two polygons, the triangle must be discarded. This restriction on the triangles assures nonoverlapping triangles over the entire area of data collection.

To continue, the side S is now substituted for the original line in the case that one of the first two triangles is chosen. The side S is substituted for the original two sides of the outer polygon if the third case is selected. The search is continued, again moving in a counterclockwise

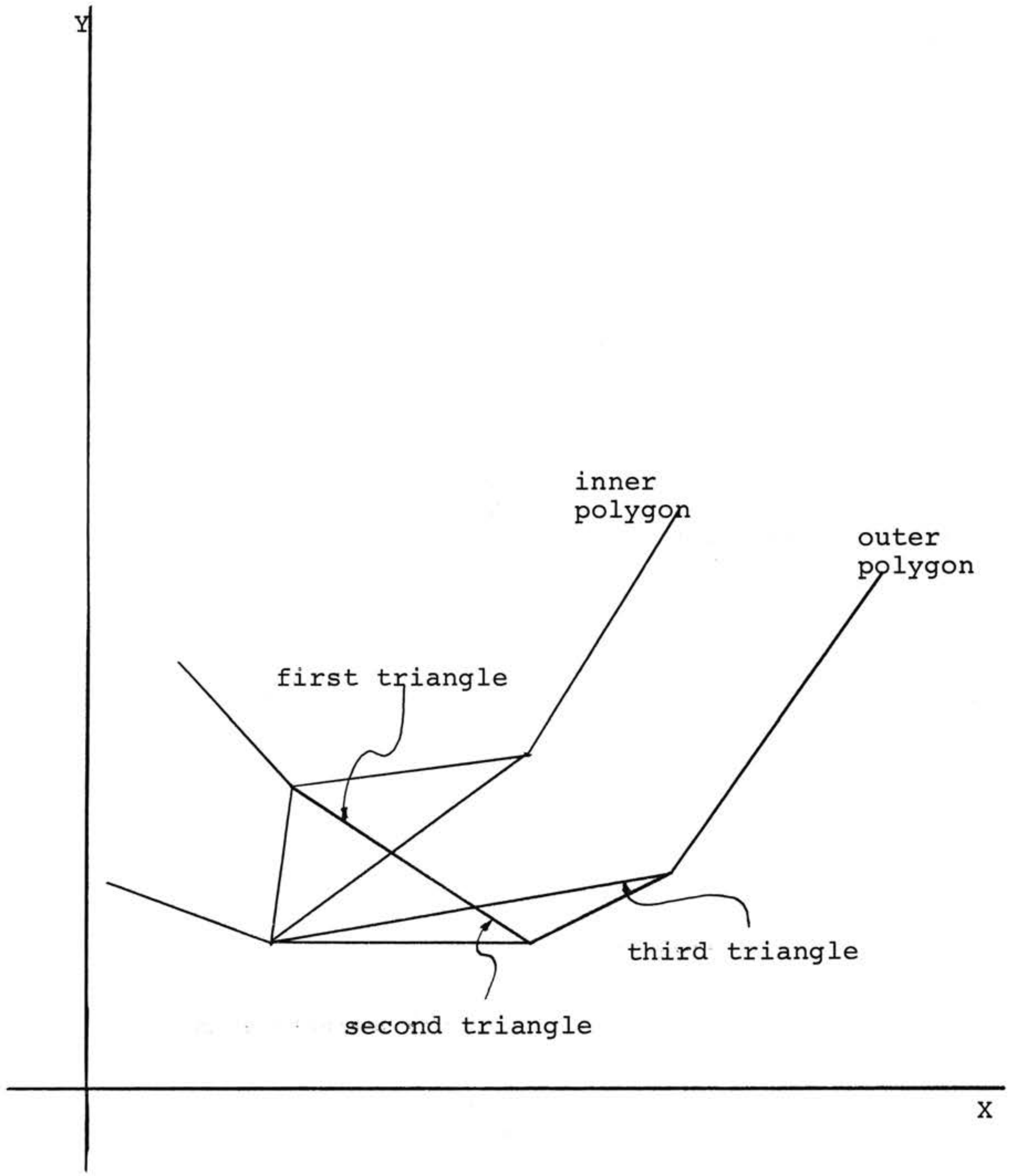


Figure 5. Triangulation Between Two Convex Polygons

direction. Once all triangles between the two outermost polygons are established, the procedure moves in to consider the next two adjacent polygons.

When triangles have been drawn in all pairs of adjacent polygons, a check must be made on the area of the innermost polygon. If this area is nonzero, triangles must be drawn within the innermost polygon in order to complete the configuration. If the polygon is a triangle, no further steps are necessary. However, if the polygon has four or more sides, a series of steps is repeated until the area is divided into triangles. Consider four points in series around the polygon and form a convex quadrilateral by joining the first and last points as in Figure 6. The shorter diagonal of this quadrilateral is found and becomes the third side of a triangle within the polygon. Deleting this triangle from the polygon, these steps are repeated until the entire polygon is divided into triangles.

Interpolation

Once the triangles have been generated, the third major step of interpolating to get contour estimates may be performed. This procedure is the same as described in Chapter II. The vertices are ordered according to increasing responses to aid the procedure.

In practice, the computer program for this procedure does not find all the triangles before proceeding to interpolation. In order that no computer storage be wasted in

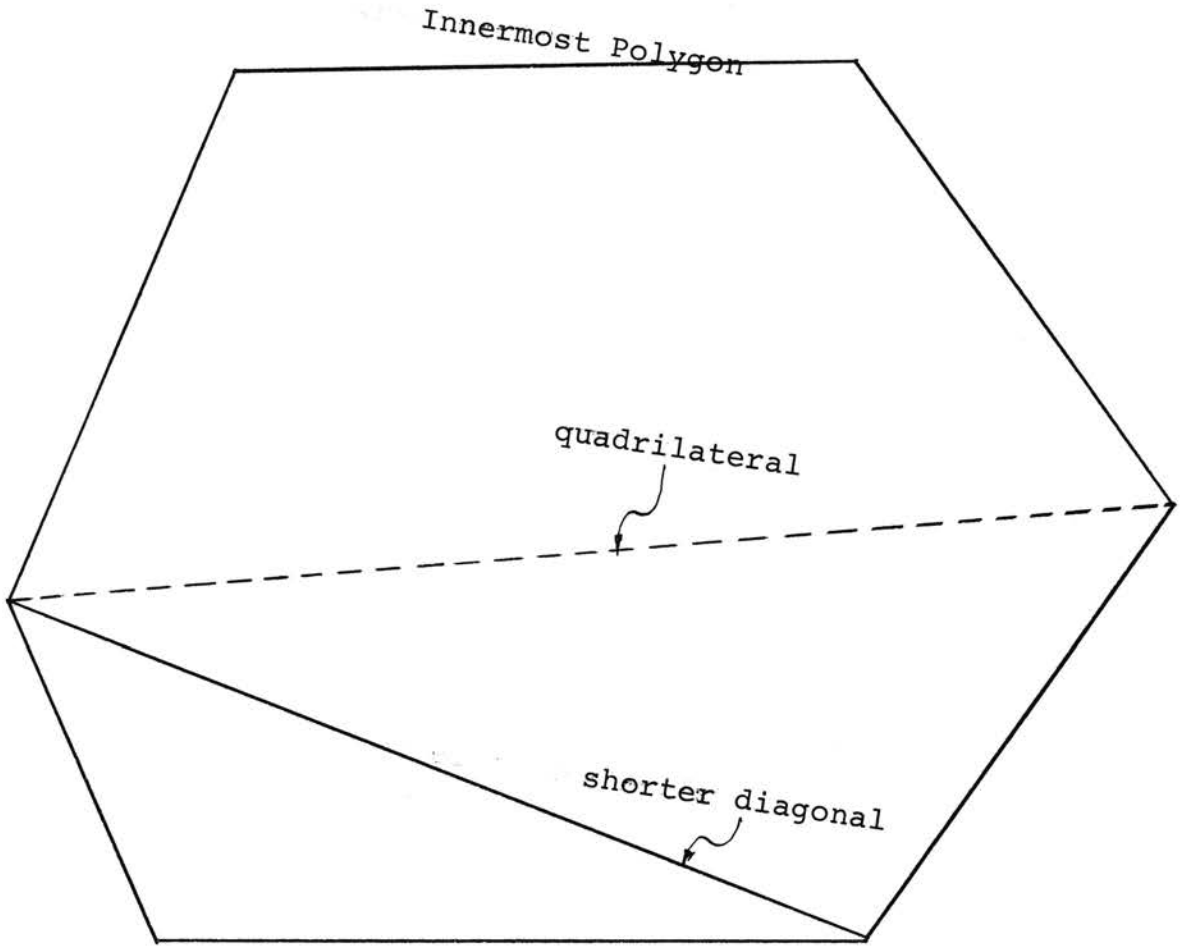


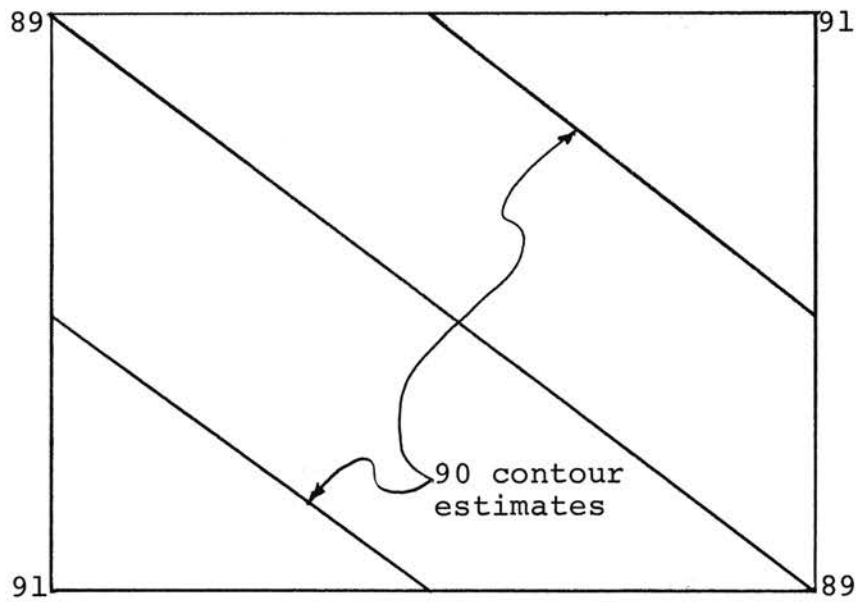
Figure 6. Triangulation of Innermost Polygon

saving the triangles, the program interpolates to find the estimates immediately after each triangle is generated. The contour estimates are recorded on auxiliary storage, again saving internal memory. The program then returns to generate another triangle and find contour estimates, continuing until all triangles have been generated.

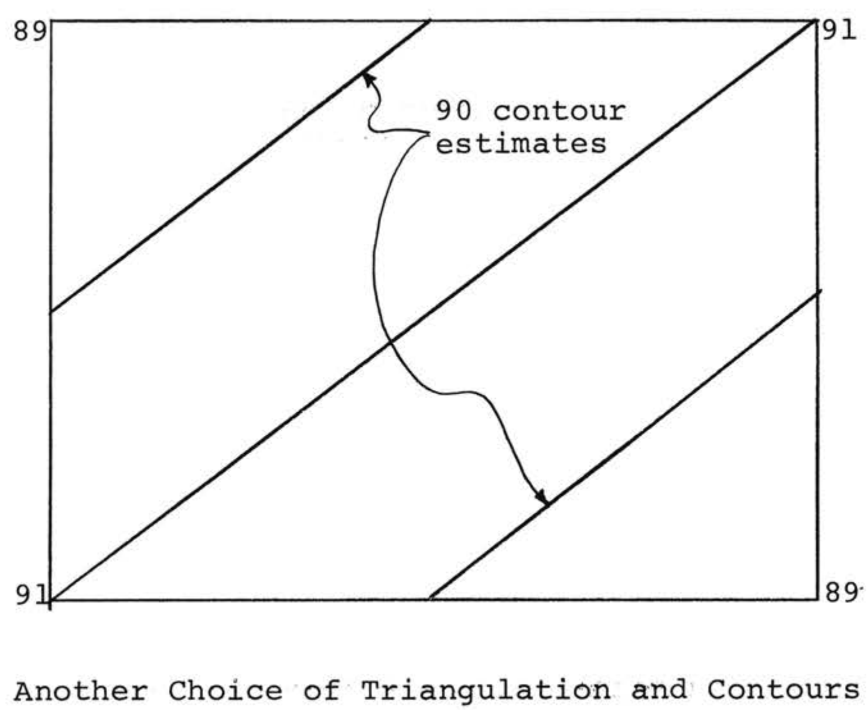
Analysis of the Method

Consider now the reasons why this method of estimation is preferred over those of Chapter II. First, only original data points are used directly in linear interpolation. Second, there is no need for a preanalysis of the data to find a convenient grid size with which to work. The fact that the linear interpolation to the estimates is made directly from the data points contributes to the accuracy of the procedure, and guarantees that the contour estimates lie on the proper side of the data points. Furthermore, no two contours will cross. One last benefit of the method is that it enables one to find an estimate of the accuracy of the contours. This subject will be discussed in Chapter IV.

Of course, there are disadvantages to this method. The use of convex polygons restricts the configuration of triangles in such a way that the ideal configuration may not be found. This represents a compromise in order to keep computer time within reasonable bounds. Also, in the course of deciding which line is to be used in defining a triangle, there may be a situation where two lines have equal merit, and the choice of one line may produce different results than if the other line is selected. See Figure 7.



One Choice of Triangulation and Contours



Another Choice of Triangulation and Contours

Figure 7. Nonunique Contour Estimates

CHAPTER IV

ERROR ANALYSIS

As mentioned in Chapter I, an estimation procedure cannot be considered complete without some analysis of the errors involved. The errors incurred in computer oriented contour estimation procedures such as the ones discussed in this study are these: round-off error, experimental error, lack of fit error, and procedural error. By examining these errors, the quality of both the method of data collection and the estimation procedure may be determined. To the author's knowledge, no error analysis has been made for the grid and quadrilateral methods of Chapter II. This chapter is concerned with error analysis for the estimation procedure developed in Chapter III.

Notation

For ease of reading, a summary of the notation to be used in this chapter is presented in Table I. For a given response surface $z(x,y)$ Table I lists and defines briefly each basic symbol to be used. Figure 8 illustrates these terms for a hypothetical response surface.

TABLE I

NOTATION

Symbol	Definition
p_i	ith datum point in the x-y plane
n	number of data points, $i=1,2,\ldots,n$
r	number of observations at each datum point, $j=1,2,\ldots,r$
z_{ij}	jth observed response at p_i
z_i	observed response at p_i when r=1
$\bar{z}_{i.}$	mean of the r z_{ij}'s at p_i
e_{ij}	experimental error in z_{ij}
s_{iz}^2	variation among z_{ij}'s at p_i
s_z^2	pooled variation of z_{ij}'s
$\hat{z}_i$	calculated response at p_i
z_{iT}	true response at p_i
K	number of independent quadrilaterals formed over data area, $k=1,2,\ldots,K$
$p_{13} \equiv p_{24}$	point where line p_1p_3 intersects the line p_2p_4
$\hat{z}_{13}$	response at p_{13} calculated from $\hat{z}_1$ and $\hat{z}_3$
$\hat{z}_{24}$	response at p_{24} calculated from $\hat{z}_2$ and $\hat{z}_4$
$z_{13T} \equiv z_{24T}$	true response at p_{13} or p_{24}
$\hat{e}_{13}$	error at p_{13} interpolated from e_1 and e_3
$\hat{e}_{24}$	error at p_{24} interpolated from e_2 and e_4
d_{13}	component of error at p_{13}
d_{24}	component of error at p_{24}
$\hat{z}_{13T}$	response at p_{13} calculated from z_{1T} and z_{3T}
$\hat{z}_{24T}$	response at p_{24} calculated from z_{2T} and z_{4T}

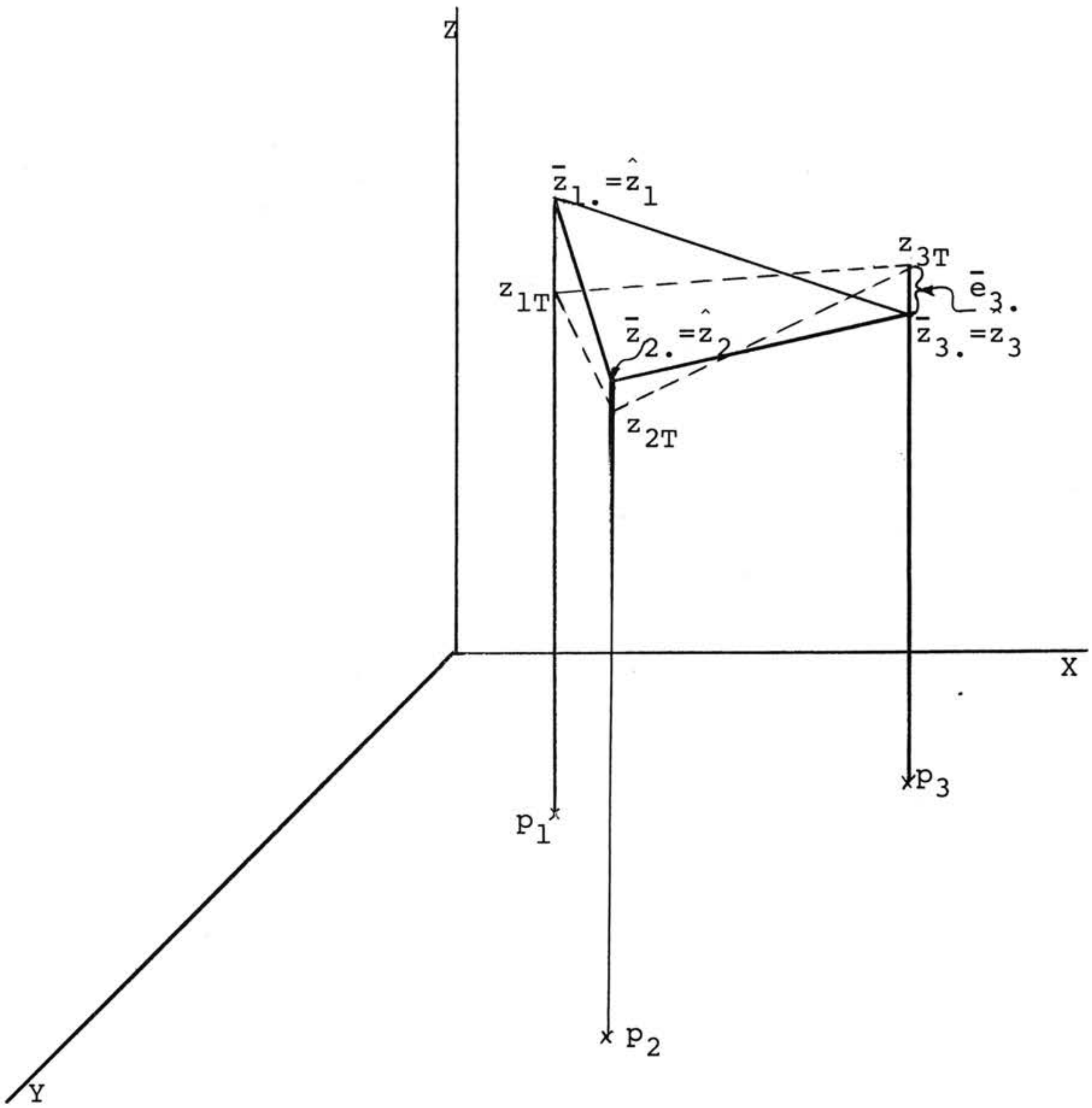


Figure 8. Illustration of Terms

Round-off Error

Some round-off error is inherent in any computer operation. This error may be controlled by efficient programming and careful development of the computing algorithm. The contour estimation method of Chapter III has no complicated formulas and is not iterative. Hence, an efficient program of this method will not produce any significant round-off errors.

Experimental Error

Experimental errors include many types of extraneous variation of which two main sources may be distinguished: (1) inherent variability in the experimental material, and (2) variability caused by lack of uniformity in the physical conduct of the experiment. If the responses are greatly influenced by experimental error, an analysis of this error should give some insight as to the precision of the response estimation. Such an analysis should provide the conditions under which it is possible to estimate the variance of this type of error.

For the contour estimation problem it will be assumed that each observed response z_{ij} is composed of a constant z_{iT}, sometimes known as the true response at the ith point in the x-y plane, plus an experimental error e_{ij} . That is,

$$z_{ij}=z_{iT}+e_{ij} \ , \quad i=1,2,\ldots,n \ , \quad j=1,2,\ldots,r \ .$$

Here, i represents the ith (x,y) position in the coordinate space, and j indexes the replication of data at the ith position. It is further assumed that the e_{ij}'s are normally and independently distributed with mean zero and variance σ^2. Using these assumptions, it follows that $E(z_{ij})=z_{iT}$, and $Var(z_{ij})=\sigma^2$. If there are r observations at a given datum point, it is well known that if

$$\bar{z}_{i.} = \frac{1}{r} \sum_{j=1}^{r} z_{ij} \quad ,$$

$$\text{then } E(\bar{z}_{i.})=z_{iT} \quad ,$$

$$\text{and if } s_{iz}^2 = \frac{1}{r-1} \sum_{j=1}^{r} (z_{ij}-\bar{z}_{i.})^2 \quad ,$$

$$\text{then } E(s_{iz}^2)=\sigma^2 \quad ,$$

the expectation being taken over all possible sets of r observations that can be made at the ith point.

Now, suppose there are r observations at all n data points and that the experimental errors at each of the data points have the same distribution. Then, let

$$s_z^2 = \frac{1}{n} \sum_{i=1}^{n} s_{iz}^2 \quad , \qquad [4.1]$$

from which it follows that $E(s_z^2)=\sigma^2$. s_z^2 is the minimum variance unbiased quadratic estimator for σ^2 (4).

The above formulas provide the methodology for estimating the variance of the experimental errors if the responses at each datum point are at least duplicated. However, there are cases where it is not feasible, and sometimes not possible, to provide replicated observations at a datum point. If no replications exist, the above methods are not applicable. It would be desirable, therefore, to have an estimator for σ^2 which could be used when r=1. The remainder of this section on experimental error will attempt to indicate the conditions under which an estimator for the variance of experimental error is possible.

As a preliminary step to developing a general estimator for σ^2 when r=1, consider the special case in which the mesh points of a rectangular grid system are the data points. Figures 9 and 10 are provided for orientation purposes. Suppose that the n data points determine K independent rectangles, where two rectangles are said to be independent if they have no common vertex. In figure 10, let $p_1p_2p_3p_4$ be the kth such rectangle, and let z_1, z_2, z_3, and z_4 be the responses at the points p_1, p_2, p_3, and p_4, respectively. The point $p_{13} \equiv p_{24}$ is the intersection of the two diagonals p_1p_3 and p_2p_4. Calculate, as indicated, two independent estimates of response, $\hat{z}_{13}$ and $\hat{z}_{24}$, the first associated with the points p_1 and p_3 and the second associated with the points p_2 and p_4, and then consider their difference q_k.

$$\hat{z}_{13} = \frac{1}{2}(z_1+z_3) \text{ and } \hat{z}_{24} = \frac{1}{2}(z_2+z_4) \quad .$$

Since $z_i = z_{iT} + e_i$,

$$\begin{aligned} q_k &= \hat{z}_{13} - \hat{z}_{24} \\ &= [\tfrac{1}{2}(z_{1T}+z_{3T}-z_{2T}-z_{4T})]+[\tfrac{1}{2}(e_1+e_3-e_2-e_4)] \\ &= Q_k + e_k' \quad , \end{aligned}$$

where Q_k is the first term and e_k' is the second term in brackets in the equation.

Let E_1 be the expectation over all possible replications of the experiment which provide responses at these

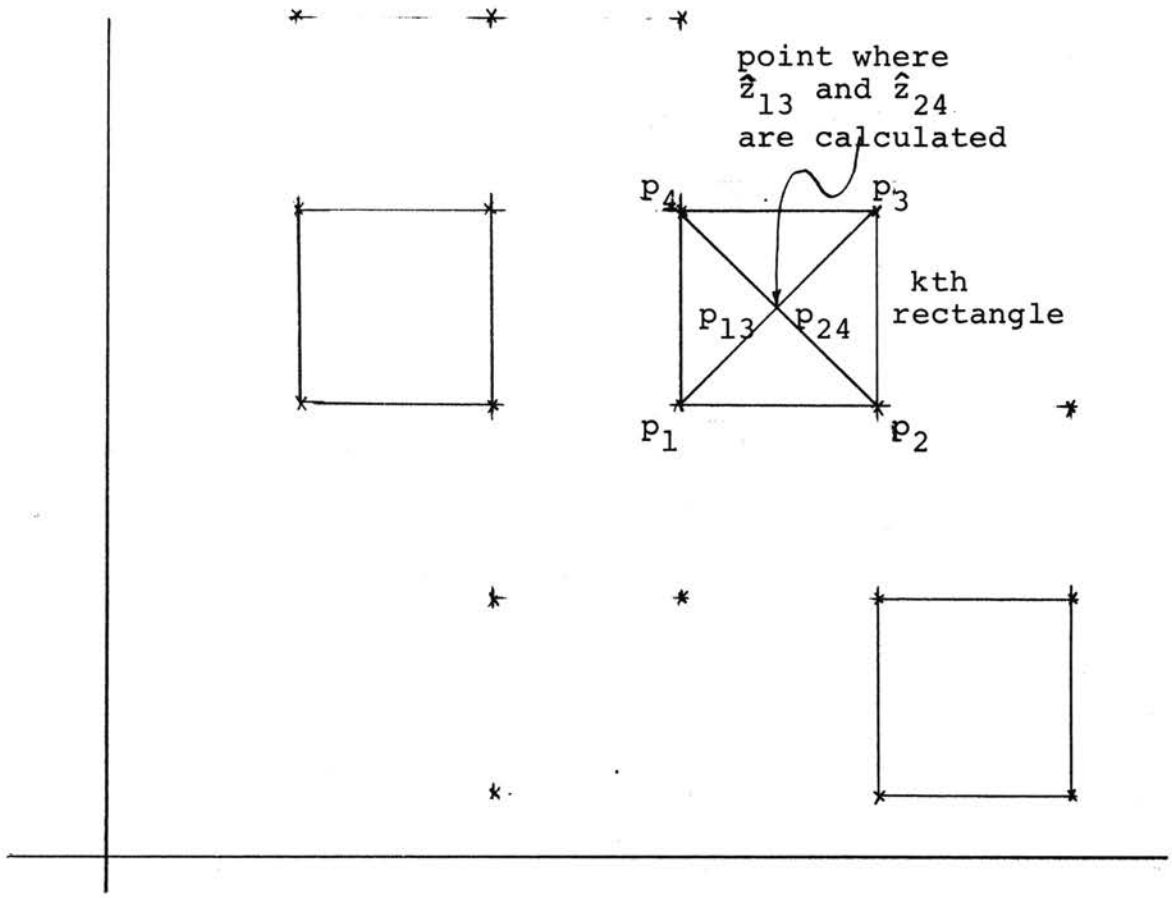


Figure 9. Formation of Quadrilaterals for Regularly Spaced Data

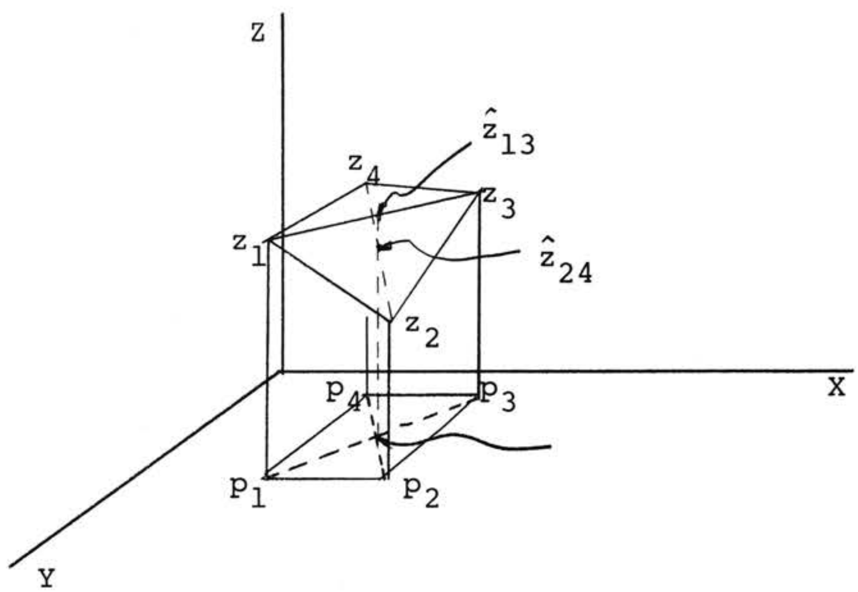


Figure 10. Observed and Estimated Responses for kth Rectangle

same specified mesh points p_1, p_2, p_3, and p_4. Then,

$$E_1(q_k) = \frac{1}{2}(z_{1T}+z_{3T}-z_{2T}-z_{4T})=Q_k \quad ,$$

$$\begin{aligned} \text{and } \operatorname{Var}(q_k) &= \operatorname{Var}(e_k') \\ &= \operatorname{Var}[\tfrac{1}{2}(e_1+e_3-e_2-e_4)] \\ &= \tfrac{1}{4}(4\sigma^2)=\sigma^2 \quad . \end{aligned}$$

The next step is to estimate the variance of q_k which is equal to the σ^2 of experimental error. However, with r=1, there is only one value for q_k, and it is not possible to estimate the variance of q_k from the kth rectangle. It shall be necessary, therefore, to resort to the consideration of the several random variables $q_k=Q_k+e_k'$, k=1,2,...,K. From the experiment there is one value for each of the K random variables q_k. Under what conditions can the values of the K random variables be considered a random sample of size K from a common distribution?

If the "true responses" at the four vertices of the kth rectangle were coplanar, then $Q_k=0$. Moreover, if this were true for each of the K rectangles, then $q_k=e_k'$, and it would be possible to consider the K values as normal deviates with mean zero and variance σ^2. Similarly, if the Q_k's were all equal, the K values could be considered as normal deviates with mean Q and variance σ^2. Under each of these conditions an estimate of the variance of the experimental errors could be obtained by using the standard format

$$s_q^2 = \frac{1}{K-1} \sum_{k=1}^{K} (q_k-\bar{q})^2 \text{ , where } \bar{q} = \frac{1}{K} \sum_{k=1}^{K} q_i \quad .$$

The distribution of the q_k obtained from repeating the experiment by obtaining responses at n mesh points will depend upon the nature of the response surface being fitted.

For surfaces with no interaction, that is, surfaces which may be written in the functional form

$$z(x,y) = f(x)+g(y) \quad ,$$

it is easily shown that $Q_k=0$. Let the coordinates of the vertices of the kth rectangle be as follows: $p_1=(x,y)$, $p_2=(x+\Delta x,y)$, $p_3=(x+\Delta x,y+\Delta y)$, and $p_4=(x,y+\Delta y)$. Then the true responses are these:

$$z_{1T}=f(x)+g(y) \qquad z_{3T}=f(x+\Delta x)+g(y+\Delta y)$$

$$z_{2T}=f(x+\Delta x)+g(y) \qquad z_{4T}=f(x)+g(y+\Delta y) \quad .$$

Now, $Q_k=1/2(z_{1T}+z_{3T}-z_{2T}-z_{4T})$ is obviously zero. Since the vertices of any rectangle will be related in the same manner as above, it may be concluded that $Q_k=0$ for all rectangles when the surface has no interaction.

For surfaces with only linear by linear interaction, Q_k will be constant for a given rectangular grid system. In this case it will be possible to write $z(x,y)=f(x)+g(y)+cxy$. If the coordinates of the vertices of the kth rectangle are the same as in the above paragraph, the true responses may be written as follows:

$$z_{1T}=f(x)+g(y)+cxy$$

$$z_{2T}=f(x+\Delta x)+g(y)+c(x+\Delta x)y$$

$$z_{3T}=f(x+\Delta x)+g(y+\Delta y)+c(x+\Delta x)(y+\Delta y)$$

$$z_{4T}=f(x)+g(y+\Delta y)+cx(y+\Delta y) \quad .$$

Now,

$$Q_k=\tfrac{1}{2}[cxy+c(x+\Delta x)(y+\Delta y)-c(x+\Delta x)y-cx(y+\Delta y)]$$

$$= \tfrac{1}{2}c\Delta x\Delta y=Q \quad .$$

Surfaces with more complicated interactions will not, in general, meet the condition that Q_k be constant for all of the K rectangles. One may, in certain problems where there is no prior information about any interactions, be willing to assume that the conditions on the distribution of the q_k are approximately correct, and that the resulting inflation of the estimate of the variance of the experimental error is negligible relative to the variance of experimental error itself.

To return to the general problem, let the data points be randomly spaced over the x-y plane. Join the data points to form K independent convex quadrilaterals such that no datum point lies within a quadrilateral. With these restrictions, it may not be possible to use all the data points. Number the vertices of the kth quadrilateral as illustrated in Figure 11. Estimate the responses $\hat{z}_{13}$ and $\hat{z}_{24}$ at the point where the line between p_1 and p_3 crosses the line between p_2 and p_4 using the forms

$$\hat{z}_{13}=\alpha z_1+(1-\alpha)z_3 \quad \text{and}$$
$$\hat{z}_{24}=\beta z_2+(1-\beta)z_4 \quad .$$

Then, since $z_i=z_{iT}+e_i$,

$$\hat{z}_{13}=\alpha z_{1T}+(1-\alpha)z_{3T}+\alpha e_1+(1-\alpha)e_3=\hat{z}_{13T}+\hat{e}_{13}$$
$$\hat{z}_{24}=\beta z_{2T}+(1-\beta)z_{4T}+\beta e_2+(1-\beta)e_4=\hat{z}_{24T}+\hat{e}_{24} \quad .$$

The values α and β are weights for linear interpolation. If h_1 is the distance from p_1 to p_{13} and h_3 is the distance from p_3 to p_{13},

$$\text{then} \quad \alpha=\frac{h_3}{h_1+h_3} \text{ . Similarly,} \quad \beta=\frac{h_4}{h_2+h_4} \text{ .}$$

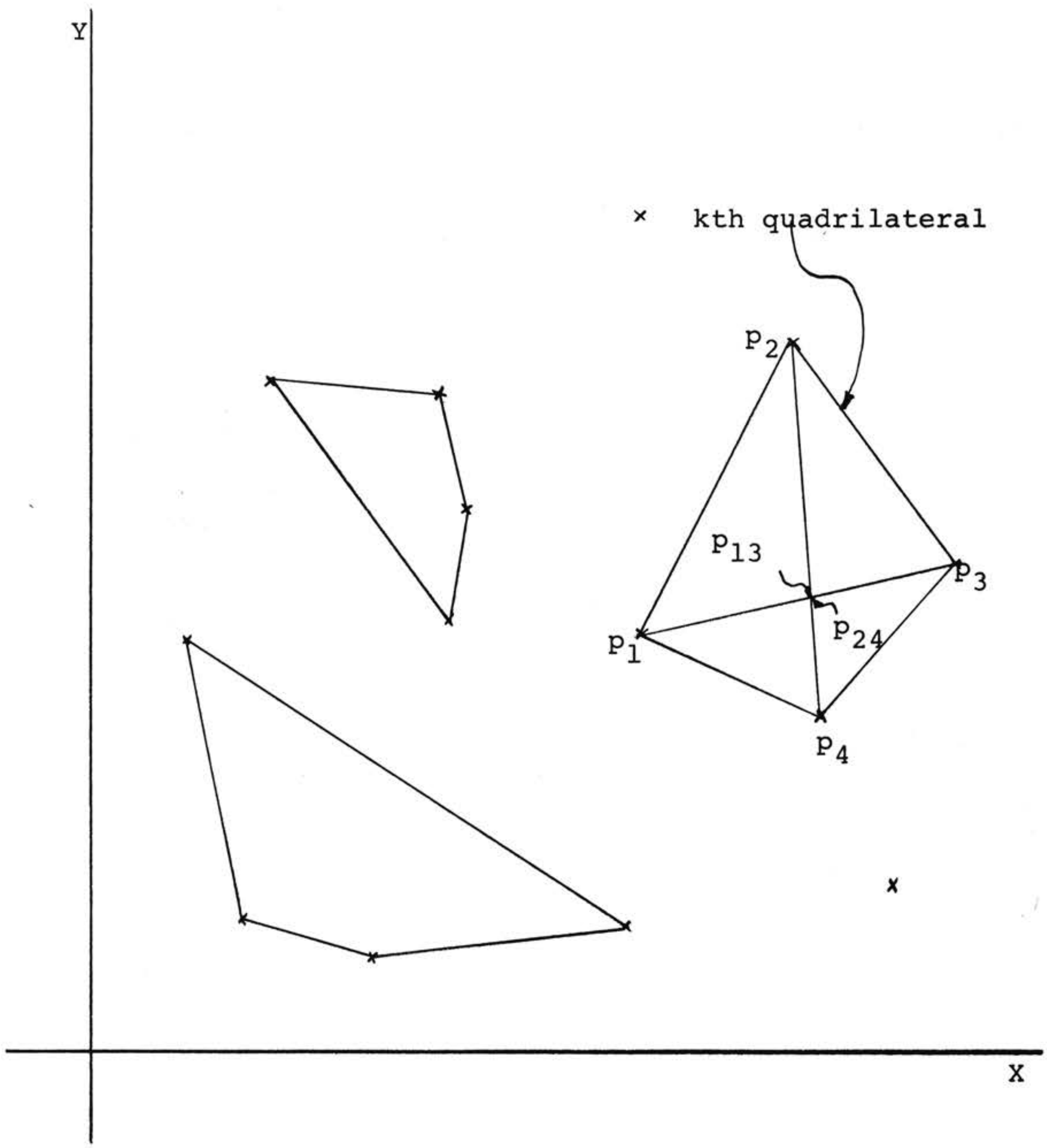


Figure 11. Formation of Quadrilaterals for Randomly Spaced Data

In comparing this general case with the special case previously discussed, one can see that here the α and β are playing the role of the constant factor 1/2 in the special case of regularly spaced data; i.e., in the special case $\alpha=\beta=1/2$. For a fixed or specified quadrilateral, α and β are constants depending only on the position of the data points. However, since the p_i are randomly spaced, the h_i are random variables, and this implies that α and β are random variables.

Again, let

$$\begin{aligned} q_k &= \hat{z}_{13} - \hat{z}_{24} \\ &= \hat{z}_{13T} - \hat{z}_{24T} + \hat{e}_{13} - \hat{e}_{24} \quad , \end{aligned}$$

and consider the conditional expectation and conditional variance of q_k, where the condition is specified as fixing the data points, and where the expectation E_1 is taken over all possible sets of responses at these four points.

$$\begin{aligned} E_1(q_k) &= \hat{z}_{13T} - \hat{z}_{24T} = Q_k \\ \text{Cond. Var}(q_k) &= \text{Cond. Var}[\alpha e_1 + (1-\alpha)e_3 - \beta e_2 - (1-\beta)e_4] \\ &= [\alpha^2 + (1-\alpha)^2 + \beta^2 + (1-\beta)^2]\sigma^2 \quad . \end{aligned}$$

If the experiment is considered to provide K independent quadrilaterals as indicated above, there are then K values of α and β, and upon repeating the experiment one must consider the distribution of α_k and β_k. An empirical study, the results of which are found in Appendix A, has shown that the joint distribution of α_k and β_k is approximately uniform; i.e., $f(\alpha_k, \beta_k)=1$, $0<\alpha_k<1$, $0<\beta_k<1$. Since α_k and β_k are independent, it follows that $E_2(\alpha_k)=E_2(\beta_k)=1/2$. This implies that

$$E_2[\text{Cond. Var}(q_k)] = E_2[\alpha_k^2+(1-\alpha_k)^2+\beta_k^2+(1-\beta_k)^2]\,\sigma^2$$
$$= \frac{4}{3}\sigma^2 \quad ,$$

E_2 being taken over all possible repetitions of the experiment.

Now, the problem of estimating the variance of the experimental error is still difficult due to the lack of sufficient evidence about the values of the q_k for the quadrilaterals. Any study of the nature of the q_k for this case is much more complicated than the study encountered in the special case of regularly spaced data. Certain assumptions can be made, or an empirical study can be performed for specific types of surfaces. However, this thesis includes no special advice about the type of assumptions that should be made, nor does it report on any empirical studies in this area. In short, for randomly spaced data points, the recommendation of this thesis is that some replication of data should be included in the experiment to provide an estimate of the variance of experimental error by conventional methods.

Lack of Fit Error

In addition to experimental error, it is desirable to investigate the failure of the triangulation model to estimate the true response surface. Since the responses at the n data points all lie on the fitted model, there are no observations other than those used in the fitting procedure to measure the failure of the fitted model, unless additional observations can be obtained. These additional observations are either a part of the primary collection of data or are obtained over and above the points deemed necessary for the surface fitting process.

If both experimental error and lack of fit error are to be assessed, one may take repeated observations at each datum point as well as the additional observations at points other than the n data points. These data could be used to estimate both sources of error using standard technique (2). Since replication or additional observation may not be feasible, it would be desirable to investigate these errors when $r=1$. The rest of this section will be devoted to this problem.

To indicate the nature of the problem of estimating lack of fit error when $r=1$, a procedure which could be used for measuring this error if data were available shall be examined. Suppose $p_1p_2p_3$ is one of the triangles produced by the triangulation procedure of Chapter III. Let p_{13} be a point on the side of this triangle determined by p_1 and p_3. Let $\alpha = h_3/(h_1+h_3)$, where h_1 and h_3 are the line segments determined by p_{13}. Figure 12 provides visualization of this situation and the following relations. The estimated response at p_{13} is $\hat{z}_{13}=\alpha z_1+(1-\alpha)z_3$. Suppose that it were possible to measure the responses at p_1 and p_3 without error. In this case, $\hat{z}_{13}=\alpha z_{1T}+(1-\alpha)z_{3T}=z_{13T}$. In addition, suppose that one were able to observe the true response at p_{13}; that is, $z_{13}=z_{13T}$. Then, the quantity $d_{13}=\hat{z}_{13}-z_{13}=\hat{z}_{13T}-z_{13T}$

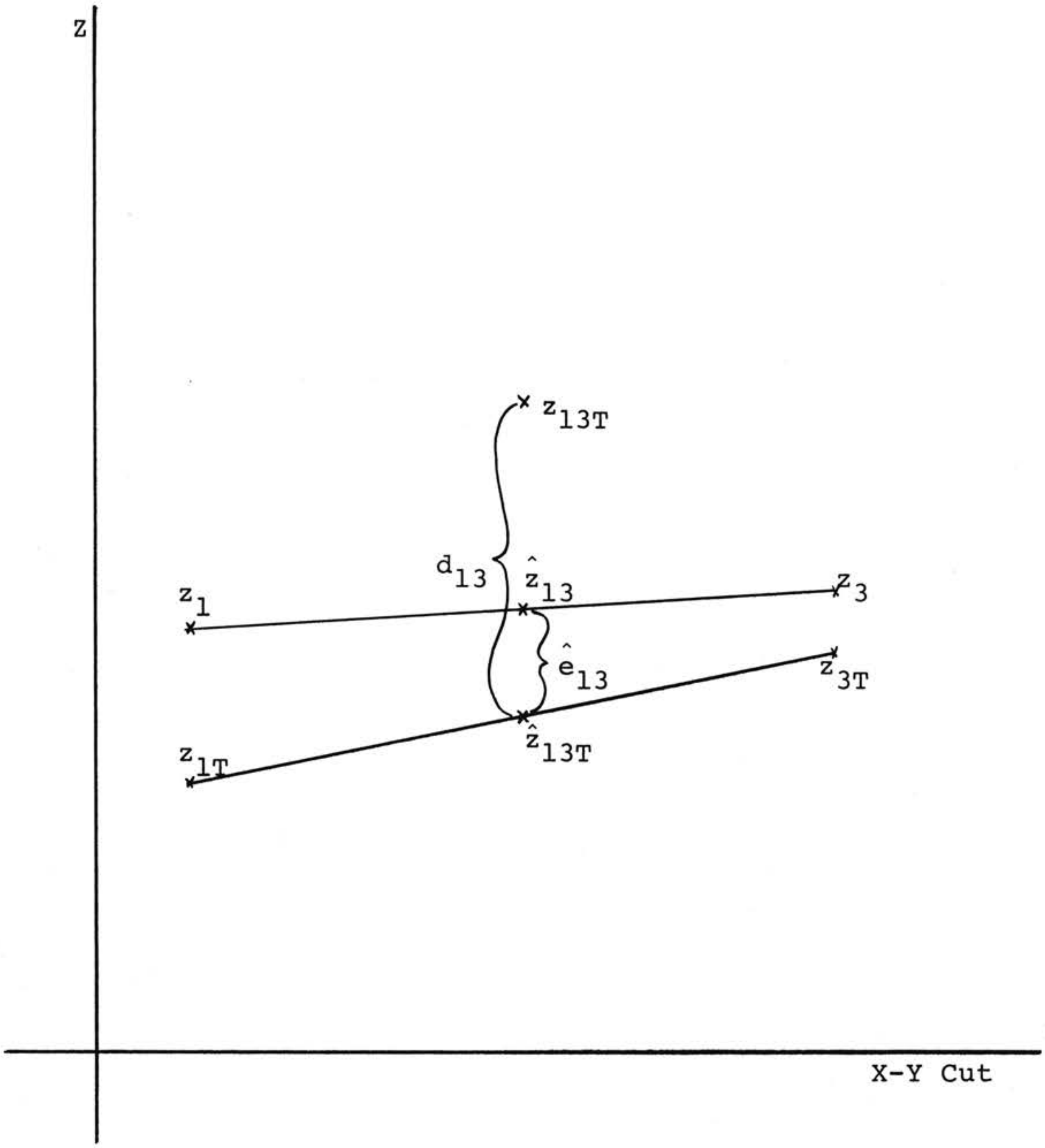


Figure 12. Diagram of Errors as Observed in a Cross Section of the Response Surface

would be a measure of lack of fit error. Of course, it is not possible to measure the responses without error. Rather, one would observe $z_1=z_{1T}+e_1$ and $z_3=z_{3T}+e_3$. The estimated response is then

$$\begin{aligned}\hat{z}_{13}&=\alpha z_1+(1-\alpha)z_3\\&=[\alpha z_{1T}+(1-\alpha)z_{3T}]+[\alpha e_1+(1-\alpha)e_3]\\&=\hat{z}_{13T}+\hat{e}_{13}\quad.\end{aligned}$$

If an observation were available at p_{13}, it would have the form $z_{13}=z_{13T}+e_{13}$. Let

$$\begin{aligned}\hat{d}_{13}&=\hat{z}_{13}-z_{13}\\&=(\hat{z}_{13T}-z_{13T})+(\hat{e}_{13}-e_{13})\quad.\end{aligned}$$

Then, taking the expectation over repetitions of the experiment at these same points,

$$\begin{aligned}E(\hat{d}_{13})&=E[\hat{z}_{13}-z_{13}]\\&=E[(\hat{z}_{13T}-z_{13T})+(\hat{e}_{13}-e_{13})]\\&=\hat{z}_{13T}-z_{13T}=d_{13}\quad,\end{aligned}$$

$$\begin{aligned}\text{and } \operatorname{Var}(\hat{d}_{13})&=\operatorname{Var}[\hat{e}_{13}-e_{13}]\\&=\operatorname{Var}[\alpha e_1+(1-\alpha)e_3-e_{13}]\\&=[\alpha^2+(1-\alpha)^2+1]\sigma^2\quad.\end{aligned}$$

The quantity $\hat{d}_{13}$ is then an estimator for d_{13}, the measure of lack of fit error. Since an observation at p_{13} is not available, it is not possible to estimate d_{13}. However, subsequent results in this chapter require additional knowledge about d_{13}. If the point p_{13} is selected at random from the points on the side p_1p_3 of a triangle which is selected at random, d_{13} may be considered as a random variable. To find the distribution of d_{13}, it is necessary to restrict considerations to response surfaces with the property of equal concavity and convexity. That is, select at random any point on the surface, and slice the surface, in a random direction, with a plane perpendicular to the x-y plane through the selected point. If the function described by the intersection of the plane and the surface has equal probability of being concave or convex at the selected point, the surface is said to be equally concave and convex. An example of such a surface is the hyperbolic paraboloid.

For surfaces with equal concavity and convexity, the results of empirical studies, shown in Appendix A, indicate that it is reasonable to assume that the d_{13}'s are independently normally distributed with mean zero and variance σ_d^2. The results which follow in this chapter make use of this assumption; any application of the results may strictly be made, therefore, only to surfaces of equal concavity and convexity.

Procedural Error

The procedural error is the failure of the triangulation model to give a unique result in fitting the response surface. It is obvious that, for a given set of n points, there are many possible triangulations, and each triangulation might result in a different fitted surface. The variation among these different fitted surfaces will be referred to as procedural variation. The program for the method of Chapter III will provide the same triangulation if it is applied repeatedly to the same set of data points because the computer is instructed to begin forming the first triangle at a specific point. If the computer had been instructed to begin triangulation at random, each repetition of the program on the same data could, conceivably, provide a different triangulation. The program was written to start at a specific point in an attempt to minimize the lengths of the sides of the triangles which are used in linear interpolation. In addition, the amount of computer time needed to permit the computer to begin triangulation at random is so large that the method would be impractical.

In the following section on interval estimation it will be necessary to consider procedural variation. At a given point in the x-y plane, let $\hat{z}'$ be an estimate of the response surface obtained from applying the method to one triangulation, and let $\hat{z}''$ be an estimated response obtained from using a second triangulation. The difference in these estimated responses, $q=\hat{z}'-\hat{z}''$, represents procedural variation at the given point.

Confidence Intervals

Return now to the situation described in the section on experimental error where the sample data points are randomly spaced over the x-y plane, and from K independent convex quadrilaterals as before. For the kth quadrilateral, $\hat{z}_{13k}$ and $\hat{z}_{24k}$ are estimates of response at the crossing point of the diagonals of the quadrilateral, $\hat{z}_{13k}$ being obtained by linear interpolation between the responses at the vertices p_1 and p_3 and $\hat{z}_{24}$ being obtained in a similar manner from the responses at p_2 and p_4. This means, essentially, that $\hat{z}_{13k}$ is an estimate of the response obtained from one triangulation, and that $\hat{z}_{24k}$ is an estimated response obtained from a second triangulation. The difference, $q_k=\hat{z}_{13k}-\hat{z}_{24k}$, represents procedural variation at the point $p_{13}\equiv p_{24}$.

The experiment will yield one value of q_k for each of the K quadrilaterals. Consider now the statistic obtained by performing a sum of squares of these K values of q_k. That is, let

$$H=\frac{1}{K}\sum_{k=1}^{K}(\hat{z}_{13k}-\hat{z}_{24k})^2=\frac{1}{K}\sum_{k=1}^{K}q_k^2 \quad .$$

If the mean of the q_k is assumed to be zero, and if the distribution of the q_k is assumed to be the same for each of the K quadrilaterals, then the statistic H is an estimator for the variance of the q_k.

Empirical studies could be made to determine the types of surfaces for which the above assumptions are true. This thesis does not provide any such studies, nor does it attempt to state conditions under which the assumptions may be approximately correct.

Supposing that these assumptions are correct, it would be desirable to obtain the expected value of H over replications of the experiment for all possible sets of K convex quadrilaterals.

From the section on lack of fit error,

$$\hat{z}_{13k}=\hat{z}_{13T}+\hat{e}_{13} \quad .$$

Add and subtract z_{13T} from the right hand side of the equation. Then

$$\hat{z}_{13k}=z_{13T}+(\hat{z}_{13T}-z_{13T})+\hat{e}_{13}$$
$$=z_{13T}+d_{13}+\hat{e}_{13} \quad .$$

Similarly, $\hat{z}_{24k}$ may be written as

$$\hat{z}_{24k}=z_{24T}+d_{24}+\hat{e}_{24} \quad .$$

Assume that the d_{ij}'s are independently normally distributed with mean zero and variance σ_d^2, and that the e_{ij}'s are distributed independently of the d_{ij}'s as normal variates with mean zero and variance σ^2. Under the condition that the n data points are fixed,

$$\begin{aligned}\text{Cond. Var}(q_k)&=\text{Cond. Var}(\hat{z}_{13k}-\hat{z}_{24k})\\ &=\text{Cond. Var}[(z_{13T}-z_{24T})+(d_{13}-d_{24})+(\hat{e}_{13}-\hat{e}_{24})]\\ &=\text{Cond. Var}[(d_{13}-d_{24})+(\hat{e}_{13}-\hat{e}_{24})]\\ &=2\sigma_d^2+\sigma^2[\alpha_k^2+(1-\alpha_k)^2+\beta_k^2+(1-\beta_k)^2] \quad .\end{aligned}$$

If the random variables α_k and β_k are distributed jointly and independently as uniform variates on the unit square,

$$E_2(H)=E_2[\text{Cond. Var}(q_k)]=2\sigma_d^2+\tfrac{4}{3}\sigma^2 \quad ,$$

where E_2 is taken over repetitions of the experiment.

Under the above assumptions about the distribution of the random variables d_{ij}, e_{ij}, α_k, and β_k, the statistic

$$\frac{KH}{2\sigma_d^2+\frac{4}{3}\sigma^2}$$

has a Chi-square distribution with K degrees of freedom. A data point cannot be used in more than one quadrilateral if the terms in H are to be independent. The above statistic is an approximate χ^2 if data points are used in more than one quadrilateral.

Using this result, confidence intervals may be set on the true response at any point on an estimated contour. Since the response model for a point on a side of a triangle which lies on an estimated contour is

$$\hat{z}_{13}=z_{13T}+d_{13}+\hat{e}_{13} \quad ,$$

then

$$\frac{\hat{z}_{13}-z_{13T}}{\sqrt{\sigma_d^2+\frac{2}{3}\sigma^2}} \sim N(0,1) \quad .$$

Therefore, the quantity

$$\frac{\hat{z}_{13}-z_{13T}}{\sqrt{H/2}}$$

is a t distributed variable with K degrees of freedom. Knowing this distribution one may perform a t-test or calculate a confidence interval as follows:

$$P[\hat{z}_{13}-t_{\delta/2}(K)\sqrt{H/2} < z_{13T} < \hat{z}_{13}+t_{\delta/2}(K)\sqrt{H/2}]=1-\delta \quad .$$

The interval

$$(\hat{z}_{13}-t_{\delta/2}(K)\sqrt{H/2} \ , \ \hat{z}_{13}+t_{\delta/2}(K)\sqrt{H/2}) \qquad [4.2]$$

is a δ confidence interval about z_{13T}.

If replications exist, the variance of $\hat{e}_{13}$ changes. However, the resulting confidence interval is the same since the correction is made in both the normal and the Chi-square variables.

For a point of a contour which does not lie between two data points another confidence interval is necessary. Though the general equation for estimating the response at such a point involves four data points, the response is actually calculated only from the three points making up the triangle around the point. Therefore two of the points in the equation are equal, so that if $p_1=p_4$, say, then

$$\hat{z}_{1234}=\gamma(\alpha z_1+(1-\alpha)z_3)+(1-\gamma)(\beta z_2+(1-\beta)z_4)$$
$$=z_{1234T}+d_{1234}+\gamma\alpha e_1+\gamma(1-\alpha)e_3+(1-\gamma)\beta e_2+(1-\gamma)(1-\beta)e_4.$$
$$E(\hat{z}_{1234})=z_{1234T} \quad .$$
$$\mathrm{Var}(\hat{z}_{1234})=E[d_{1234}+\gamma\alpha e_1+\gamma(1-\alpha)e_3+(1-\gamma)\beta e_2+(1-\gamma)(1-\beta)e_4]^2$$
$$=\sigma_d^2+\sigma^2[\tfrac{1}{9}+\tfrac{1}{9}+\tfrac{1}{9}+\tfrac{1}{9}+2(\tfrac{1}{4}+\tfrac{1}{12}-\tfrac{1}{6}-\tfrac{1}{8})]$$
$$=\sigma_d^2+\sigma^2(\tfrac{19}{36}) \quad . \qquad [4.3]$$

The substitution of 2/3 for 19/36 in equation [4.3] will give an approximate confidence interval which will then be the same as that calculated in equation [4.2].

Empirical tests on the validity of the confidence intervals are found in Appendix C.

CHAPTER V

SUMMARY AND CONCLUSIONS

Nonlinear precise-data-fit techniques constitute one of the basic ways of approaching the problem of contour mapping of response surfaces. The two existing precise-data-fit methods have undesirable properties which, for the most part, result because the methods require the use of a rectangular grid system. To remedy this situation, a new estimation method which does not use a grid system has been developed.

A computer program has been written to implement the triangular contour estimation method described in Chapter III. This program will also perform the error analysis of Chapter IV. For easy conversion between machines, the program was written entirely in Fortran IV. Two existing versions of the program allow operation on either the IBM 7040 or the IBM System 360/40G. Appendix E provides the information necessary for the operation of these programs.

Results obtained by using the triangular method may be seen in the four plots displayed in Appendix B. The plots are contour maps of two representative surfaces: a circular paraboloid and a hyperbolic paraboloid. The data used in constructing each plot were sampled from the known function with no experimental error in the responses. For each function, one plot was constructed from responses at data points selected on a rectangular grid, whereas the other plot was constructed from responses at data points which were randomly spaced.

To support the triangular estimation method of Chapter III, an error analysis was performed. In experiments with replication it was noted that s_z^2 (equation 4.1) is the minimum variance unbiased estimator of the variance of experimental error. In the case of regularly spaced data when there is no replication, it has been shown that the variance of experimental error may be estimated by conventional methods for surfaces with no interaction of order higher than linear by linear, if the researcher is willing to assume that other errors in an estimated response are negligible relative to experimental error.

For surfaces of equal concavity and convexity it was possible to develop confidence intervals on the value of the response at any point on a computed contour. To test the validity of this error analysis, empirical tests were conducted on the confidence intervals. The results of this testing are shown in Appendix C. In a series of experiments with a hyperbolic paraboloid, 95% confidence intervals on the response were computed at various points on the contours. The average number of points which fell within the appropriate confidence intervals was about 90%. The failure of this percentage to be nearer 95% can be explained as the accumulated results of the assumptions made in developing the t variable used in compilation of the confidence intervals. That is, as shown in Appendix A, the d_{ij}'s are not distributed exactly as normal variables, and they are not independent since the correlation between d_{13} and d_{24} is approximately 0.1. Confidence intervals were also computed for functions which are not of equal concavity and convexity. For 95% confidence intervals, the average number of points which fell within the confidence interval on the response was approximately 82%. The failure of this figure to be 95% is largely due to the fact that the surfaces are not equally concave and convex. From these results, one may conclude that in any application of the confidence intervals to surfaces of equal concavity and convexity, the confidence levels are somewhat inflated. For other surfaces, the confidence levels may be severely inflated. However, from the results in Appendix C, it is evident that the intervals are at least an "educated guess" and should not be completely discounted.

Before the triangular procedure of Chapter III is accepted in preference to the older methods of contour estimation, an investigation and comparison is in order. The triangular method represents the response surface by a series of plane segments which have the property that the responses at data points are fitted exactly. This means that the surface is as free as an n-termed polynomial in fitting the data and has the added advantage of being a simple surface between data points. From this point of view the triangular method is superior to least squares polynomial fitting techniques.

To aid in comparing the triangular method with the grid and quadrilateral methods, each method was applied to the same set of semiregularly spaced data. The results of each method are plotted in Appendix D. For the grid method plot, note that about half of the 403 responses fall on each side of the 400 contour. In the quadrilateral method plot there is one point with a 403 response which does not lie between the 400 and 450 contours. In both cases the failure of the points to be on the proper side of contours is the result of averaging datum point responses for use in linear interpolation. In equation [2.1] it can be seen that if all four data points lie on one side of the mesh point, the average may be biased. For the triangular method all responses lie on the expected side of the contours. Consider now regions A and B on each plot. In these regions contours for the grid method curve in a manner which is unexplained by the data. These deviations are caused by deriving responses for linear interpolation in areas where no observed responses are available. The quadrilateral method does not produce contour estimates in the regions A and B. The algorithm for constructing quadrilaterals fails to include one datum point in each of these regions. Thus, the failure of the method to produce contours can be attributed to loss of information. In the triangular method plot there is one irregularity in the 450 and 500 contours which is caused by poor selection of triangles.

When the data points are irregularly or randomly spaced, the contours produced by the grid and quadrilateral methods will contain more regions similar to regions A and B of Figures 21 and 22. There will also be more points which do not lie on the proper side of the contours. Krumbein(7) states, in essence, that a set of n regularly spaced data points contains more information than a set of n randomly spaced points. This follows because of the tendency of randomly spaced points to form clusters which are nearer to replication than additional information. It is obvious from the figures of Appendix B that the triangular method will suffer from lack of information along with the other methods when the data points are not regularly spaced, but the triangular method does not discard any information or attempt to create information where none exists. The contour estimates of the triangular method are an accurate representation of the data and do not give rise to the irregularities which are possible in the grid and quadrilateral methods.

BIBLIOGRAPHY

(1) Batcha, J. P. and Reese, J. R. "Surface Determination and Automatic Contouring for Mineral Exploration, Extraction, and Processing." Quarterly of the Colorado School of Mines, Vol. 59. (October, 1964), 1-14.

(2) Cochran, William G. and Cox, Gertrude M. Experimental Designs. New York: John Wiley, 1950, 352.

(3) Conte, S. D. Elementary Numerical Analysis. New York: McGraw-Hill, 1965, 75-76.

(4) Draper, N. R. and Smith, H. Applied Regression Analysis. New York: John Wiley, 1966, 17-32.

(5) Hildebrand, F. B. Introduction to Numerical Analysis. New York: McGraw-Hill, 1956, 258.

(6) Karlin, Samuel. Mathematical Methods and Theory in Games, Programming, and Economics. London: Addison-Wesley, 1959, 399-400.

(7) Krumbein, W. C. "Trend Surface Analysis of Contour-type Maps with Irregular Control-point Spacing." Journal of Geophysical Research, Vol. 64. (July, 1959), 823-834.

(8) Moise, Edwin E. Elementary Geometry from an Advanced Standpoint. London: Addison-Wesley, 1963, 324-350.

(9) Salvadori, Mario G. and Baron, Melvin L. Numerical Methods in Engineering. Englewood Cliffs, N. J.: Prentice-Hall, 1964, 87-89.

(10) 1620 Numerical Surface Techniques and Contour Map Plotting, White Plains, N. Y.: IBM Publications.

APPENDIX A

EMPIRICAL STUDY

To find the nature of certain unknown distributions, some testing was done using a computer. The results of this testing can be found following a short discussion.

The first experiment was performed to study the distribution of the random variables α and β defined in Chapter IV. To do this, four points were chosen at random in the unit square on the x-y plane. That is, the eight coordinates of the four points were independently uniformly distributed on the interval $[0,1)$. If the points did not define the four vertices of a convex quadrilateral, the set was discarded. Of the five thousand sets of four points used in the testing, about 3400 sets formed convex quadrilaterals. For each convex quadrilateral, corresponding values of α and β were calculated. The histograms obtained for α and β are shown in Figures 13 and 14.

The assumptions about d_{13} and d_{24} were also tested in the experiment above. To do this, a hyperbolic paraboloid was assumed as a true response surface over the unit square in the x-y plane. For each convex quadrilateral, the response surface was evaluated at the four vertices and at the point where the diagonals of the quadrilateral crossed.

Estimated responses at the crossing point of the diagonals were then computed by linear interpolation between the two pairs of opposite vertices. It was possible to obtain the true values of d_{13} and d_{24} since the response surface was known. Histograms of the results are shown in Figures 15 and 16.

$E(\alpha)=0.50217$
$Var(\alpha)=0.07228$
$\rho(\alpha, \beta)=-0.01202$

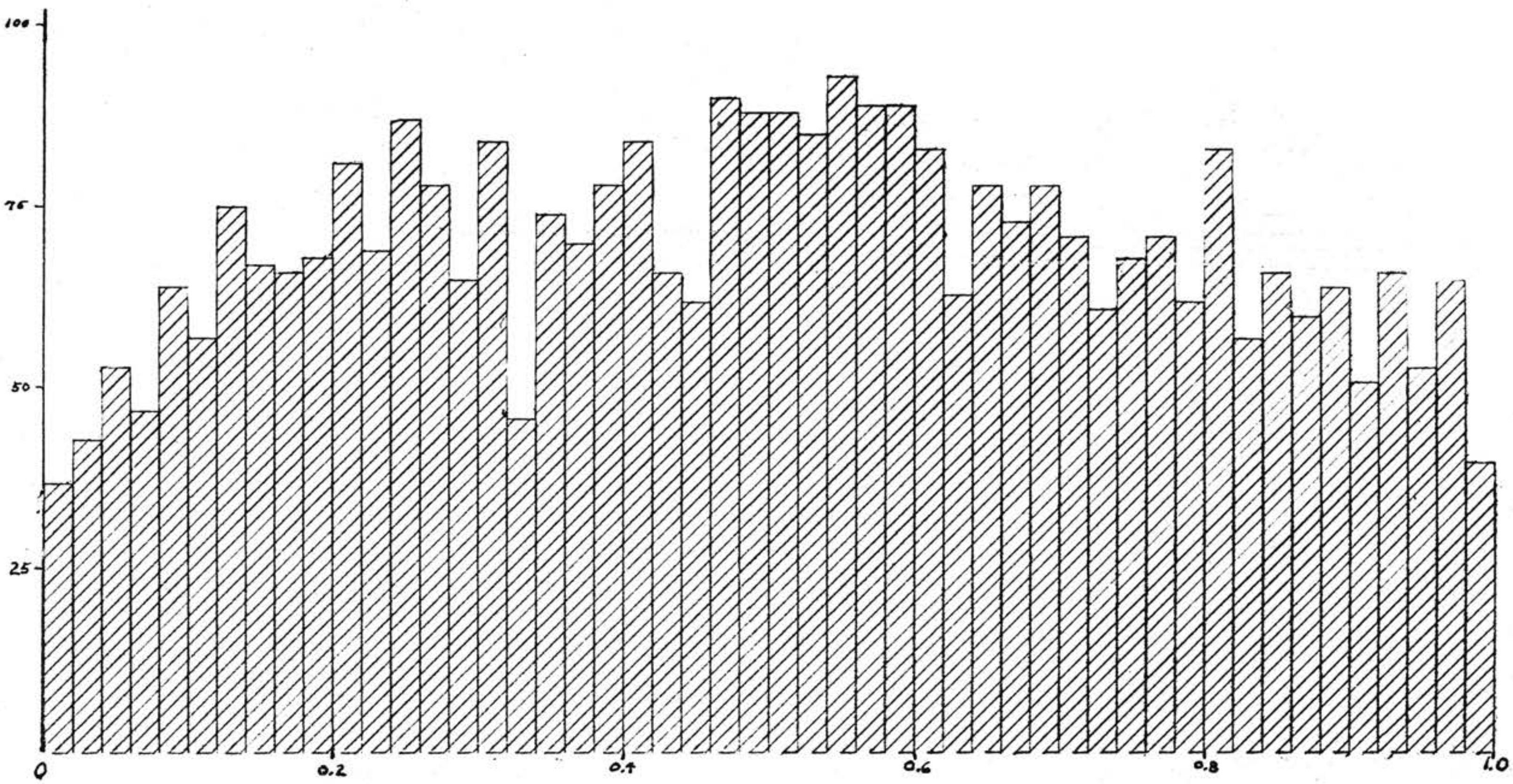


Figure 13. Histogram for α

$E(\beta)=0.50329$
$Var(\beta)=0.07091$
$\rho(\beta,\alpha)=-0.01202$

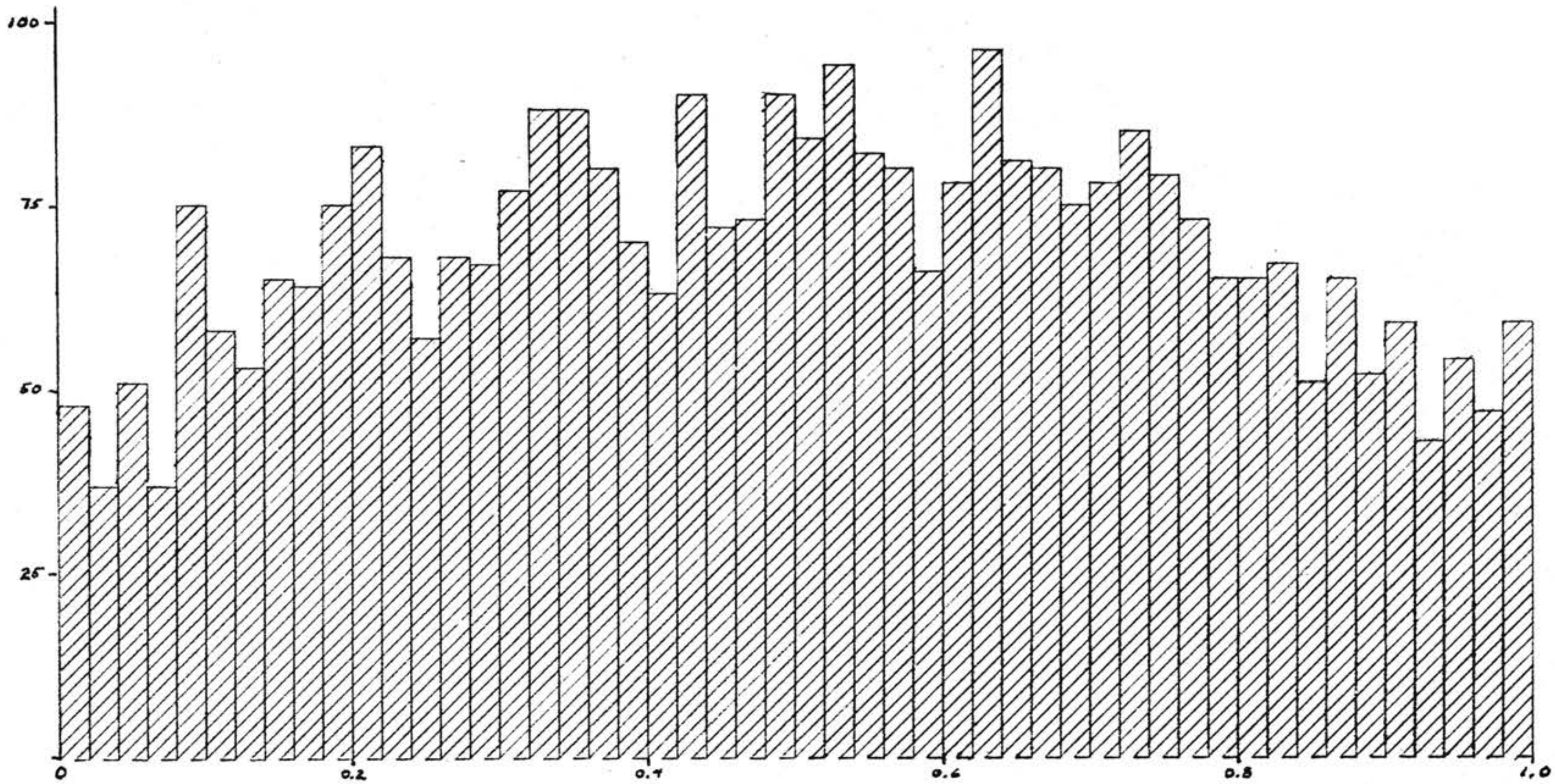


Figure 14. Histogram for β

$E(d_{13})=-0.00129$
$Var(d_{13})=0.00428$
$\rho(d_{13},d_{24})=-0.16539$

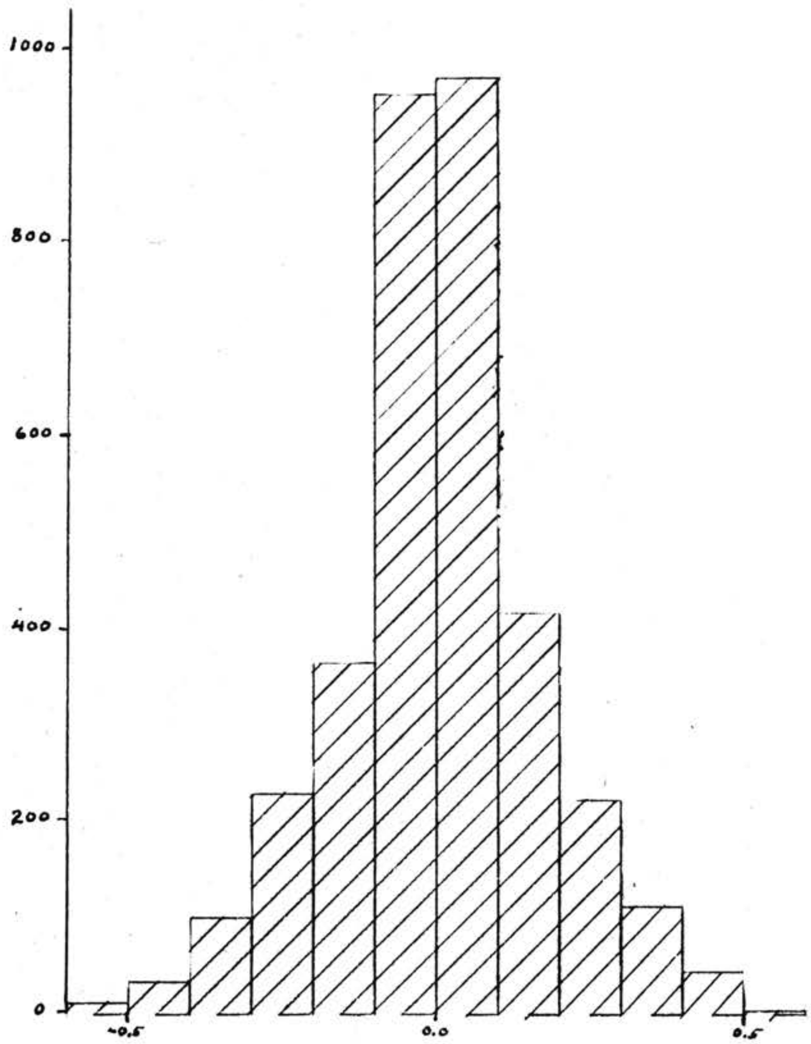


Figure 15. Histogram for d_{13}

$E(d_{24})=0.00148$
$Var(d_{24})=0.00434$
$\rho(d_{24},d_{13})=-0.16539$

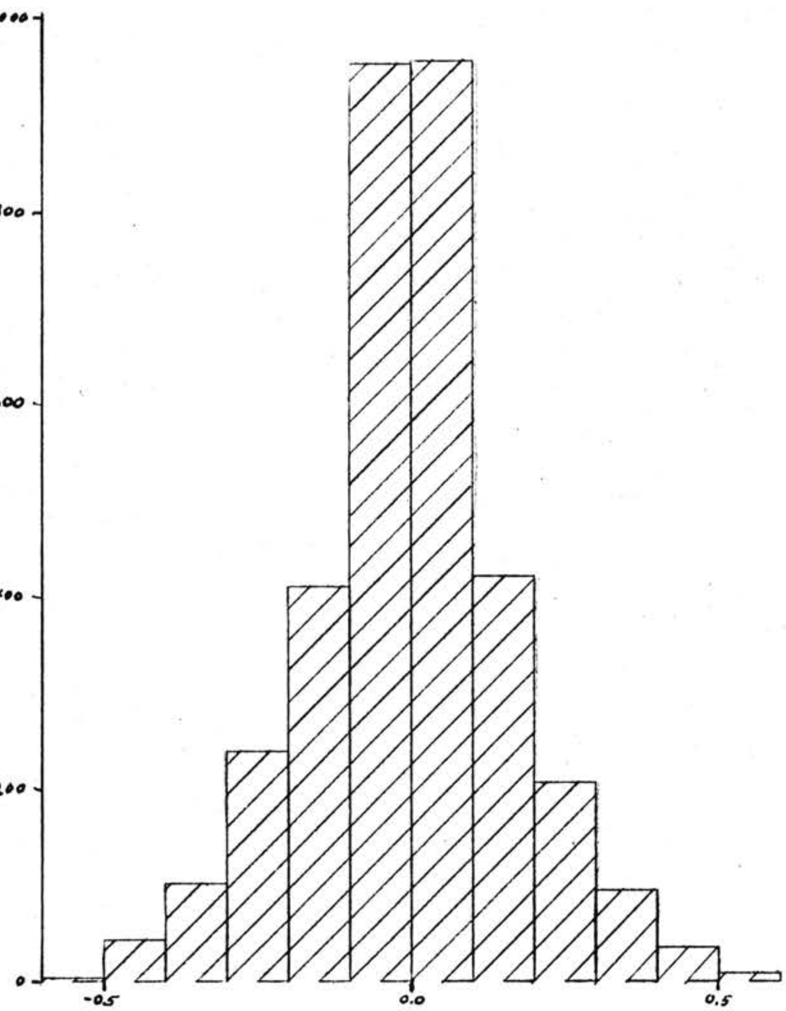


Figure 16. Histogram for d_{24}

APPENDIX B

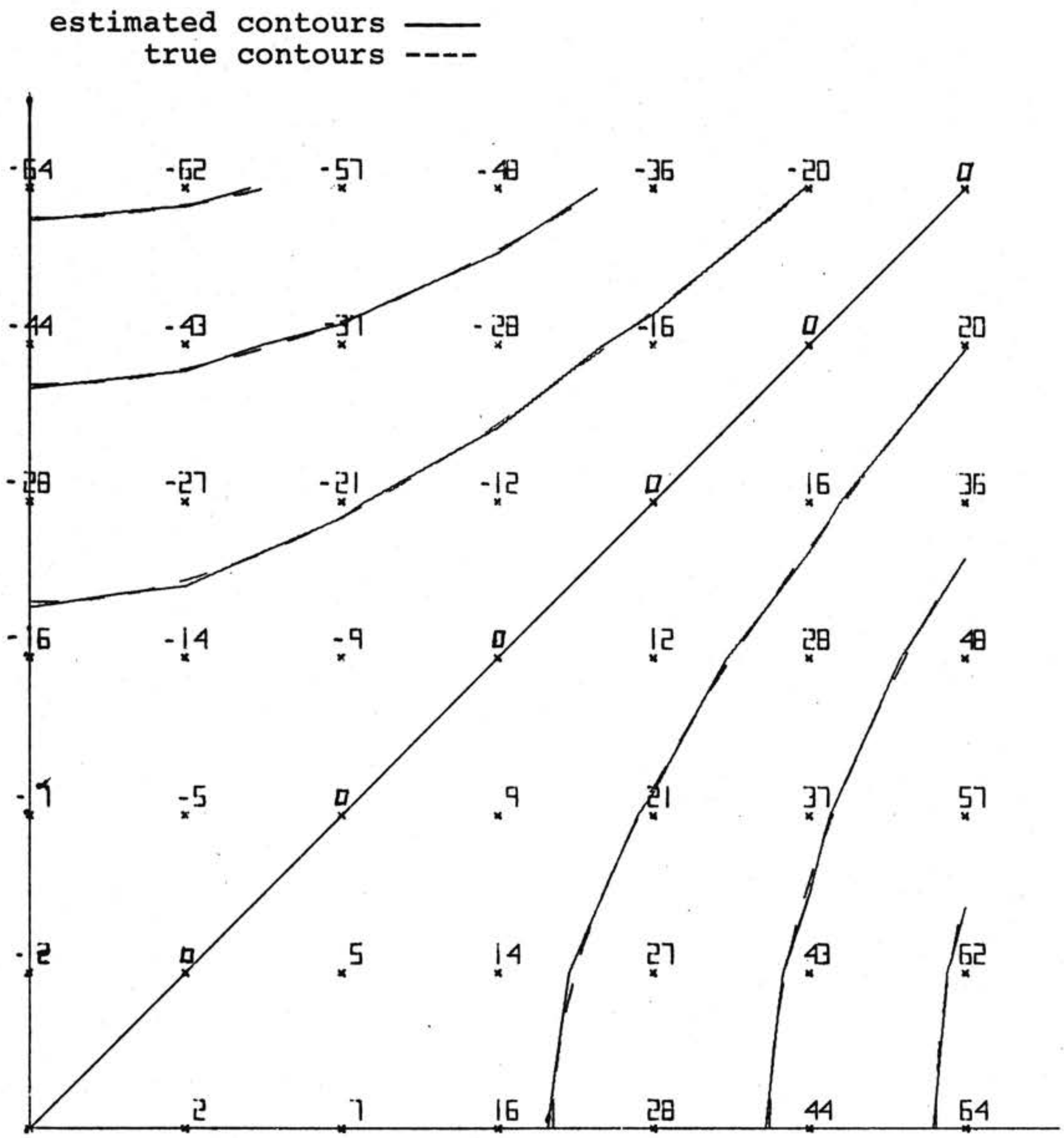


Figure 17. Triangular Contour Estimation for Hyperbolic Paraboloid With Regularly Spaced Data

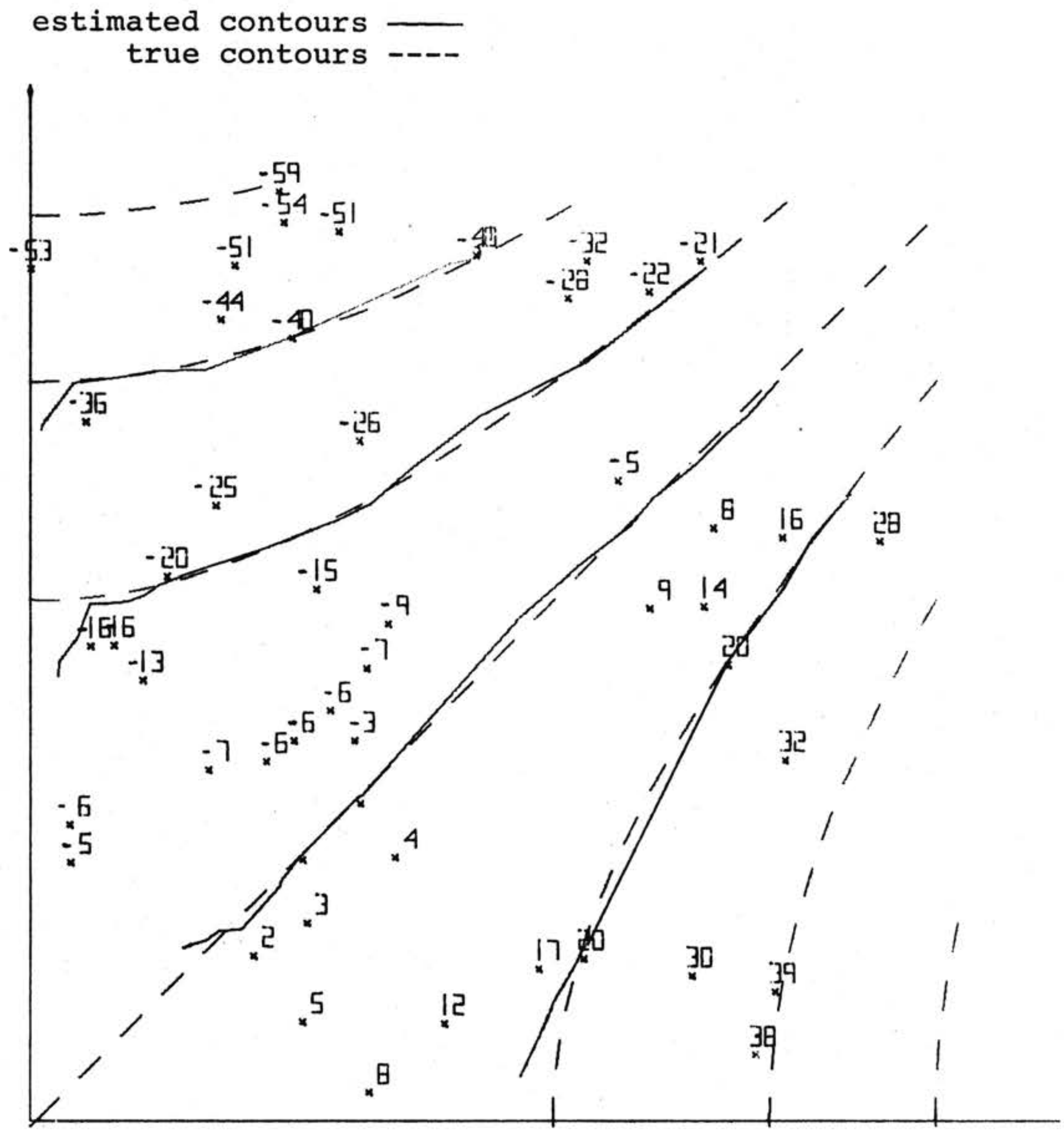


Figure 18. Triangular Contour Estimation for Hyperbolic Paraboloid With Randomly Spaced Data

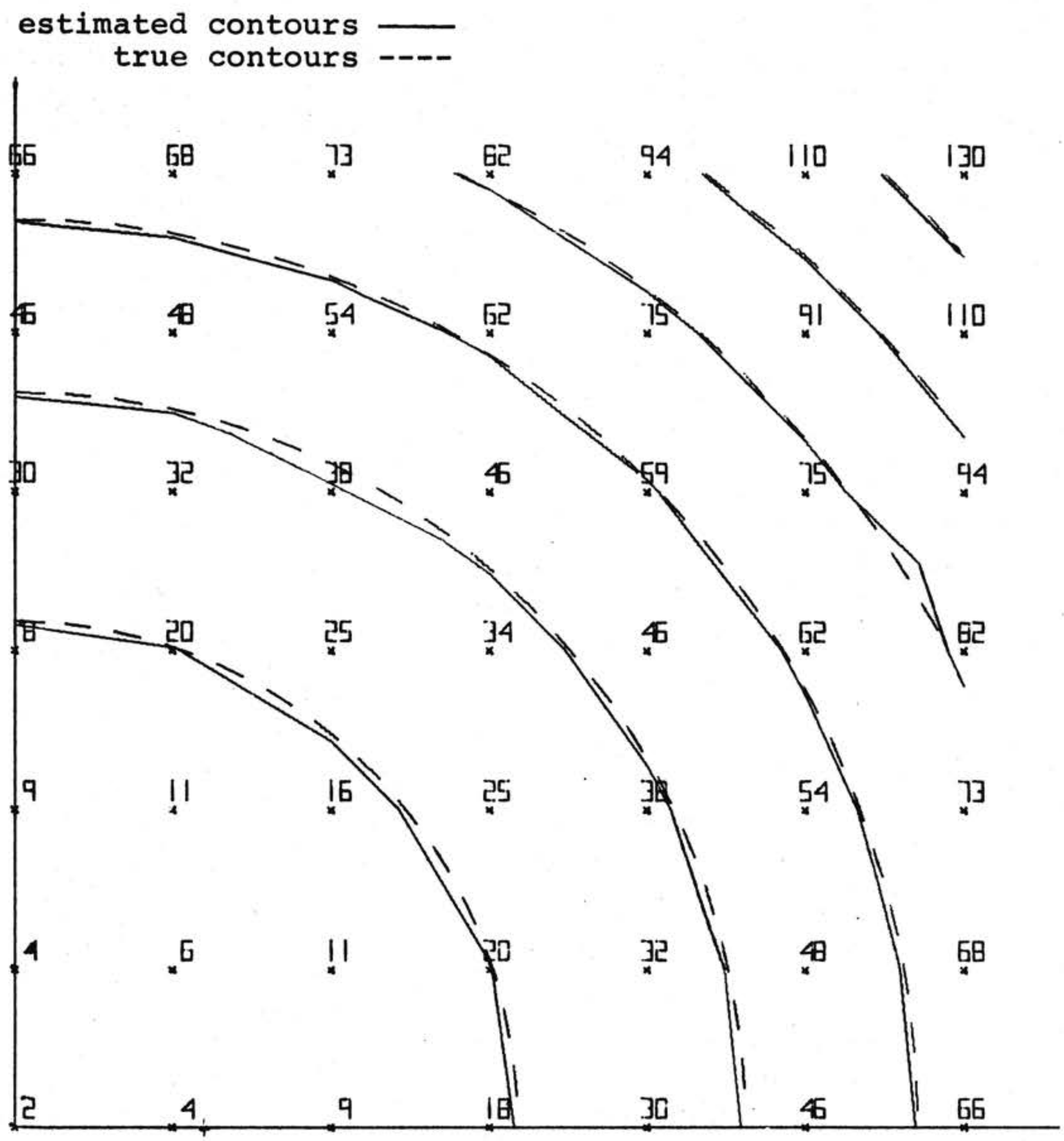


Figure 19. Triangular Contour Estimation for Circular Paraboloid With Regularly Spaced Data

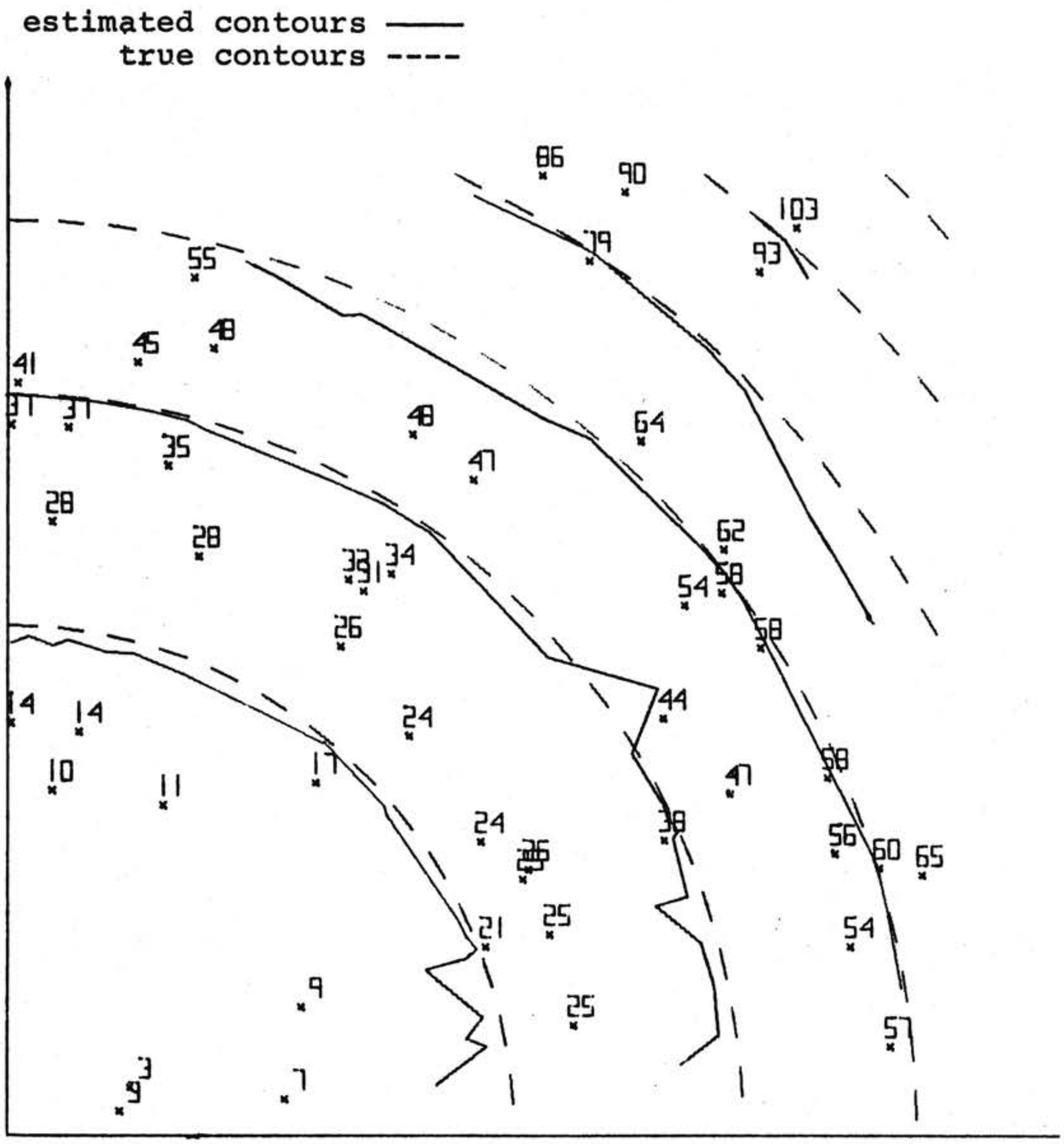


Figure 20. Triangular Contour Estimation for Circular Paraboloid With Randomly Spaced Data

APPENDIX C

EMPIRICAL STUDY OF CONFIDENCE INTERVALS

To test the validity of the confidence intervals developed in Chapter IV, an error analysis was performed on sample data selected from several known surfaces. For each of six response models, ten samples of size 50 were drawn at random from the region $R=\{(x,y)\,|\,0<x<8,\ 0<y<8\}$. The response models used were these:

$$z_i=x_i^2-y_i^2 \qquad z_i=x_i^2-y_i^2+e_i$$
$$z_i=x_i^2+y_i^2 \qquad z_i=x_i^2+y_i^2+e_i$$
$$z_i=x_i^2+8y_i^2 \qquad z_i=x_i^2+8y_i^2+e_i$$

For the models with experimental error, $e_i \sim NID(0,1/4)$.

Contours were estimated for every twenty units in the response variable z. For example, in the hyperbolic paraboloid $z=x^2-y^2$, the contours of z= -60, -40, -20, 0, 20, 40, and 60 were estimated. Confidence intervals were computed on the true response at the end points and midpoints of the broken line segments making up an estimated contour. The true response at each of the above points was calculated, and a check was made to see whether or not this response fell inside the computed confidence interval. The results of this check are tabulated for each of the six models in Tables II, III, and IV.

The four columns of each table are described as follows:

Col. 1 K, the degrees of freedom of H

Col. 2 Mean square H defined in Chapter IV.

Col. 3 % of true responses at end points which fell inside the confidence interval.

Col. 4 % of true responses at midpoints which fell inside the confidence interval.

TABLE II

RESULTS OF CONFIDENCE INTERVAL TESTING FOR THE HYPERBOLIC PARABOLOID, $z=x^2-y^2$

Without Experimental Error				With Experimental Error			
K	H	End Point %	Mid-Point %	K	H	End Point %	Mid-Point %
34	2.14101	90	91	34	2.77206	79	78
39	4.07525	88	91	34	1.95098	91	93
36	5.16563	91	91	33	5.15863	83	88
44	3.04004	98	98	41	5.48978	88	87
33	3.99320	97	97	35	4.57380	94	96
36	2.67606	82	84	28	2.65088	91	90
35	5.00882	79	80	36	2.93224	98	95
36	4.07378	94	94	31	4.32608	96	99
37	4.40000	82	78	38	2.25550	91	86
32	3.60938	91	90	36	6.57157	93	93
Average %		89.2	89.4	Average %		90.4	90.5

TABLE III

RESULTS OF CONFIDENCE INTERVAL TESTING FOR THE

CIRCULAR PARABOLOID, $z=x^2+y^2$

Without Experimental Error				With Experimental Error			
K	H	End Point %	Mid-Point %	K	H	End Point %	Mid-Point %
33	1.55617	76	74	37	2.20099	84	84
42	2.11937	76	78	35	3.12220	81	78
39	4.78946	90	91	37	2.71459	77	77
36	4.37617	84	83	40	1.91029	75	73
34	1.77424	85	87	36	3.67513	92	92
31	2.88453	83	80	37	2.13270	88	82
34	2.28204	78	70	37	2.94569	91	87
41	4.02237	91	87	36	3.06212	85	83
32	4.36839	76	81	41	2.64807	86	83
41	2.91650	87	87	33	4.95650	89	90
Average %		82.6	81.8	Average %		84.8	82.9

TABLE IV

RESULTS OF CONFIDENCE INTERVAL TESTING FOR THE

ELLIPTIC PARABOLOID, $z=x^2+8y^2$

Without Experimental Error				With Experimental Error			
K	H	End Point %	Mid-Point %	K	H	End Point %	Mid-Point %
42	170.72200	86	83	35	86.79742	92	92
30	138.17799	88	91	37	36.80585	71	71
36	47.02765	82	81	33	50.48413	83	81
36	62.62494	87	87	31	79.30219	89	92
36	77.86133	83	79	29	64.18718	69	65
40	236.49677	85	83	28	24.50569	67	67
35	73.42567	92	91	34	46.49091	59	57
33	26.07060	69	69	35	39.79968	73	71
31	75.15530	78	78	44	39.16206	77	77
36	75.84314	76	76	45	124.76241	94	98
Average %		82.6	81.8	Average %		77.5	77.1

APPENDIX D

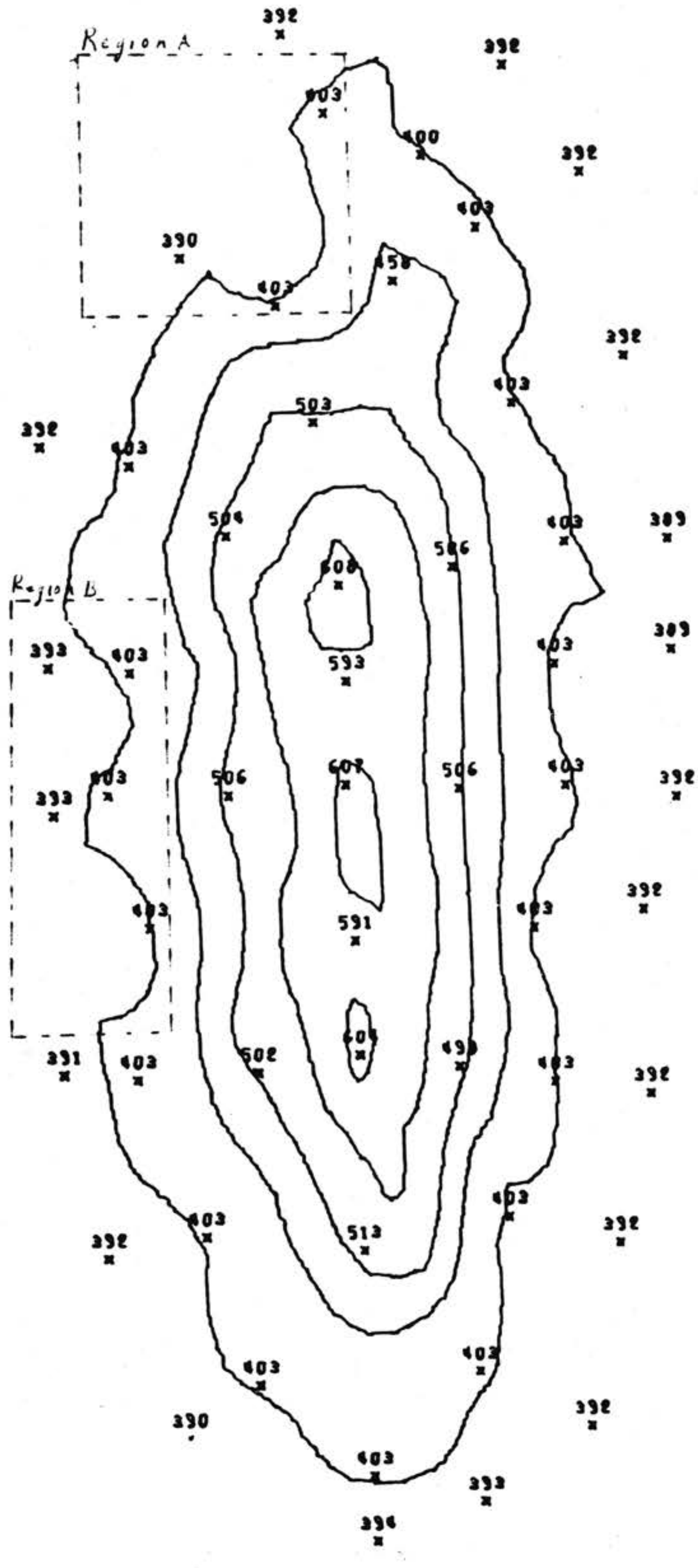


Figure 21. Grid Method Plot

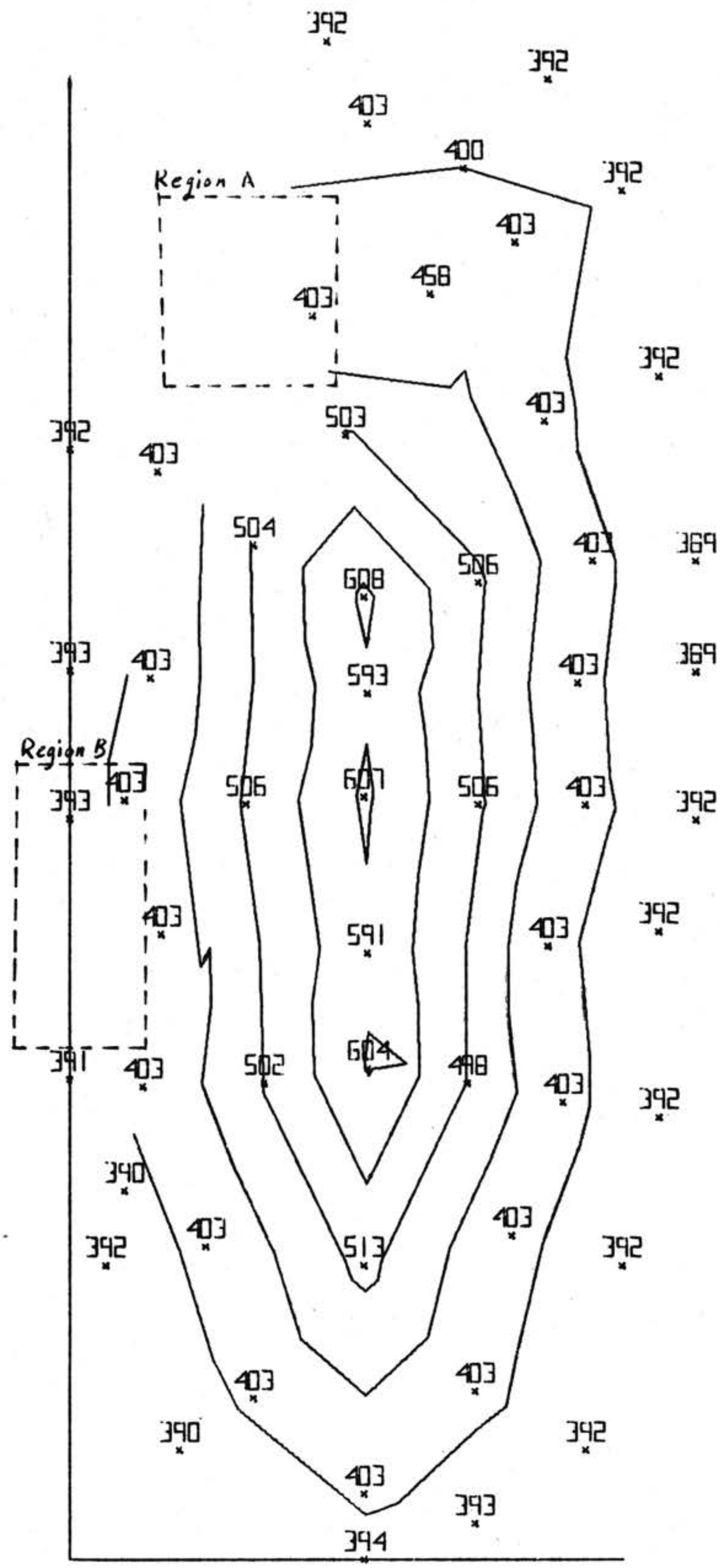


Figure 22. Quadrilateral Method Plot

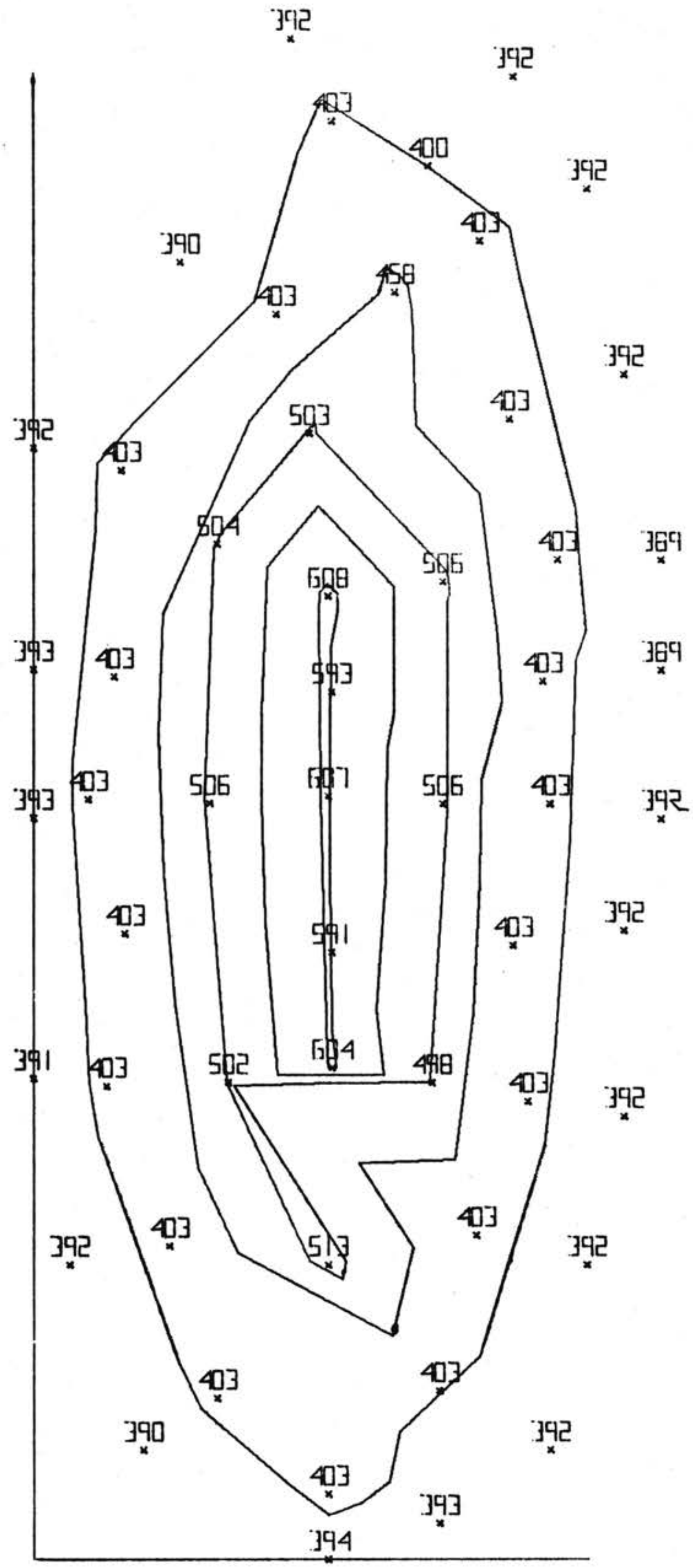


Figure 23. Triangular Method Plot

APPENDIX E

OPERATION OF THE PROGRAM FOR TRIANGULAR CONTOUR ESTIMATION

To aid in the operation of the computer program for the triangular method of contour estimation, a general description of the input and output for the program is given here. Details of the input may be obtained from comments at the beginning of the program.

Input

The input to this program consists of two parts. Control cards give the following information:

a) Problem identification

b) Number of observed responses

c) Minimum x, y, and z

d) Maximum x, y, and z

e) Scaling factor of x and y or size of axis for x and y

f) Distance between response values of contour estimates

Some of these values are optional because they can be found from the input data or from other control values.

For the data it is required to have the coordinates (x,y,z) for each observation. The Fortran programming system allows many changes in the exact card format to be made easily.

Output

The final output of the program is a listing of the error analysis and a data set which is on disk or tape, giving the (x,y) coordinates of the line segments which make up a contour. To put these segments on paper, another program is needed. A separate program is used here because of the variety of ways the contours can be plotted.

Two devices for plotting are considered. The first will print asterisks on continuous form listing paper. These asterisks will simulate very roughly the line for each contour. If the resulting plot is too wide for one sheet, the listing is made so that sheets may be attached to form the entire picture.

Another method for plotting is the punching of cards giving the (x,y) coordinates of the contours in such a form that a simple IBM 1620 program, which is written, can be used to plot the curves with a Calcomp plotter connected on-line to the 1620.

VITA

Edgar Louis Butler

Candidate for the Degree of

Doctor of Philosophy

Thesis: CONTOUR PLOTTING OF RESPONSE SURFACES

Major Field: Mathematics and Statistics

Biographical:

Personal Data: Born in Fort Worth, Texas, November 30, 1939, the son of Corwin A. and Edith P. Butler.

Education: Attended grade school and junior high school in Blackwell, Oklahoma; graduated from Blackwell Senior High School in 1958; received the Bachelor of Science degree from the Oklahoma State University, with a major in electrical engineering, in May, 1963; received the Master of Science degree from the Oklahoma State University, with a major in mathematics, in May 1965; completed requirements for the Doctor of Philosophy degree in May, 1967.

Professional experience: Employed as a student assistant, graduate assistant, programmer, and staff assistant in the Computer Center at Oklahoma State University from June, 1959 to September, 1966, except for the summers of 1961, 1962, and 1963, when employed as an Assistant Programmer by International Business Machines Corp. in Endicott, New York in the Applied Programming Department of the General Products Division; employed, since September, 1966, as Assistant Professor of Computer Science and Mathematics at the University of Missouri at Rolla.