Became SLAC-PUB 391

CGTM Number 213
June 1991

OBSOLETE

GKS-EZ
Graphic Algorithms for FORTRAN-77

by
Robert C. Beach
Computation Research Group
Stanford Linear Accelerator Center
Stanford, California 94309

WARNING

This document is only a proposal for a project. Although it has been written as if the subroutines already existed, they are not actually ready for use. They will not be available until SLAC gets a usable GKS system.

Contents

Chapter 1

Introduction

This document describes a number of mathematical subroutines that can be useful in GKS graphic applications programmed in FORTRAN-77. The algorithms described here include algorithms for projecting two-dimensional and three-dimensional space onto two-dimensional space, drawing smooth curves, and producing two-dimensional projections of complex three-dimensional objects. FORTRAN-77 is described in *American National Standard, Programming Language, FORTRAN* [ANS78]. GKS is described in *American National Standard for Information Systems: Computer Graphics - Graphical Kernel System (GKS) Functional Description* [ANS85a] and the FORTRAN-77 interface is described in *American National Standard for Information Systems: Computer Graphics - Graphical Kernel System (GKS) FORTRAN Binding* [ANS85b].

These subroutines can most conveniently be used with GKS-EZ. GKS-EZ is a subroutine package that attempts to make GKS more usable by people who are not computer graphic specialists and is described in *GKS-EZ Programming Manual for FORTRAN-77* [Bea87]. Actually the connection of these subroutines with GKS-EZ is quite tenuous; about the only part of GKS-EZ that they use is a subroutine to print error messages. However, their design and purpose is very similar.

To be consistent with GKS-EZ, all of the subroutine names that will be described in this document begin with the letters "GZ."

Many concepts will have to be defined in the following chapters. When a concept is first encountered, it will be given in *italics*. The information around the italicized word or phrase may be taken as its definition.

1.1. The Availability of the Subroutines

The subroutines described in this document are available on the IBM mainframe computers running at the Stanford Linear Accelerator Center. These computers are running under the VM/XA operating system. Executable versions of the subroutines are contained in the file

```
GKSEZ TXTLIB U.
```

They may therefore be used by anyone at this installation who uses the proper `TXTLIB` statement. The `TXTLIB` contains the GKS-EZ subroutines as well as the subroutines described in the following chapters.

The source code is also available for those people who have to use the subroutines on another computer. The file

```
GKSGAUTL FORTRAN U
```

contains a group of mathematical subroutines that are used by the other subroutines described here. The file

`GKSGATRN FORTRAN U`

contains the transformation subroutines of Chapter 2. The file

`GKSGACRV FORTRAN U`

contains the curve drawing subroutines of Chapter 3. Finally, the file

`GKSGASUR FORTRAN U`

contains the surface drawing subroutines of Chapter 4.

Since the these subroutines are written in strict FORTRAN-77, they themselves are transportable to any computer with a FORTRAN-77 compiler and a GKS system. One possible modification is that of a control value, `INFN`, that appears in a number of subroutines. That value is used to check for things like singular matrices and to guard against division by zero. The value may have to be changed for computers with differing word size or precision. In fact, it may be necessary to change this value on the host computer as more experience is accumulated with these subroutines.

Chapter 2

Projective Transformations

This chapter describes a group of subroutines that may be used to define projective transformations from two-dimensional or three-dimensional space into two-dimensional space. Subroutines are provided which generate the transformations and encode them as a matrix. Other subroutines are then provided that take a point, in two-dimensional or three-dimensional space, and project them into two-dimensional space. The mathematical derivation of all of these projective transformation algorithms is given in *An Introduction to the Curves and Surfaces of Computer-Aided Design* [Bea91].

One use of the two-dimensions to two-dimensions transformation is in digitizing photographs. If the photograph contains a figure of known dimensions then the transformation from the coordinate system of the photograph to real two-dimensional space can often be determined. A projective transformation is also the physically correct transformation if the optical system of the camera approximates a pinhole camera.

The three-dimensions to two-dimensions transformations are useful whenever two-dimensional images of three dimensional objects are required.

These transformations have many desirable properties. One of the most important is that they transform straight lines into straight lines. Another advantage is that neither the generation of the transformation nor the projection of a point is computationally expensive.

If one of the transformation generating subroutines determines that the transformation does not exist, it sets an error indicator and returns to the caller. The subroutines that project a point should always work unless they are supplied with extremely large coordinates.

2.1. Two-dimensions to Two-dimensions Projective Transformations

This section describes a means of generating and using a projective transformation from two-dimensional space to two-dimensional space. The transformation is defined by giving four points in the source coordinate system and the corresponding four points in the target coordinate system. The resulting projective transformation will always be computable provided no three of the points lie on a straight line in either coordinate system.

There is, however, a problem with points that transform into a *point at infinity.* To understand this problem, refer to Figure (2.1). In this figure, the four points on the irregular quadrilateral, $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, and $\mathbf{P}_4$, are to be transformed into the

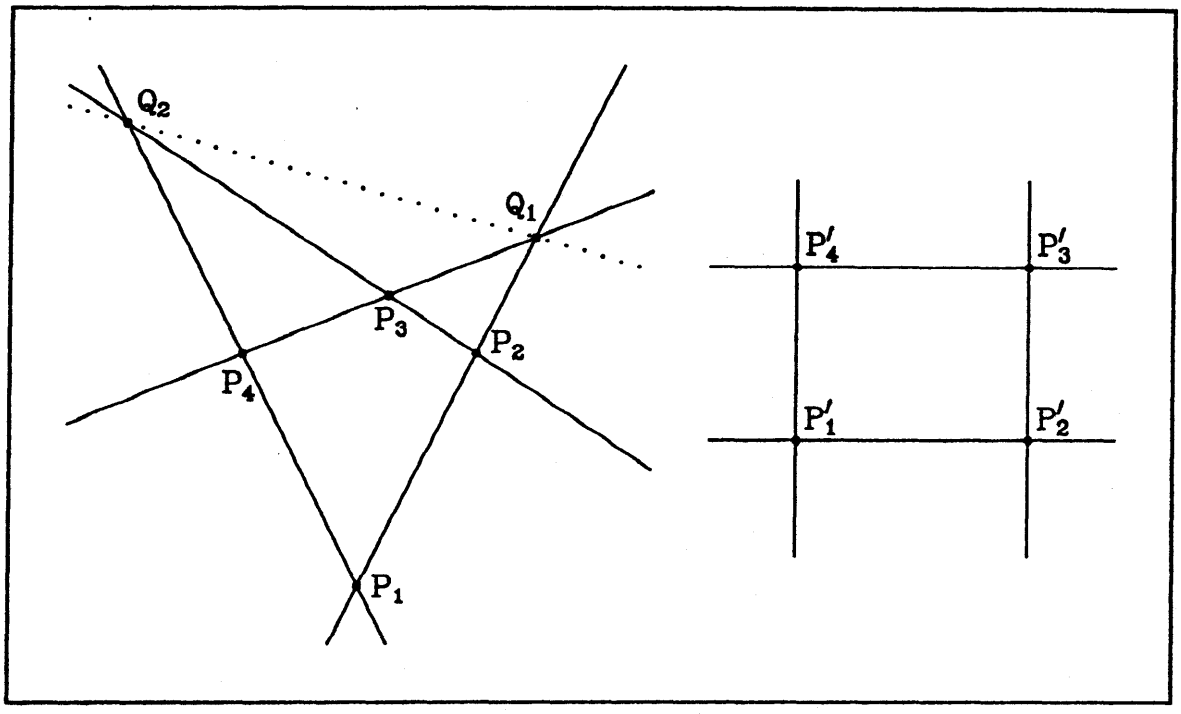

Figure 2.1. A two-dimensions to two-dimensions projective transformation

rectangle described by $\mathbf{P}'_1$, $\mathbf{P}'_2$, $\mathbf{P}'_3$, and $\mathbf{P}'_4$. The line through the points $\mathbf{P}_1$ and $\mathbf{P}_2$ intersects the line through the points $\mathbf{P}_3$ and $\mathbf{P}_4$ at $\mathbf{Q}_1$. The lines through the corresponding $\mathbf{P}'_i$ points do not intersect, or rather, they intersect at infinity. The point $\mathbf{Q}_1$ therefore transforms into a point at infinity. The point $\mathbf{Q}_2$ similarly transforms into a point at infinity. Since straight lines are preserved under the transformation, all of the points on the dotted line through $\mathbf{Q}_1$ and $\mathbf{Q}_2$ transform into points at infinity. The subroutine that transforms a point from one coordinate system to another will determine if the given point transforms into a point at infinity and warn the caller.

2.1.1. Subroutine GZ22PJ: Generate a Transformation

This subroutine may be used to generate a two-dimensions to two-dimensions projective transformation that carries four given points into four given points.

The calling sequence is:

```
CALL GZ22PJ(PXAS,PYAS,PXAT,PYAT,IERR,PTRN)
```

The input parameters are:

PXAS — A real array of dimension 4 containing the x coordinates of the source points.

PYAS — A real array of dimension 4 containing the y coordinates of the source points.

PXAT A real array of dimension 4 containing the x coordinates of the target points.

PYAT A real array of dimension 4 containing the y coordinates of the target points.

The output parameters are:

IERR An integer giving an error flag. A nonzero value means the transformation could not be computed.

PTRN A real array of dimension (3,3) containing the projective transformation.

2.1.2. Subroutine GZ22TR: Transform a Point

This subroutine may be used to transform a point using a two-dimensions to two-dimensions projective transformation. A flag indicates if the projected point is a finite point or a point at infinity.

The calling sequence is:

```
CALL GZ22TR(PTRN,PAS,PAP,FLAG)
```

The input parameters are:

PTRN A real array of dimension (3,3) containing the projective transformation.

PAS A real array of dimension 2 containing the source point.

The output parameters are:

PAP A real array of dimension 2 containing the projected point.

FLAG A real value that indicates whether a finite point or a point at infinity has been computed. If this value is nonzero, PAP contains the finite coordinates of the projected point. If this value is zero, PAP is a unit vector pointing in the direction of the point at infinity.

2.2. Three-dimensions to Two-dimensions Projective Transformations

This section describes two ways to generate a three-dimensions to two-dimensions projective transformation.

In the first case the projection of a point in three-dimensional space is defined by an eye point and a projection plane as shown in Figure (2.2). The plane is defined by an origin point, **O**, on the plane, and two direction vectors, **H** and **V**. **H** is the "horizontal" direction and **V** is the "vertical" direction. These two direction vectors will often be perpendicular to each other but that is not necessary in the subroutines. A point on the plane, **Q**, is found by starting at **O**, and moving parallel to **H** the necessary distance and then parallel to **V** the necessary distance. Thus, **Q** is represented as

$$\mathbf{Q} = \mathbf{O} + \xi\mathbf{H} + \eta\mathbf{V}.$$

Thus the vectors **H** and **V** impose a coordinate system on the plane. The projection of a point **P** onto the plane is then obtained by drawing a straight line through the

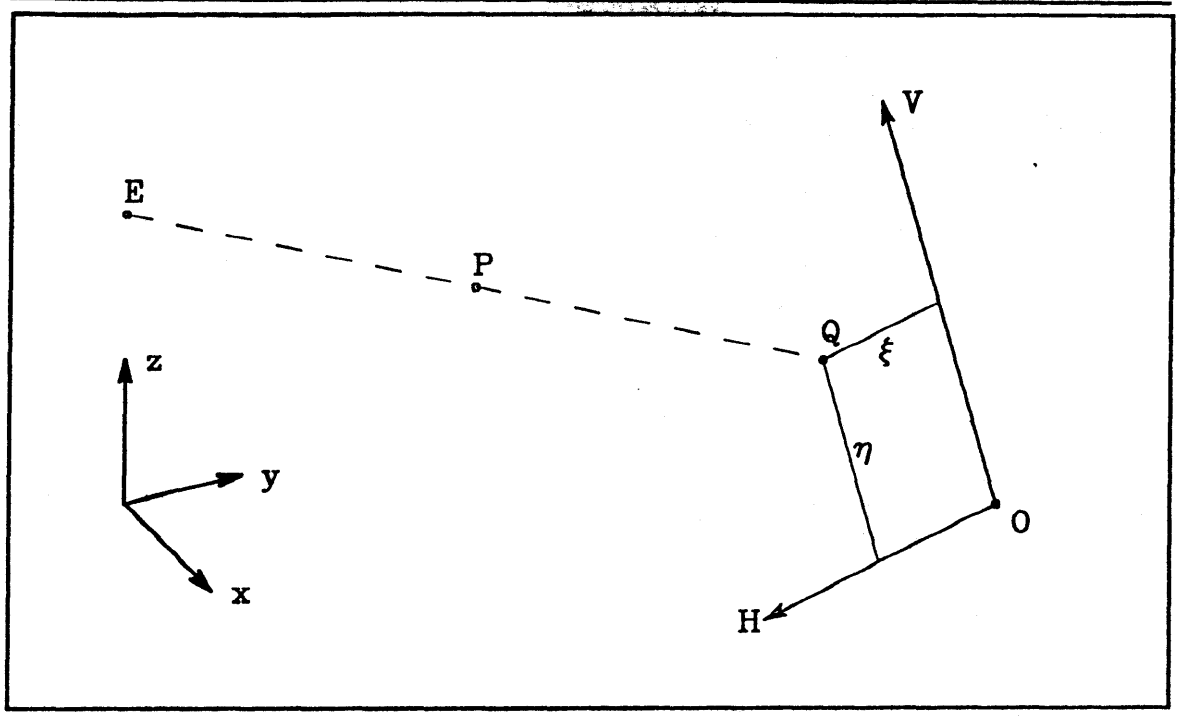

Figure 2.2. A three-dimensions to two-dimensions perspective transformation

eye point, E, and the point P until it meets the plane. The ξ and η values of the intersection point are the coordinates of the projected point in two-dimensional space. In many applications the vector from E to O is perpendicular to both H and V but that is not necessary in these subroutines. This type of transformation is known as a *perspective transformation*. These transformations are best understood by imagining a viewer at the eye point, looking toward the origin point.

The second type of three-dimensions to two-dimensions transformation that is described here is known as a *parallel transformation*. It is formed by projecting a given point P parallel to a fixed direction D as shown in Figure (2.3). It is again common to have D perpendicular to both H and V.

In the case of a perspective transformation, the horizontal and vertical directions must be distinct and neither may point at the eye point. In a parallel transformation, the horizontal, vertical, and projection directions must all be distinct.

A perspective transformation can also produce points at infinity. All of the points on the plane through the eye point and parallel to the projection plane map into points at infinity except for the eye point itself. The eye point has no corresponding point. A parallel transformation never produces points at infinity.

2.2.1. Subroutine GZ32PT: Generate a Perspective Transformation

This subroutine may be used to generate a three-dimensions to two-dimensions perspective transformation. The transformation is defined by defining the projection plane and an eye point. The projection plane is specified by giving a point on the plane and a horizontal and vertical direction within the plane.

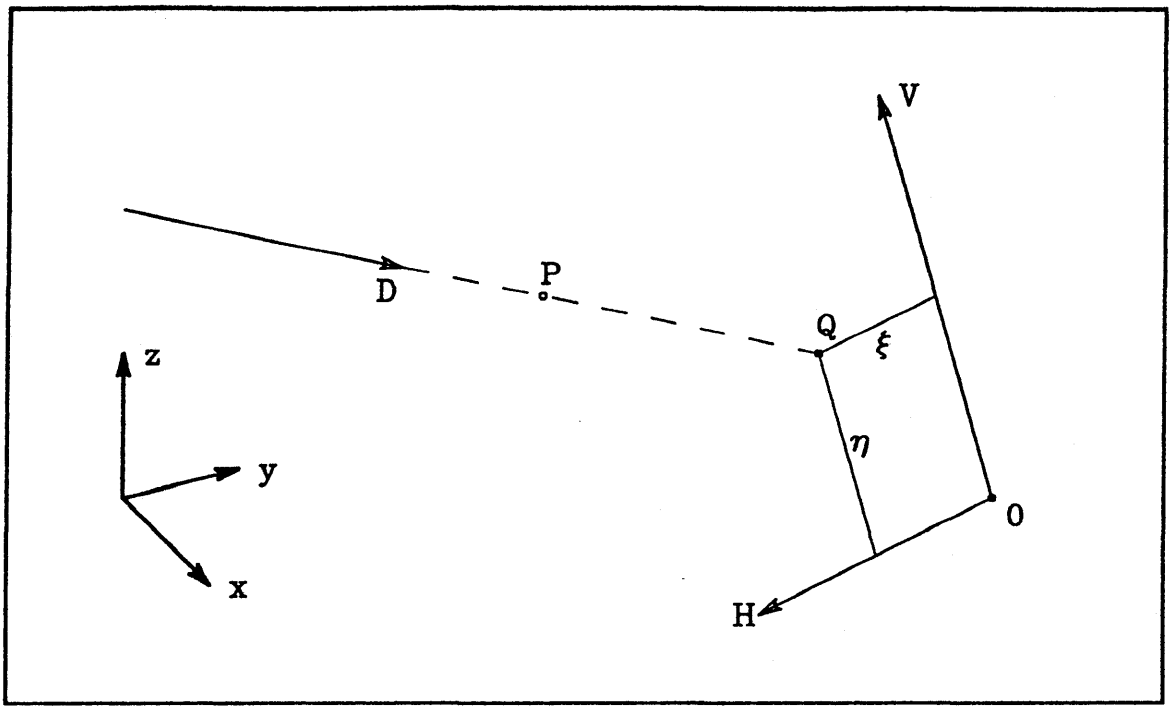

Figure 2.3. A three-dimensions to two-dimensions parallel transformation

The calling sequence is:

```
CALL GZ32PT(PO,HD,VD,PE,IERR,PTRN)
```

The input parameters are:

- `PO` A real array of dimension 3 containing the origin point on the projection plane.
- `HD` A real array of dimension 3 containing the horizontal direction in the projection plane.
- `VD` A real array of dimension 3 containing the vertical direction in the projection plane.
- `PE` A real array of dimension 3 containing the eye point.

The output parameters are:

- `IERR` An integer giving an error flag. A nonzero value means the transformation could not be computed.
- `PTRN` A real array of dimension (3,4) containing the projective transformation.

2.2.2. Subroutine GZ32PL: Generate a Parallel Transformation

This subroutine may be used to generate a three-dimensions to two-dimensions parallel transformation. The transformation is defined by defining the projection plane and a projection direction. The projection plane is specified by giving a point on the plane and a horizontal and vertical direction within the plane.

The calling sequence is:

```
CALL GZ32PL(PO,HD,VD,PD,IERR,PTRN)
```

The input parameters are:

- PO — A real array of dimension 3 containing the origin point on the projection plane.
- HD — A real array of dimension 3 containing the horizontal direction in the projection plane.
- VD — A real array of dimension 3 containing the vertical direction in the projection plane.
- PD — A real array of dimension 3 containing the projection direction.

The output parameters are:

- IERR — An integer giving an error flag. A nonzero value means the transformation could not be computed.
- PTRN — A real array of dimension (3,4) containing the projective transformation.

2.2.3. Subroutine GZ32TR: Transform a Point

This subroutine may be used to transform a point using a three-dimensions to two-dimensions projective transformation. A flag indicates if the projected point is a finite point or a point at infinity.

The calling sequence is:

```
CALL GZ32TR(PTRN,PAS,PAP,FLAG)
```

The input parameters are:

- PTRN — A real array of dimension (3,4) containing the projective transformation.
- PAS — A real array of dimension 3 containing the source point.

The output parameters are:

- PAP — A real array of dimension 2 containing the projected point.
- FLAG — A real value that indicates whether a finite point or a point at infinity has been computed. If this value is nonzero, PAP contains the finite coordinates of the projected point. For a perspective transformation, a positive value indicates the source point is in front of the viewer while a negative value indicates it is behind the viewer. In these cases, the magnitude of FLAG is proportional to the distance from the eye point to the projected point; it can be used as the projected distance from the eye point to the source point. If this value is zero, PAP is a unit vector pointing in the direction of the point at infinity. If the source point is the eye point of a perspective transformation, both components of PAP and the value of FLAG will be zero.

Chapter 3

Curve Drawing Algorithms

This chapter describes a group of subroutines that may be used to draw smooth curves. The curves are defined by supplying control points and other control information to the subroutines. The curves are drawn by breaking them down into small straight line segments and then calling the GKS polyline subroutine, GPL. The user has control over the number of line segments generated. The mathematical derivation of all of these curve drawing algorithms is given in *An Introduction to the Curves and Surfaces of Computer-Aided Design* [Bea91].

Mathematically, all of these curves are defined *parametrically*, that is, the x and y coordinates are defined as functions of a parameter, t. In effect, a user may specify the parameter at each of the control points. Different assignments of the parameter values at the control points usually results in different curves. There are two schemes that are commonly used to define the values of the parameter associated with the given control points. These two schemes produce curves that are known as *uniform* and *nonuniform* curves. For uniform curves, the parameter is set to zero at the first point and increases by one for each succeeding point. For nonuniform curves, the parameter may be set to any increasing sequence of positive values.

The uniform scheme is very simple mathematically but often does not produce acceptable curves if the points are not nearly equally spaced. A nonuniform scheme that usually produces good results is based on accumulated chord length along the sequence of points. The parameter is set to zero for the first point and increases by an amount equal to the distance between consecutive points for each point. For later reference, we display the increments in the parameter for this nonuniform case

$$\begin{aligned} D_1 &= \text{distance from point 1 to point 2,} \\ D_2 &= \text{distance from point 2 to point 3,} \\ &\cdots \\ D_{\mathtt{N}-2} &= \text{distance from point } (\mathtt{N}-2) \text{ to point } (\mathtt{N}-1), \\ D_{\mathtt{N}-1} &= \text{distance from point } (\mathtt{N}-1) \text{ to point } \mathtt{N}, \end{aligned} \tag{3.1}$$

where N is the number of given control points. The subroutines described in this chapter all start the parameter at zero and expect the user to supply the increments in parameter value, explicitly or implicitly, along the curve.

In the following subroutines, the parameter values are supplied by two arguments; the first, NP, is an integer and the second, PA, is a real array. If NP is positive, the dimension of PA must be NP. The increments in parameter values are

then obtained from the PA array. If more parameter values are needed than are contained in PA, then they are obtained cyclically from PA. That is, the values PA(1), ..., PA(NP) are obtained and then this sequence is repeated. This makes it very easy to specify the uniform curve; NP is simply given an integer value of one while PA is given a real value of one. It is also easy to specify the nonuniform curve with the parameter value based on accumulated chord length. This is done by giving NP a value of zero. In this case, PA is ignored and the subroutine calculates the parameter internally.

Most of the algorithms described here produce curves by using concatenations of simple parametric polynomials. The parametric polynomials are usually of low degree (normally two or three). The points at which consecutive polynomials join are known as ***knots***.

In addition to the simple polynomial form of these algorithms, some also have a *rational* form. The rational form consists of x and y being defined as quotients of polynomials. In certain applications, the rational form can be more useful. For example, the only conic the polynomial form can ever match exactly is the parabola. It is impossible for the polynomial form to exactly match a simple circle although it can come arbitrarily close. On the other hand, a rational parametric quadratic can exactly match any conic.

Two distinct types of curves, *interpolation curves* and *design curves*, may be produced by these subroutines. Interpolation curves pass through all of their control points while design curves do not necessarily do this.

The description of each subroutine will include figures showing examples of curves produced by the subroutines. In these figures, the given control points are joined by straight lines between consecutive points. This open polygon is known as the ***control polygon***. The reader will notice that these figures do not display the coordinate axes. The reason for this is that all of the curves described here are ***isotropic***, that is, they are independent of the coordinate system in which they are defined. In fact, the reader may draw a set of coordinate axes anywhere in these figures and label the axes in any units. The figures also do not label the points so the reader cannot tell which end of the curve corresponds to the first point. The reason for this is that these curve drawing algorithms are all *symmetric*, that is, they do not depend on which end of the control polygon is the starting end.

If one of these subroutines detects an error in the data supplied to it, the subroutine prints an error message and returns without producing any graphic output.

3.1. Bessel's Method of Local Cubic Interpolation

Bessel's method is a cubic interpolation algorithm. Between each pair of points is a segment of a parametric cubic. Adjacent cubic segments join at the control points and have tangent vectors at those points which have the same direction. The method is also local in that a cubic segment is completely determined by the two control points on either side of it. In addition to the usual parameter values that are associated with the line segments in the control polygon, there are additional

parameters associated with the tangent vectors at the points. This combination of parameters gives the user a substantial amount of control over the final interpolation curve.

The two subroutines that are described here differ in the type of control that the user has over the ends of the curve. In the first subroutine, the user must supply and extra point beyond the actual ends of the curve. In the second subroutine, the user may specify the tangent direction at the end points or request that the curvature be zero. In this later case, the end conditions may be mixed, that is, the user may specify a tangent vector at one end and request zero curvature at the other.

3.1.1. Subroutine GZBESL: Draw a Parametric Bessel's Curve (I)

This subroutine may be used to draw a curve through a sequence of points using Bessel's method. In this scheme, the ends of the curve are controlled by an extra point. The actual curve, therefore, extends from the second control point to the second point from the end of the curve. Either a uniform curve, or a nonuniform curve may be drawn. In the case of a nonuniform curve, a simple means to base the line segment parameters on accumulated chord length is provided.

The calling sequence is:

```
CALL GZBESL(N,PXA,PYA,NP,PA,NT,TA,NS)
```

The input parameters are:

`N`	An integer giving the number of control points.
`PXA`	A real array of dimension `N` containing the x coordinates of the control points.
`PYA`	A real array of dimension `N` containing the y coordinates of the control points.
`NP`	An integer giving the number of parameter values associated with line segments in the control polygon. If this value is not positive, accumulated chord length will be used to generate the parameter. If this parameter is positive, values are selected cyclically from the next parameter. In this case, a total of (`N` $-$ 1) values are needed.
`PA`	If `NP` is positive, this is a real array of dimension `NP` containing the given parameter values associated with the line segments.
`NT`	An integer giving the number of parameter values associated with tangent vectors at the interior points. This value must be positive and the values are selected cyclically from the next parameter. A total of (`N` $-$ 2) values are needed.
`TA`	A real array of dimension `NT` containing the given parameter values associated with the tangent vectors.
`NS`	An integer giving the number of straight line segments into which each curve segment is to be divided.

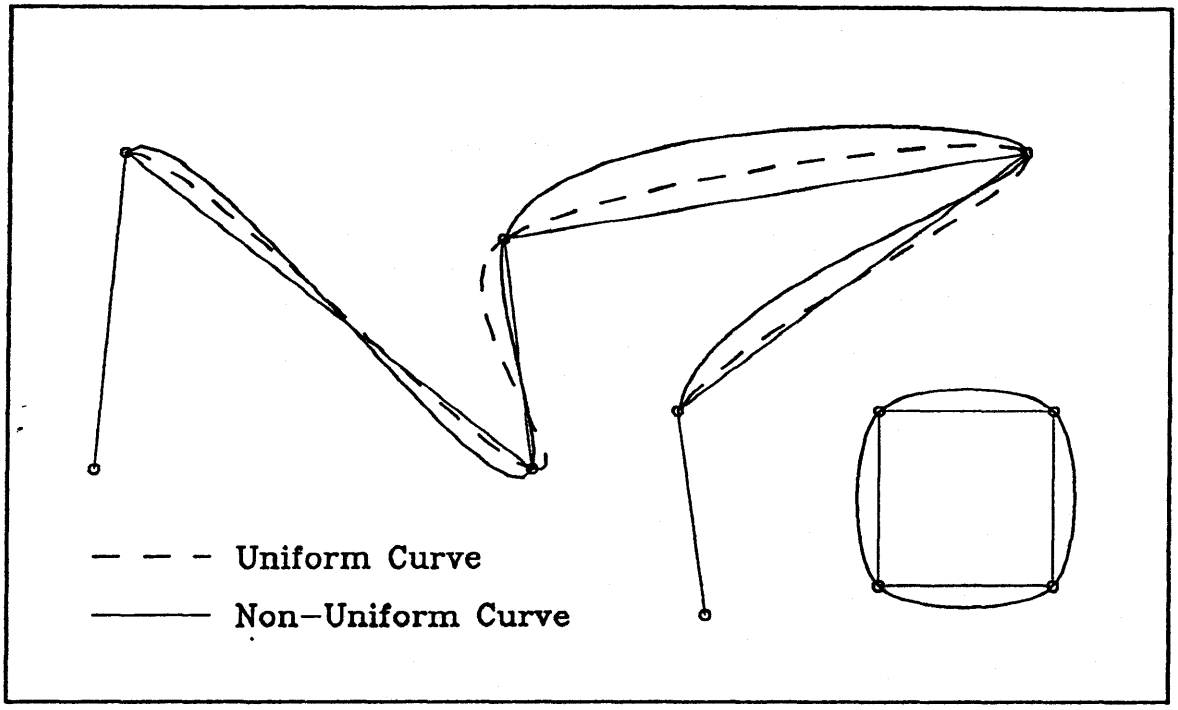

Figure 3.1. Examples of interpolation by Bessel's method (I)

Figure (3.1) includes an example of a nonuniform curve where accumulated chord length has been used as the parameter. The TA values have all been set to one. The circular curve at the lower right of Figure (3.1) was formed by specifying seven points at the corners of the square in sequence. Since the chord segments are equal, the uniform and nonuniform curves based on accumulated chord length are identical.

Figure (3.2) illustrates the affect the PA values have on the curve. The figure illustrates the manipulation the PA value associated with the central line segment of the control polygon. It shows that reducing the value of PA(3) causes the curve to move closer to the chord between the third and fourth points. In this case, the tangent vectors at the third and fourth points also rotate to become closer to the chord. Large values of PA(3) cause the curve to move away from the chord and a cusp or loop can form if it is made too large. Figure (3.2) also illustrates the local properties of the interpolation because all three composite curves are tangent to each other at their ends; any continuation of the curve beyond its current ends will not be affected by the change in the parameter.

Figure (3.3) illustrates the manipulation of the TA values. The natural value of the TA values is one. As TA(2) is reduced, the influence of the tangent vector at the middle point is reduced and the curve pulls away from the tangent vector and approaches the adjacent chords. However, in this case, the tangent direction at the middle point does not change. If TA(2) is made large, the influence of the tangent vector at the middle point becomes strong. This forces the interpolation curve to flatten and follow the direction of the tangent vector longer. In general, when the

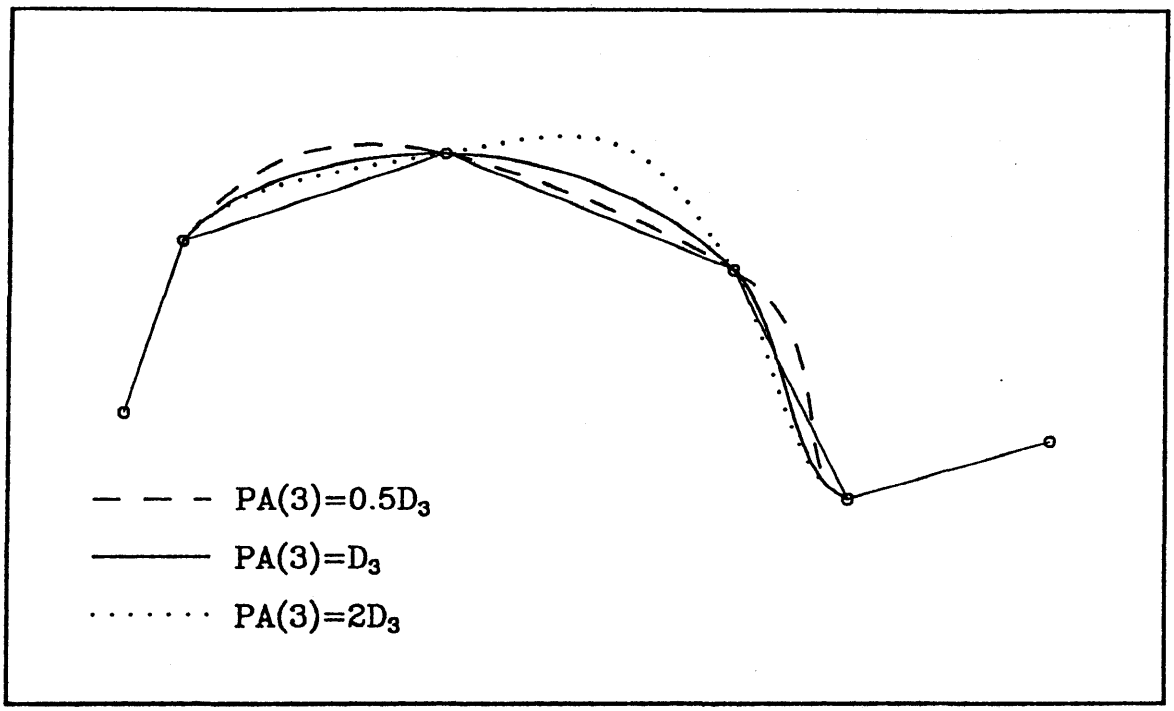

Figure 3.2. Examples of interpolation by Bessel's method (II)

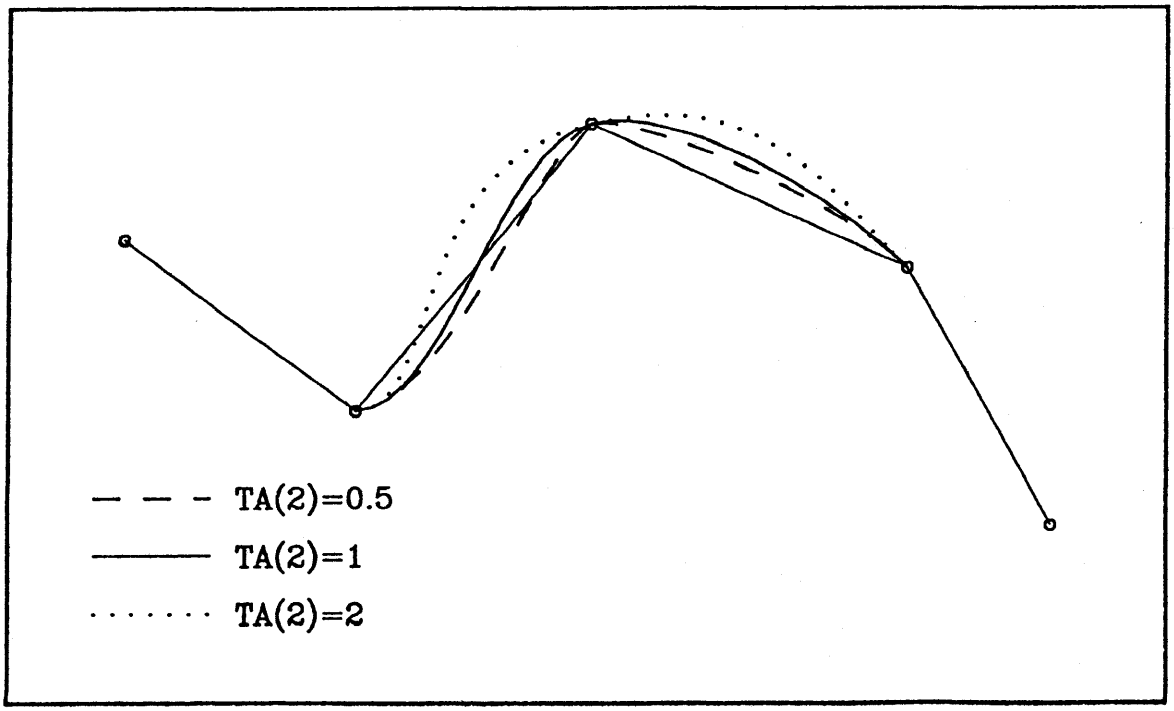

Figure 3.3. Examples of interpolation by Bessel's method (III)

TA values are reduced, the curve moves closer to the adjacent chords and becomes taut; increasing the TA values allows the curve to relax and bow out.

3.1.2. Subroutine GZBESE: Draw a Parametric Bessel's Curve (II)

This subroutine may be used to draw a curve through a sequence of points using Bessel's method. In this scheme, the ends of the curve are controlled by specifying the end tangents or by requesting zero curvature at the ends. Either a uniform curve, or a nonuniform curve may be drawn. In the case of a nonuniform curve, a simple means to base the line segment parameters on accumulated chord length is provided.

The calling sequence is:

```
CALL GZBESE(N,PXA,PYA,V1,V2,NP,PA,NT,TA,NS)
```

The input parameters are:

- `N` An integer giving the number of control points.
- `PXA` A real array of dimension `N` containing the x coordinates of the control points.
- `PYA` A real array of dimension `N` containing the y coordinates of the control points.
- `V1` A real array of dimension 2 containing the given tangent vector at the initial end. This argument should usually be a unit vector or a zero vector. If it is a zero vector, then zero curvature is imposed at the end.
- `V2` A real array of dimension 2 containing the given tangent vector at the terminal end. This argument should usually be a unit vector or a zero vector. If it is a zero vector, then zero curvature is imposed at the end.
- `NP` An integer giving the number of parameter values associated with line segments in the control polygon. If this value is not positive, accumulated chord length will be used to generate the parameter. If this parameter is positive, values are selected cyclically from the next parameter. In this case, a total of (`N` -1) values are needed.
- `PA` If `NP` is positive, this is a real array of dimension `NP` containing the given parameter values associated with the line segments.
- `NT` An integer giving the number of parameter values associated with tangent vectors at the points. This value must be positive and the values are selected cyclically from the next parameter. A total of `N` values are needed.
- `TA` A real array of dimension `NT` containing the given parameter values associated with the tangent vectors.
- `NS` An integer giving the number of straight line segments into which each curve segment is to be divided.

Figure (3.4) shows examples of interpolation by Bessel's method when tangents at the ends of the curve are supplied. In this case the curve is not, strictly speaking, symmetric. Since the tangents at the ends are supplied, they must point in the direction of the curve so this curve was drawn from the left to the right. To draw the curve in the other direction, the directions of the tangent vectors must be

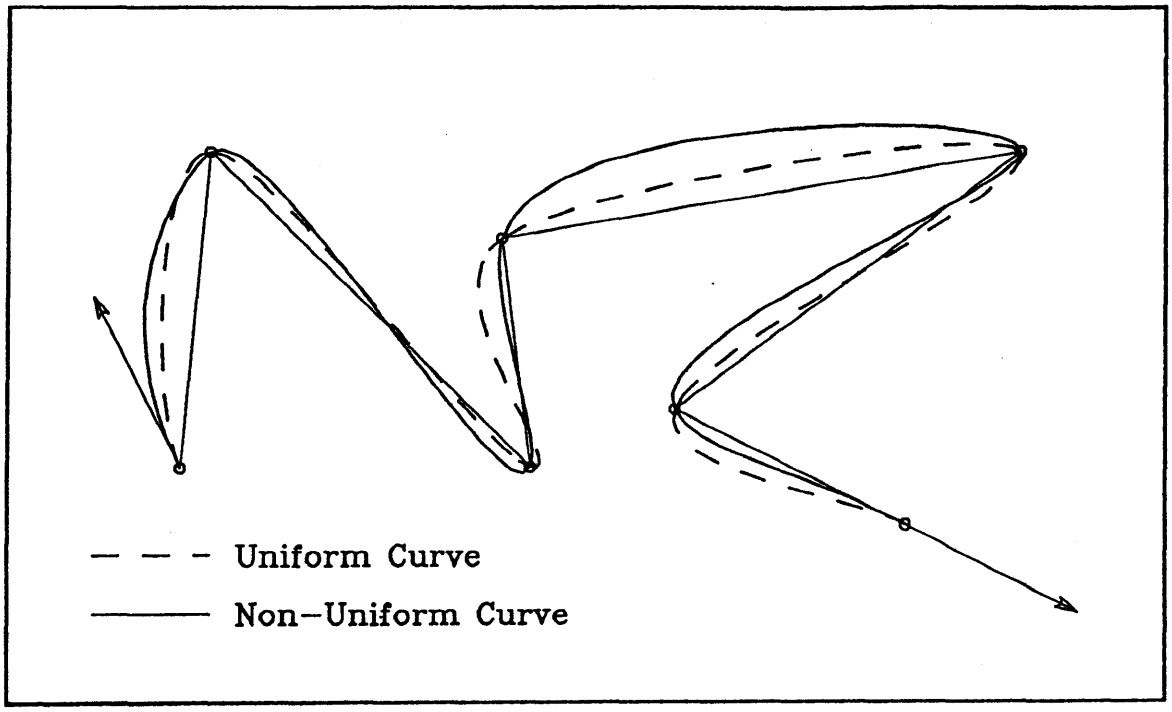

Figure 3.4. Examples of interpolation by Bessel's method with end tangents given

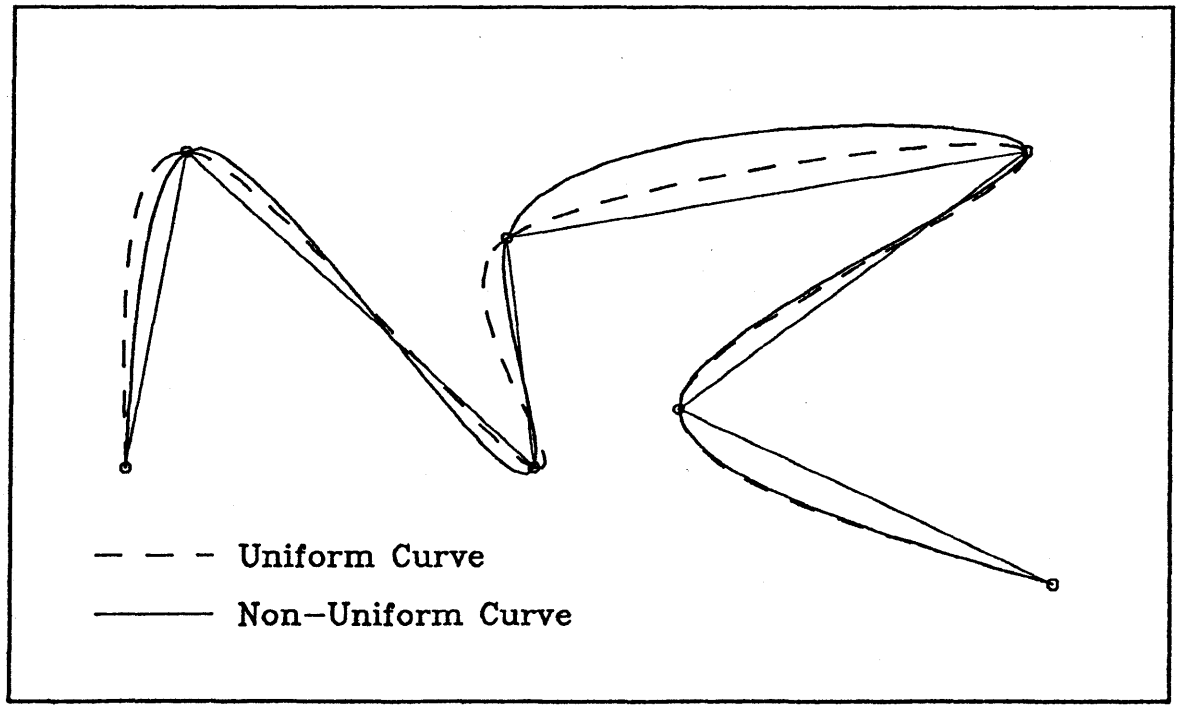

Figure 3.5. Examples of interpolation by Bessel's method with zero curvature at the ends

reversed. In the nonuniform curve, accumulated chord length has been used as the

parameter.

In Figure (3.5) the curvature at the end points has been constrained to be zero. The nonuniform curve again has accumulated chord length as its parameter.

3.2. Cubic Spline Interpolation

This section describes a subroutine that may be used to draw a parametric cubic spline. A cubic spline is an interpolation curve consisting of parametric cubic polynomial segments. The segments join at the knots with equal first and second derivatives. However, the curve is not local in nature; changing one control point modifies the entire curve.

There is a limit on the number of control points that may be supplied to this subroutine.

3.2.1. Subroutine GZSPLN: Draw a Parametric Cubic Spline

This subroutine may be used to draw a parametric cubic spline curve through a sequence of points. The ends of the curve are controlled by specifying the end tangents or by requesting zero curvature at the ends. Either a uniform curve, or a nonuniform curve may be drawn. In the case of a nonuniform curve, a simple means to base the parameter on accumulated chord length is provided.

The calling sequence is:

```
CALL GZSPLN(N,PXA,PYA,V1,V2,NP,PA,NS)
```

The input parameters are:

N	An integer giving the number of control points. The maximum number of points that are allowed is 32.
PXA	A real array of dimension N containing the x coordinates of the control points.
PYA	A real array of dimension N containing the y coordinates of the control points.
V1	A real array of dimension 2 containing the given tangent vector at the initial end. This argument should usually be a unit vector or a zero vector. If it is a zero vector, then zero curvature is imposed at the end.
V2	A real array of dimension 2 containing the given tangent vector at the terminal end. This argument should usually be a unit vector or a zero vector. If it is a zero vector, then zero curvature is imposed at the end.
NP	An integer giving the number of parameter values associated with line segments in the control polygon. If this value is not positive, accumulated chord length will be used to generate the parameter. If this parameter is positive, values are selected cyclically from the next parameter. In this case, a total of (N − 1) values are needed.
PA	If NP is positive, this is a real array of dimension NP containing the given parameter values associated with the line segments.

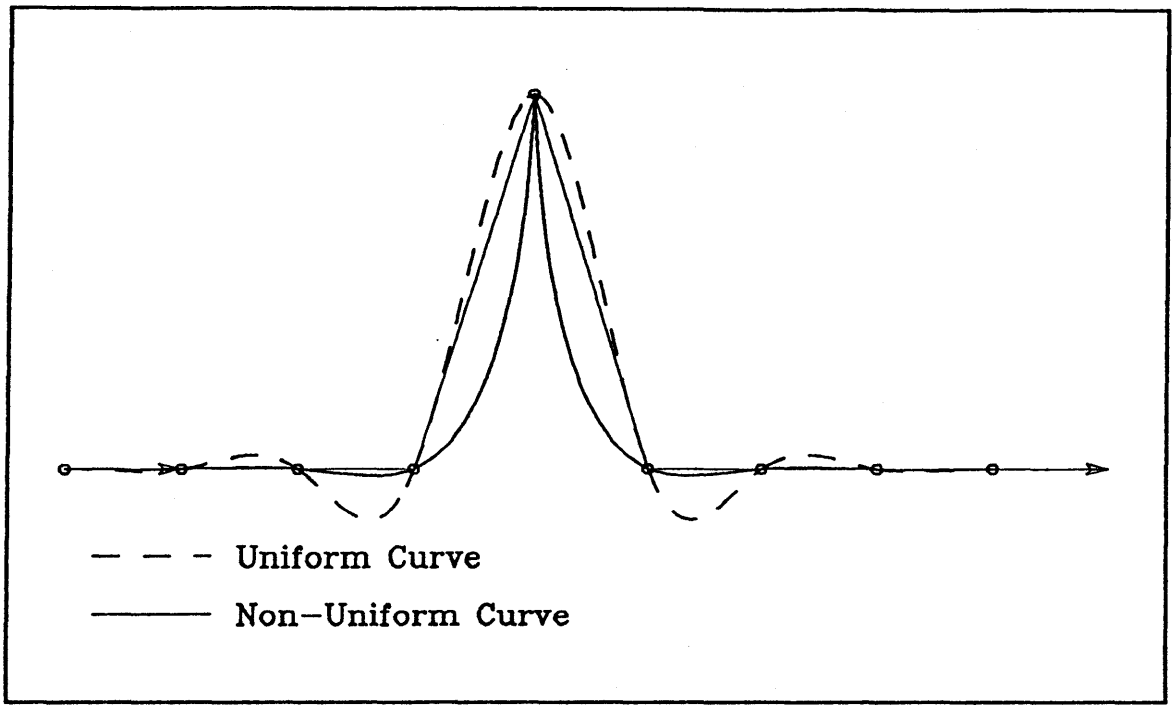

Figure 3.6. Parametric cubic splines with end tangents given

NS An integer giving the number of straight line segments into which each curve segment is to be divided.

Figure (3.6) shows examples of cubic splines with the tangents given at the end points. In this case the curve is not, strictly speaking, symmetric. Since the tangents at the ends are supplied, they must point in the direction of the curve so this curve was drawn from the left to the right. To draw the curve in the other direction, the directions of the tangent vectors must be reversed. The figure also shows the oscillatory behavior that is often a problem in spline curves.

In Figure (3.7), the spacing of the points was deliberately chosen to have large variation in the chord lengths. As a result, the uniform curve exhibits oscillatory problems at the top center of the figure.

Figure (3.8) illustrates how the PA values can be used to control the shape of the curve. In this case, chord lengths have been used for the parameters except that the PA value associated with the central line segment of the control polygon has been manipulated. As we have seen before, reducing a PA value causes the curve to move closer to the associated line segment. Figure (3.8) also shows that changes like these are not local; they affect the entire curve.

3.3. Bézier Curves

A Bézier curve is a design curve and not an interpolation curve. It does, however, pass through its first and last control points and is tangent to the first and last

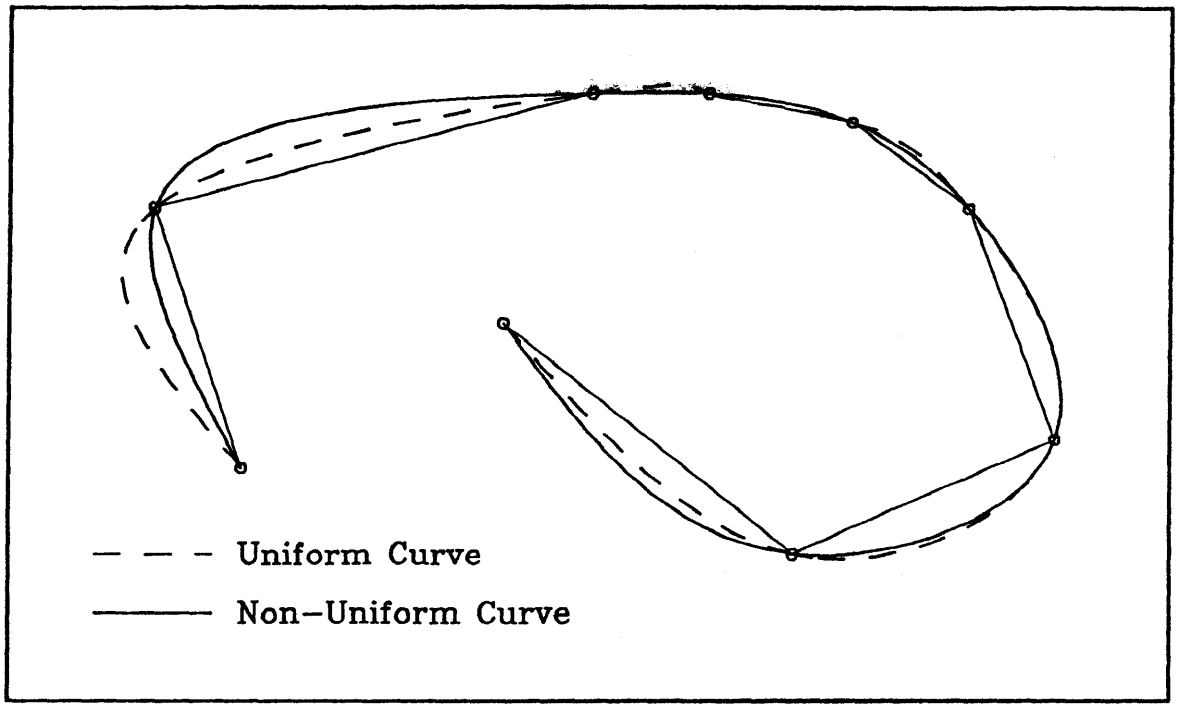

Figure 3.7. Parametric cubic splines with zero end curvature (I)

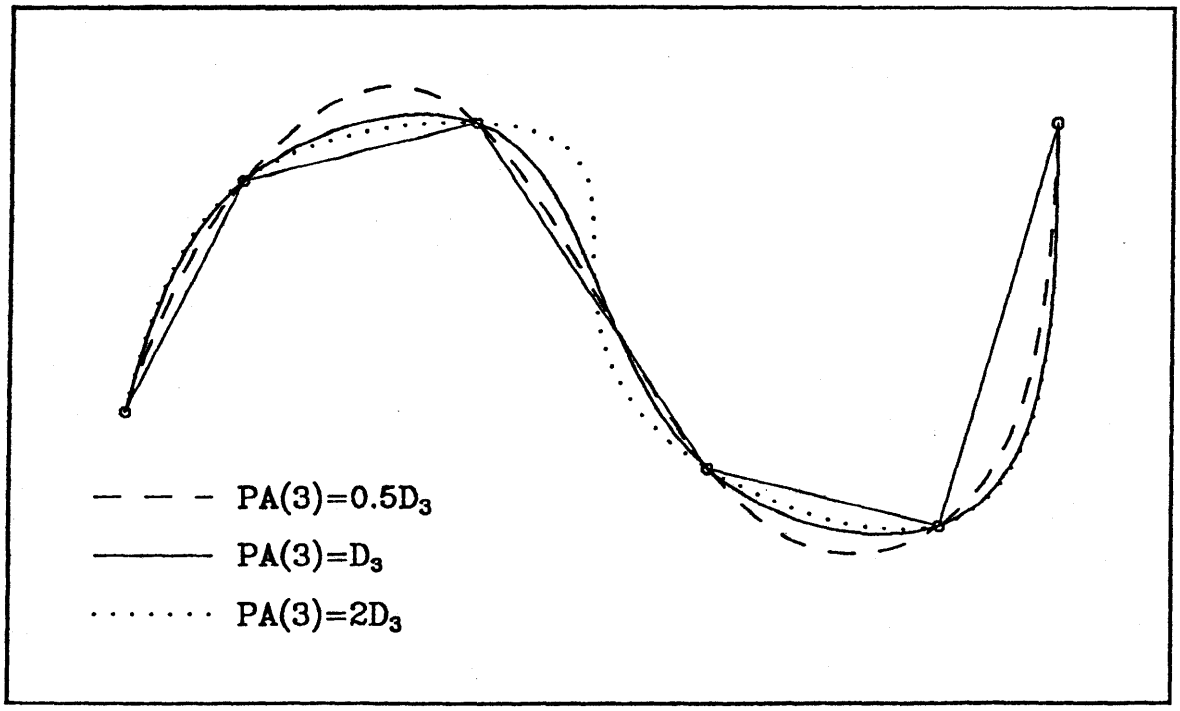

Figure 3.8. Parametric cubic splines with zero end curvature (II)

straight line segment in the control polygon. The Bézier curve is a parametric polynomial of large degree (in fact the degree is the number of control points minus

one). Although using polynomials of large degree is usually a dangerous thing to do, the Bézier curve is unusually well behaved.

The Bézier curve is available in both a simple polynomial and a rational form. The polynomial form does not have any user control except for the positioning of the control points. The rational form has control variables called *weights*. The weights may be any positive values. If the weights are all equal, the polynomial form of the Bézier curve is produced.

There is a limit on the number of control points that may be supplied to these subroutines.

3.3.1. Subroutine GZBEZR: Draw a Bézier Curve

This subroutine may be used to draw a Bézier curve of arbitrary degree determined by a sequence of points.

The calling sequence is:

```
CALL GZBEZR(N,PXA,PYA,NS)
```

The input parameters are:

- **N** An integer giving the number of control points. The maximum number of points that are allowed is 32.
- **PXA** A real array of dimension N containing the x coordinates of the control points.
- **PYA** A real array of dimension N containing the y coordinates of the control points.
- **NS** An integer giving the number of straight line segments into which the curve is to be divided.

Figures (3.9) and (3.10) show some examples of Bézier curves. Figure (3.10) illustrates the effect of moving a single control point.

3.3.2. Subroutine GZRBEZ: Draw a Rational Bézier Curve

This subroutine may be used to draw a rational Bézier curve of arbitrary degree determined by a sequence of points.

The calling sequence is:

```
CALL GZRBEZ(N,PXA,PYA,NW,WA,NS)
```

The input parameters are:

- **N** An integer giving the number of control points. The maximum number of points that are allowed is 32.
- **PXA** A real array of dimension N containing the x coordinates of the control points.
- **PYA** A real array of dimension N containing the y coordinates of the control points.

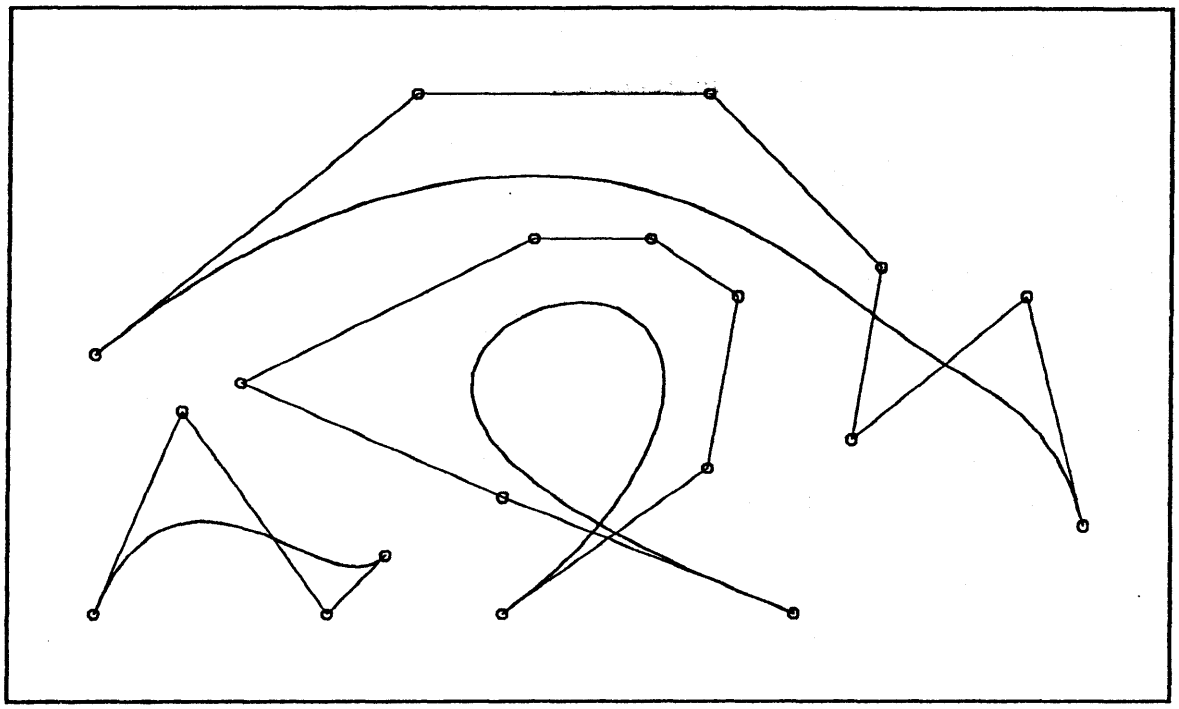

Figure 3.9. Examples of Bézier curves (I)

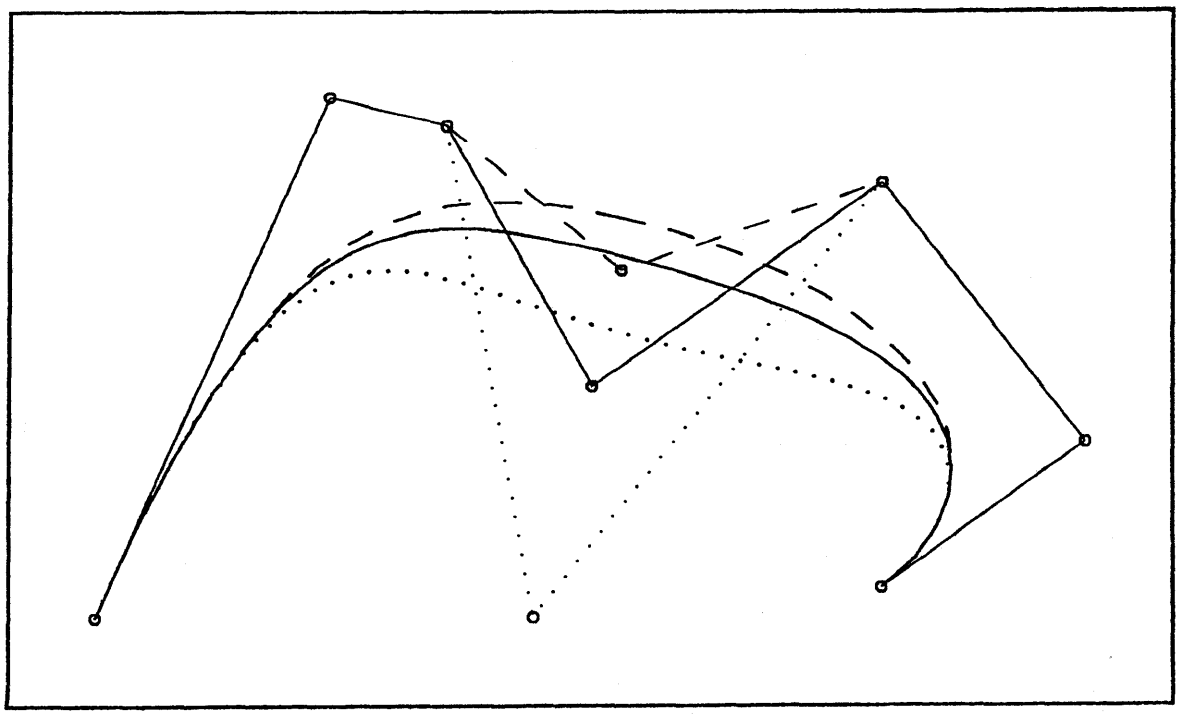

Figure 3.10. Examples of Bézier curves (II)

`NW` An integer giving the number of weights associated with the control points. This value must be positive and the weights are selected cycli-

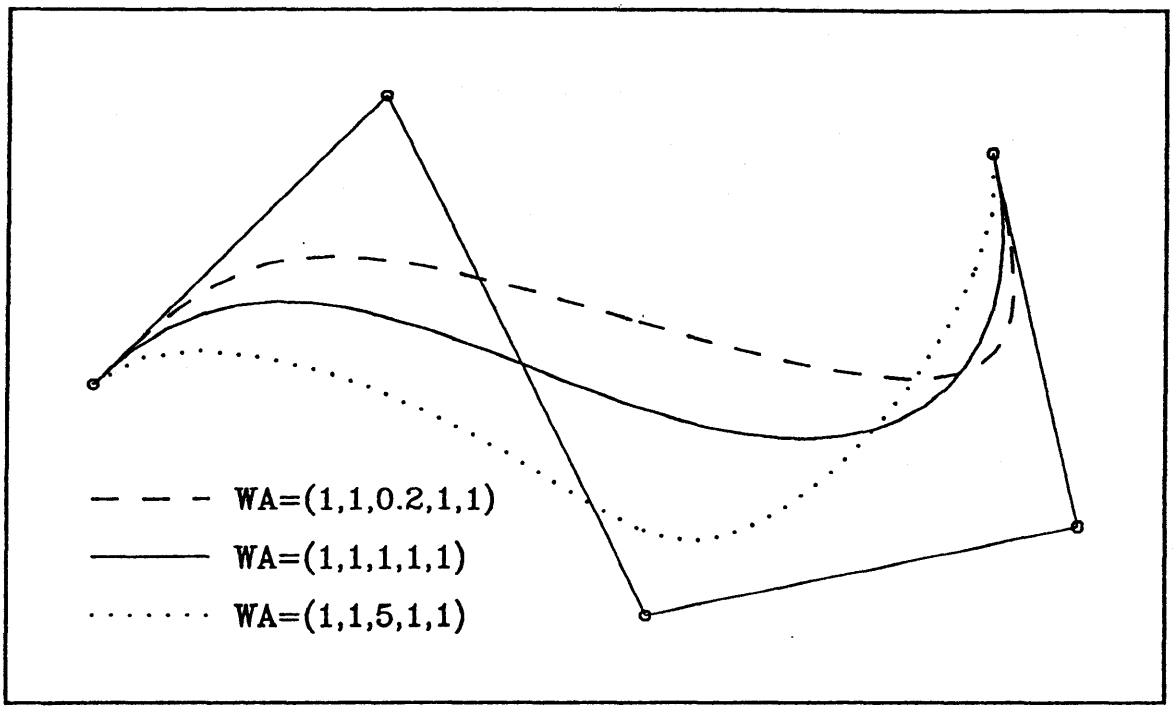

Figure 3.11. Examples of rational Bézier curves

cally from the next parameter. A total of N values are needed.

WA — A real array of dimension NW containing the given weights.

NS — An integer giving the number of straight line segments into which the curve is to be divided.

Figure (3.11) illustrates how the weights may be used to control the shape of a rational Bézier curve. In the figure, larger values of the weights cause the curve to move closer to its associated point while allowing the curve to pull away from neighboring points.

3.4. B-spline Curves

A B-spline curve is a pure design curve; it normally does not pass through any of its control points. The subroutines described here make it available in both the polynomial and rational forms in either quadratic or cubic degree. The segments of a quadratic B-spline match at the knots in ordinate and first derivative. The segments of a cubic B-spline match in ordinate, and first and second derivative. The curve also is local in nature; changing a single control point only affects a small number of curve segments.

Since the knots would not otherwise be known to the user, a facility is provided whereby the knots my be marked. This is done by calling the GKS polymarker subroutine, GPM. All of the figures in this section have had the knots marked with markers that are slightly smaller than those used for the control points.

The B-spline is actually a generalization of the Bézier curve. The proper selection of the parameter values can cause the subroutines described in this section to produce a Bézier curve.

3.4.1. Subroutine GZBSP2: Draw a Quadratic B-spline Curve

This subroutine may be used to draw a quadratic B-spline curve that is controlled by a sequence of points. Either a uniform curve, or a nonuniform curve may be drawn. In the case of a nonuniform curve, a simple means to base the parameter on accumulated chord length is provided. In addition to drawing the curve, the knots may also be marked.

The calling sequence is:

```
CALL GZBSP2(N,PXA,PYA,NP,PA,NS,MFLG)
```

The input parameters are:

N	An integer giving the number of control points.
PXA	A real array of dimension N containing the x coordinates of the control points.
PYA	A real array of dimension N containing the y coordinates of the control points.
NP	An integer giving the number of parameter values associated with line segments in the control polygon. If this value is not positive, accumulated chord length will be used to generate the parameter. If this parameter is positive, values are selected cyclically from the next parameter. In this case, a total of N values are needed.
PA	If NP is positive, this is a real array of dimension NP containing the given parameter values associated with the line segments.
NS	An integer giving the number of straight line segments into which each curve segment is to be divided.
MFLG	An integer giving a flag that indicates if the knots are to be marked. Any nonzero value will cause them to be marked.

There is, however, a problem with the generation of the PA array when accumulated chord length is used to produce it. The problem is that there are (N − 1) distances available but N values are needed. An appropriate scheme, and the one used within subroutine GZBSP2, is

$$\begin{aligned} \texttt{PA}(1) &= D_1, \\ \texttt{PA}(2) &= \tfrac{1}{2}(D_1 + D_2), \\ \texttt{PA}(3) &= \tfrac{1}{2}(D_2 + D_3), \\ &\cdots \\ \texttt{PA}(\texttt{N}-1) &= \tfrac{1}{2}(D_{\texttt{N}-2} + D_{\texttt{N}-1}), \\ \texttt{PA}(\texttt{N}) &= D_{\texttt{N}-1}. \end{aligned} \tag{3.2}$$

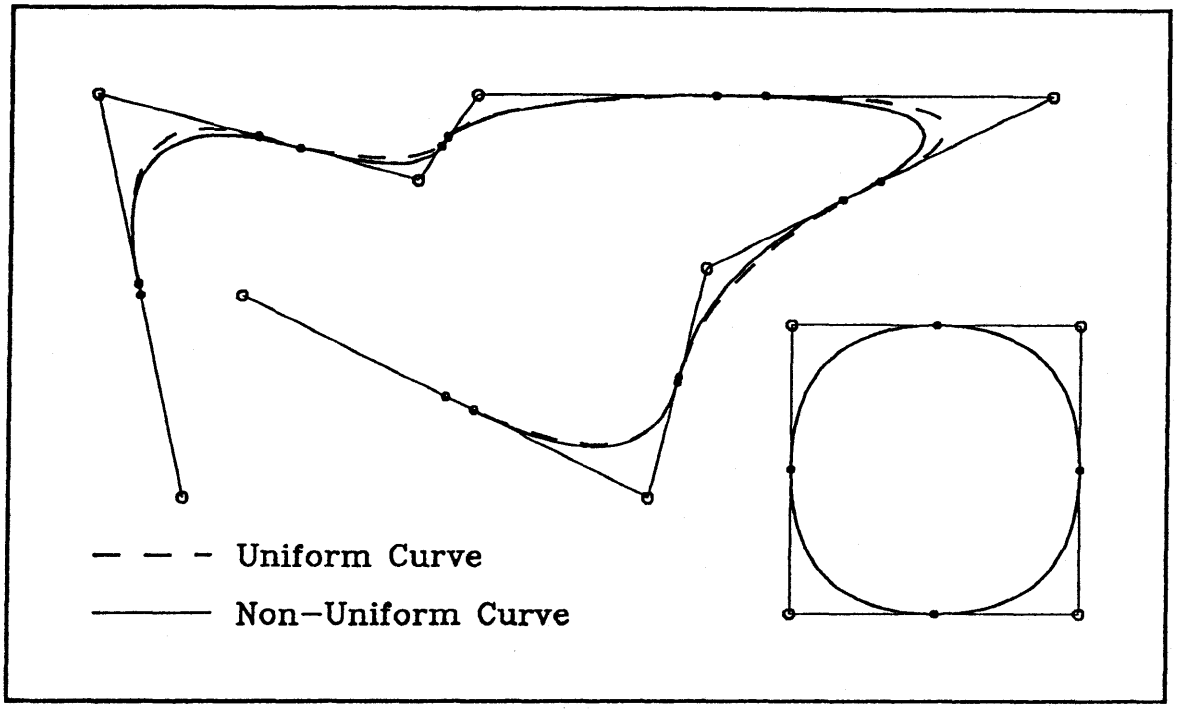

Figure 3.12. Examples of quadratic B-splines (I)

The D_i values are determined by Equations (3.1).

Figure (3.12) shows examples of uniform and nonuniform quadratic B-splines. The nearly circular curve at the lower right was formed by specifying six consecutive corner points around the square. The uniform and nonuniform curve based on chord length are equal in this case. In the other nonuniform curve, the PA values were determined from chord distances and Equations (3.2).

For the quadratic B-spline, the knots always lie on the control polygon and the curve is tangent to the control polygon at the knots. In the uniform case, the knots are at the midpoints of the line segments in the control polygon.

Figure (3.13) shows examples of how a modification of the PA values changes the curve. In this case, the PA's were also determined from Equations (3.2) and only the central one was modified. Notice how small values of this parameter cause the points of tangency on the control polygon to move closer to the associated point on the control polygon.

There is a fairly popular alternative to Equations (3.2). That alternative sets

$$\texttt{PA}(1) = 0.0,$$
$$\texttt{PA}(\texttt{N}) = 0.0,$$

with the other values set by Equations (3.2). The advantage of this scheme is that the curve now passes through the first and last control points and is tangent to the control polygon at those points. The interior of the curve has the properties described above. The problem with this formulation is that it does not reduce to the usual uniform approach.

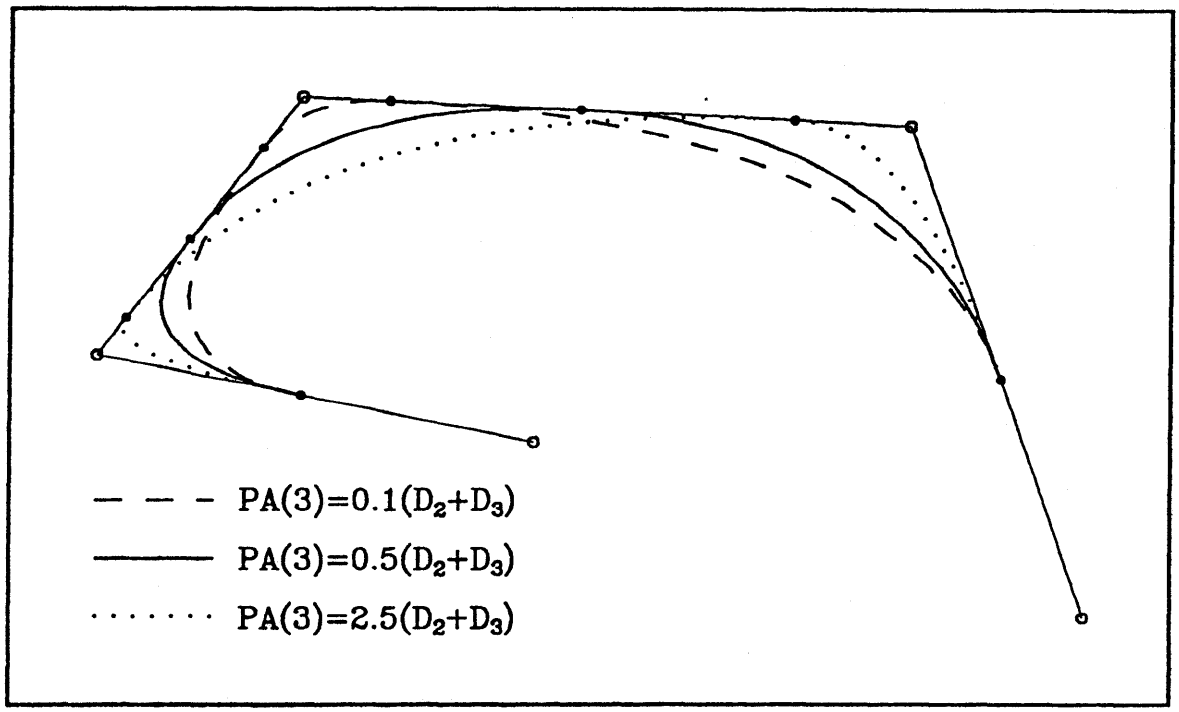

Figure 3.13. Examples of quadratic B-splines (II)

3.4.2. Subroutine GZRBS2: Draw a Rational Quadratic B-spline Curve

This subroutine may be used to draw a rational quadratic B-spline curve that is controlled by a sequence of points. Either a uniform curve, or a nonuniform curve may be drawn. In the case of a nonuniform curve, a simple means to base the line segment parameters on accumulated chord length is provided. In addition to drawing the curve, the knots may also be marked.

The calling sequence is:

```
CALL GZRBS2(N,PXA,PYA,NP,PA,NW,WA,NS,MFLG)
```

The input parameters are:

N	An integer giving the number of control points.
PXA	A real array of dimension N containing the x coordinates of the control points.
PYA	A real array of dimension N containing the y coordinates of the control points.
NP	An integer giving the number of parameter values associated with line segments in the control polygon. If this value is not positive, accumulated chord length will be used to generate the parameter. If this parameter is positive, values are selected cyclically from the next parameter. In this case, a total of N values are needed.
PA	If NP is positive, this is a real array of dimension NP containing the given parameter values associated with the line segments.

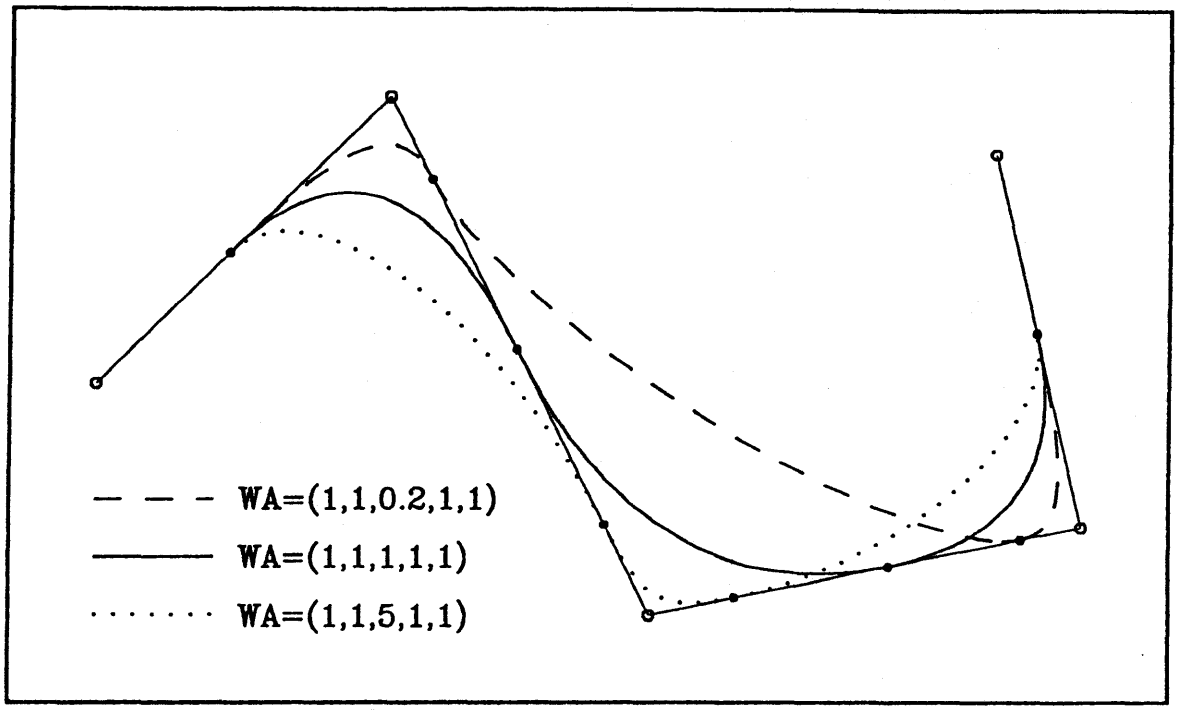

Figure 3.14. Examples of rational quadratic B-spline curves

NW	An integer giving the number of weights associated with the control points. This value must be positive and the weights are selected cyclically from the next parameter. A total of N values are needed.
WA	A real array of dimension NW containing the given weights.
NS	An integer giving the number of straight line segments into which each curve segment is to be divided.
MFLG	An integer giving a flag that indicates if the knots are to be marked. Any nonzero value will cause them to be marked.

Figure (3.14) illustrates how the weights may be used to control the shape of a rational quadratic B-spline curve. The PA values were determined by Equations (3.2). In the figure, larger values of the weights cause the curve to move closer to its associated point while allowing the curve to pull away from neighboring points.

3.4.3. Subroutine GZBSP3: Draw a Cubic B-spline Curve

This subroutine may be used to draw a cubic B-spline curve that is controlled by a sequence of points. Either a uniform curve, or a nonuniform curve may be drawn. In the case of a nonuniform curve, a simple means to base the parameter on accumulated chord length is provided. In addition to drawing the curve, the knots may also be marked.

The calling sequence is:

```
CALL GZBSP3(N,PXA,PYA,NP,PA,NS,MFLG)
```

The input parameters are:

N	An integer giving the number of control points.
PXA	A real array of dimension N containing the x coordinates of the control points.
PYA	A real array of dimension N containing the y coordinates of the control points.
NP	An integer giving the number of parameter values associated with line segments in the control polygon. If this value is not positive, accumulated chord length will be used to generate the parameter. If this parameter is positive, values are selected cyclically from the next parameter. In this case, a total of (N + 1) values are needed.
PA	If NP is positive, this is a real array of dimension NP containing the given parameter values associated with the line segments.
NS	An integer giving the number of straight line segments into which each curve segment is to be divided.
MFLG	An integer giving a flag that indicates if the knots are to be marked. Any nonzero value will cause them to be marked.

There is again a problem with the PA array when accumulated chord length is used to produce it. In this case there are (N − 1) distances available but (N + 1) values are needed. An appropriate scheme, and the one used within subroutine GZBSP3, is

$$\begin{aligned} \mathtt{PA}(1) &= D_1, \\ \mathtt{PA}(2) &= D_1, \\ \mathtt{PA}(3) &= D_2, \\ &\cdots \\ \mathtt{PA}(\mathtt{N}-1) &= D_{\mathtt{N}-2}, \\ \mathtt{PA}(\mathtt{N}) &= D_{\mathtt{N}-1}, \\ \mathtt{PA}(\mathtt{N}+1) &= D_{\mathtt{N}-1}. \end{aligned} \tag{3.3}$$

The D_i values are again determined by Equations (3.1).

Figure (3.15) shows examples of uniform and nonuniform cubic B-splines. The nearly circular curve at the lower right was formed by specifying seven consecutive corner points around the square. The uniform and nonuniform curve based on chord length are equal in this case. In the other nonuniform curve, the PA values were determined from chord distances and Equations (3.3).

Figure (3.16) shows examples of how a modification of the PA values changes the curve. In this case, the PA's were also determined from Equations (3.3) and only the central one was modified. Small values of this parameter cause the central curve segment to shrink and move closer to the associated segment of the control polygon.

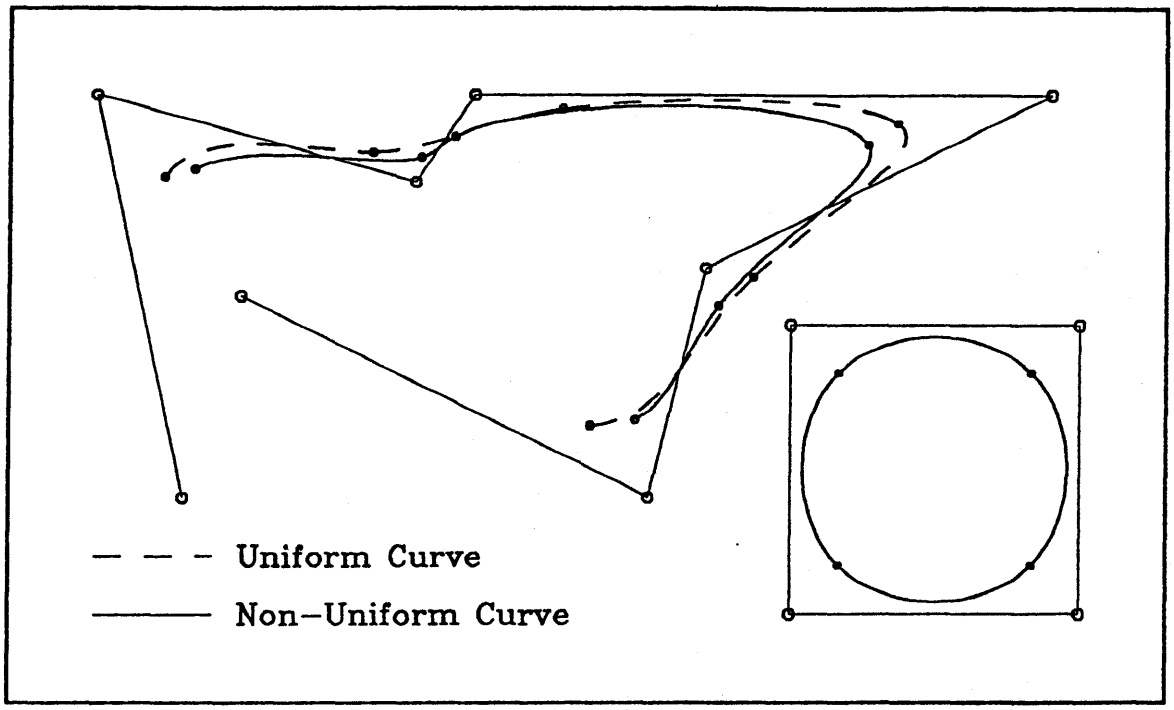

Figure 3.15. Examples of cubic B-splines (I)

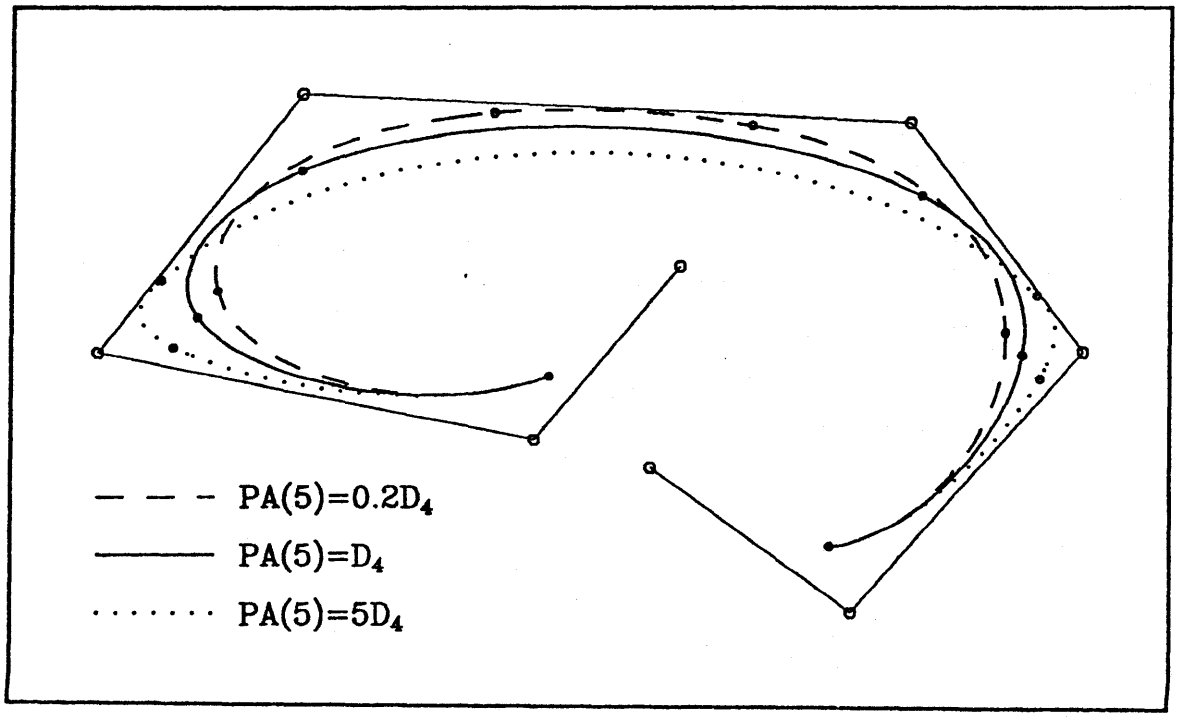

Figure 3.16. Examples of cubic B-splines (II)

As in the quadratic case, there is a popular alternative to Equations (3.3). That

alternative sets

$$\begin{aligned} \texttt{PA}(1) &= 0.0, \\ \texttt{PA}(2) &= 0.0, \\ \texttt{PA}(\texttt{N}) &= 0.0, \\ \texttt{PA}(\texttt{N}+1) &= 0.0. \end{aligned}$$

This scheme again forces the curve to pass through the first and last control points and makes it tangent to the control polygon at those points.

3.4.4. Subroutine GZRBS3: Draw a Rational Cubic B-spline Curve

This subroutine may be used to draw a rational cubic B-spline curve that is controlled by a sequence of points. Either a uniform curve, or a nonuniform curve may be drawn. In the case of a nonuniform curve, a simple means to base the parameter on accumulated chord length is provided. In addition to drawing the curve, the knots may also be marked.

The calling sequence is:

```
CALL GZRBS3(N,PXA,PYA,NP,PA,NW,WA,NS,MFLG)
```

The input parameters are:

N	An integer giving the number of control points.
PXA	A real array of dimension N containing the x coordinates of the control points.
PYA	A real array of dimension N containing the y coordinates of the control points.
NP	An integer giving the number of parameter values associated with line segments in the control polygon. If this value is not positive, accumulated chord length will be used to generate the parameter. If this parameter is positive, values are selected cyclically from the next parameter. In this case, a total of (N + 1) values are needed.
PA	If NP is positive, this is a real array of dimension NP containing the given parameter values associated with the line segments.
NW	An integer giving the number of weights associated with the control points. This value must be positive and the weights are selected cyclically from the next parameter. A total of N values are needed.
WA	A real array of dimension NW containing the given weights.
NS	An integer giving the number of straight line segments into which each curve segment is to be divided.
MFLG	An integer giving a flag that indicates if the knots are to be marked. Any nonzero value will cause them to be marked.

Figure (3.17) illustrates how the weights may be used to control the shape of a rational cubic B-spline curve. The PA values were determined by Equations (3.3).

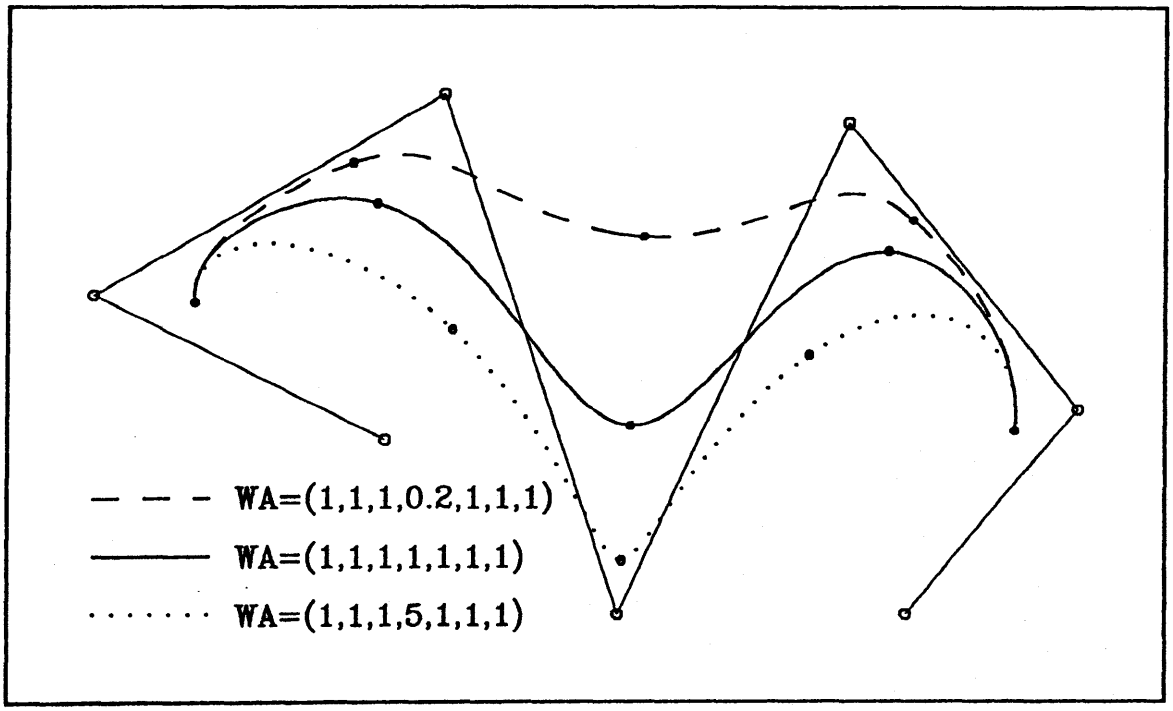

Figure 3.17. Examples of rational cubic B-spline curves

In the figure, larger values of the weights cause the curve to move closer to its associated point while allowing the curve to pull away from neighboring points.

Chapter 4

Surface Drawing Algorithms

This chapter describes a group of subroutines that may be used to draw surfaces. The surfaces are defined by supplying control points and other control information to the subroutines. The surfaces are drawn by breaking them down into simple polygons, eliminating those polygons that face away from the viewer, sorting the remainder so that the ones farthest away are first on the list, and then calling the GKS fill area subroutine, GFA, to write the polygons to the active workstations in the sorted order. Closer polygons, therefore, overlay the farther ones. If the workstation does not overlay older graphic primitives when a fill area is written, these subroutines will not work properly.

This method does have its problems. It is possible, especially when polygons of vastly differing sizes are involved, to have a large polygon to be determined to be "closer" than a small one even though the small one actually hides part of the larger. The subroutine in this chapter that deals with generalized polyhedral solids is especially vulnerable to this problem, particularly if non-convex polygons are supplied. It is also important that the polygons do not intersect each other; none of the algorithms described here can handle that problem.

There are two ways that the polygons may always be drawn so that useful pictures are produced. In the simplest method, the polygons are drawn as hollow fill areas. When this is done the pictures look like line drawn figures. A second way is to apply a light source and reflection model to obtain fairly realistic pictures. This method will only be successful on workstations that can produce a large number of colors. If the workstation only supports a small number of colors, the fill areas will all blend together and the picture will be unintelligible. In addition to these two general modes, certain algorithms may supply other options.

To understand the light source and reflection model used in these subroutines, consider Figure (4.1). This figure shows a point, **P**, on the surface and the light source and eye point. **N** is the surface normal at **P**, **L** is a vector pointing from **P** to the light source, and **E** is a vector pointing toward the eye position. **R** represents a light ray that starts at the light source and reflects off the surface. The vectors **L**, **N**, and **R** are coplanar and **L** and **R** make the same angle, θ, with **N**. The vector **E** makes an angle of α with **R**. Notice that **E** is not necessarily coplanar with **L**, **N**, and **R**.

The light source and reflection model used is

$$I = I_0 + \frac{K_d \cos\theta + K_s \cos^n \alpha}{d + K}$$

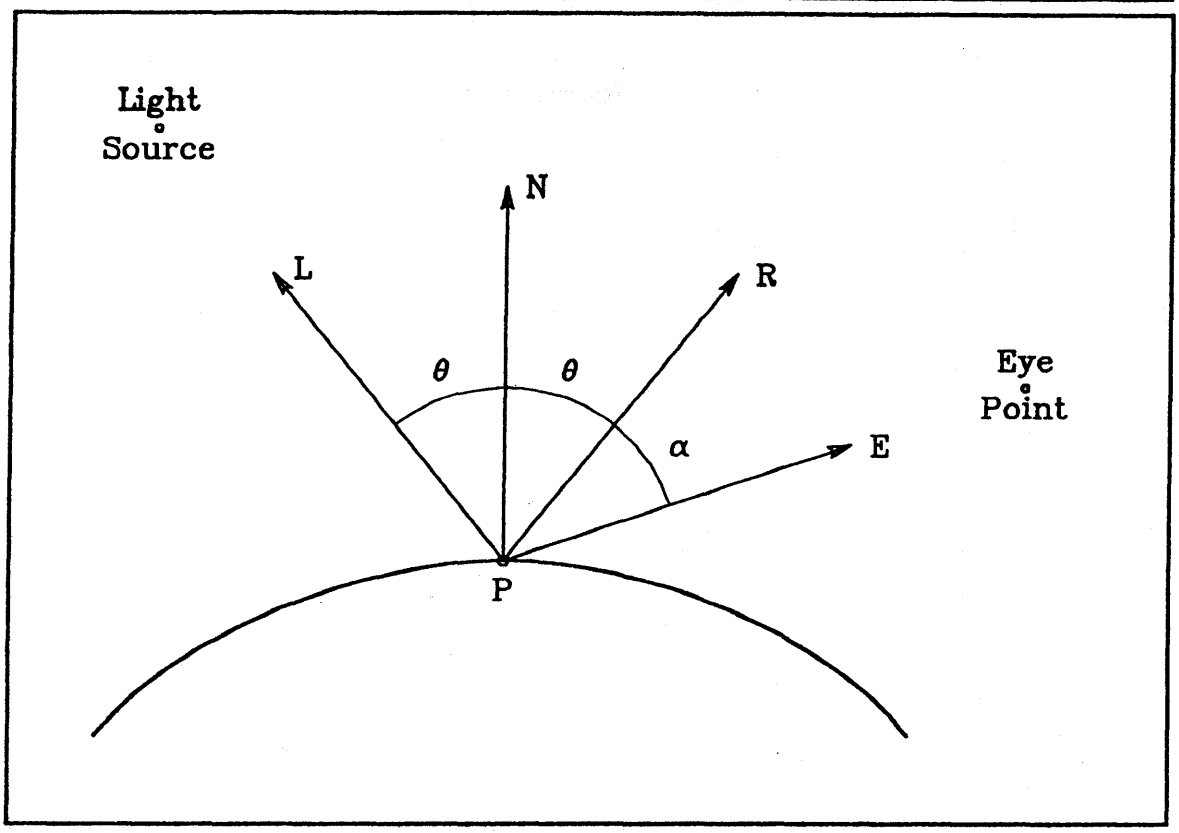

Figure 4.1. The light source and reflection model

where I_0, n, K_d, K_s, and K are parameters that the user may set. The computed value, I, is known as the *shading function* for the point. The value d is the projected distance from the eye point to the surface; its value is one on the projection plane and zero at the eye point. I_0 is the ambient illumination for the scene. K_d is the diffuse reflection constant and K_s is the specular reflection constant. Highlights on a shiny object are caused by specular reflection. Large values of n cause an object to appear shiny. A complete derivation of the model is given in Section 5.2 of *Procedural Elements for Computer Graphics* [Rog85].

The parameters of the light source and reflection model are given in a real array, `CCA`, that is common to all of the subroutines. The description of the array in this case is as follows:

- `CCA(1)` To specify this option, this value should contain a real value of one.
- `CCA(2)` A real value which specifies the GKS color index to be used to draw the polygons. This value will be converted to an integer before it is used.
- `CCA(3)` The red color value for the surface.
- `CCA(4)` The green color value for the surface.
- `CCA(5)` The blue color value for the surface.
- `CCA(6)` The x component of a vector in the direction of the light rays. That is, the direction is from the light source toward the object.

CCA(7) The y component of a vector in the direction of the light rays.
CCA(8) The z component of a vector in the direction of the light rays.
CCA(9) The ambient illumination, I_0.
CCA(10) The specular reflection exponent, n.
CCA(11) The diffuse reflection constant, K_d.
CCA(12) The specular reflection constant, K_s.
CCA(13) The distance adjustment constant, K.

Notice that the direction of the light rays as given by CCA(6), ..., CCA(8) is the reverse of that shown by the vector L in Figure (4.1). It is important to get the direction correct; it is no help if the light is shining on the bottom of the model when you expected it on the top. When the shading function value, I, is multiplied by the color values for the surface, the resulting values must be between zero and one. If they are not, they will be set to zero or one, whichever is appropriate. The values of these parameters can be difficult to select. In the absence of other information, a good place to start is $I_0 = 0.1$, $n = 2.0$, $K_d = 1.5$, $K_s = 0.3$, and $K = 1.0$.

Each of the subroutines also needs a work array. This is a real array that is used to sort the polygons. The required size depends on the problem but a maximum value is usually easy to obtain.

From the above discussion, it is apparent the there are two things that are difficult to determine when these subroutines are used. The first of these problems is the coefficients of the shading function and the second is the size of the work array. To aid in the use of these subroutines, the maximum and minimum values of the shading function and the actual size of the work array that was needed are made available to the user. These results are put into a COMMON block whose declaration is

```
C  COMMON BLOCK TO RETURN SURFACE INFORMATION.
       SAVE             /GZSINF/
       COMMON           /GZSINF/GZLMAX,GZSMIN,GZSMAX
C     MAXIMUM LENGTH OF THE WORK AREA THAT WAS USED.
       INTEGER           GZLMAX
C     MINIMUM AND MAXIMUM VALUES OF THE SHADING FUNCTION.
       REAL              GZSMIN,GZSMAX
```

The COMMON block is available after one of these subroutines has been called. If the light source and reflection model was not used, GZSMIN and GZSMAX will be zero.

The view of the surface is selected by specifying a three-dimensions to two-dimensions projective transformation. That transformation must be a perspective transformation; it cannot be a parallel transformation.

These subroutines are quite efficient when processing on the host computer only is considered. However, the amount of data that must be transmitted to the workstation can be quite large and many workstations require substantial amounts of time to process fill areas. In essence, these subroutines off-load much of the computation from the host computer to the graphic device itself.

If one of these subroutines detects an error in the data supplied to it, the subroutine prints an error message and returns, usually without producing any graphic output.

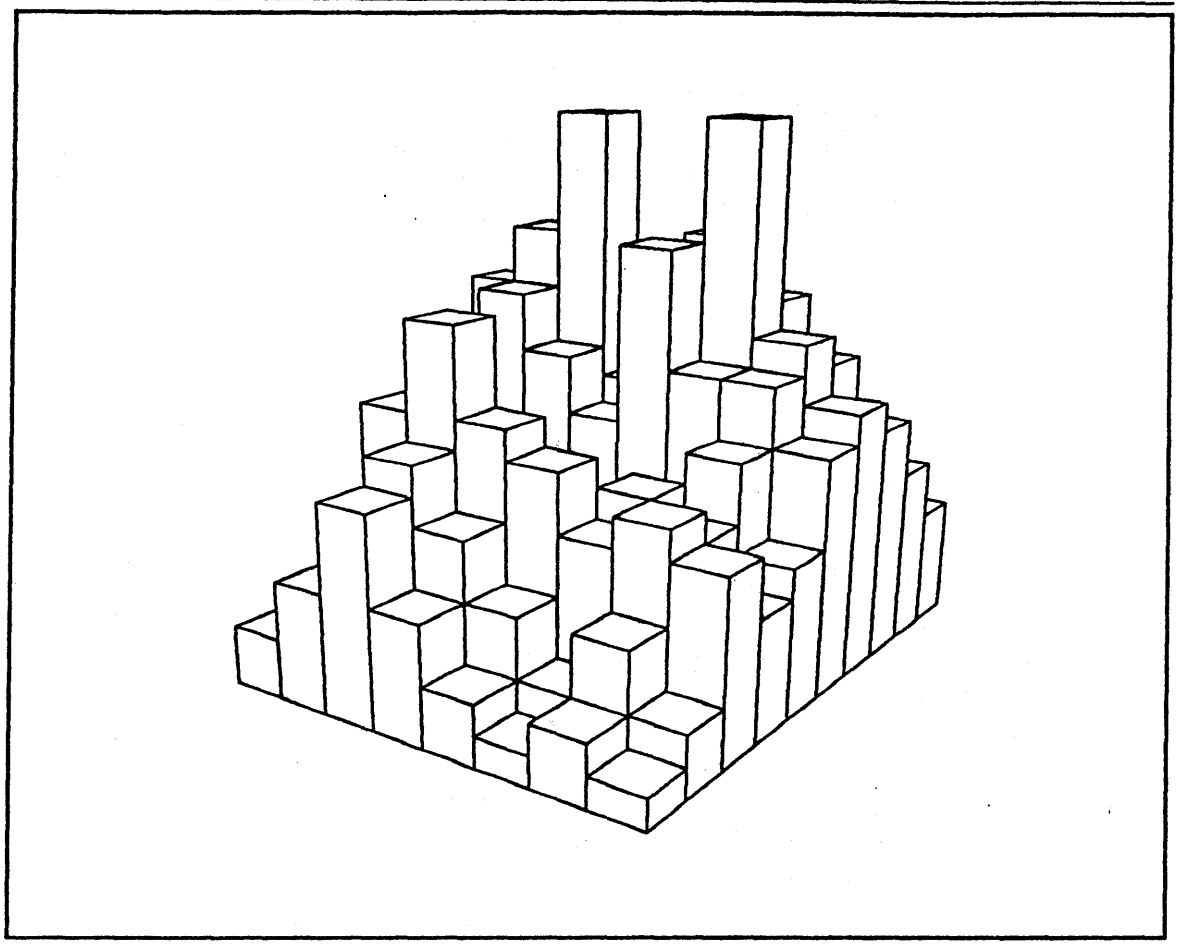

Figure 4.2. A two-dimensional histogram

4.1. Two-dimensional Histograms

A two-dimensional histogram consists of a rectangular array of rectangular columns sitting on a common base. The height of the columns can be used to represent experimental data. Pictures of this type are sometimes called *Lego* plots.

4.1.1. Subroutine GZ2DHG: Draw a Two-Dimensional Histogram

This subroutine may be used to draw a two-dimensional histogram.

The polygons that constitute the histogram may be drawn in one of three ways. In the first scheme, corresponding sides on each column are drawn in a distinct color. In the second scheme, the normal GKS-EZ setting is used to draw the polygons. This should normally be done using hollow polygons. The third scheme provides a light source and reflection model to color the polygons.

The calling sequence is:

```
CALL GZ2DHG(M,N,PXYZA,PTRN,CCA,L,WA)
```

The input parameters are:

M An integer giving the first dimension of PXYZA.

N An integer giving the second dimension of PXYZA.

PXYZA A real array of dimension (M,N) containing the x, y, and z coordinates of the two-dimensional histogram. The format of the array is

$$\begin{pmatrix} z_0 & x_1 & x_2 & \cdots & x_{N-2} & x_{N-1} \\ y_1 & z_{1,1} & z_{2,1} & \cdots & z_{N-2,1} & - \\ y_2 & z_{1,2} & z_{2,2} & \cdots & z_{N-2,2} & - \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ y_{M-2} & z_{1,M-2} & z_{2,M-2} & \cdots & z_{N-2,M-2} & - \\ y_{M-1} & - & - & \cdots & - & - \end{pmatrix}.$$

The sequences $(x_1, x_2, \ldots, x_{N-1})$ and $(y_1, y_2, \ldots, y_{M-1})$ must be monotonically increasing but do not have to be equally spaced. The two-dimensional histogram consists of $(N-2)$ columns in the x direction and $(M-2)$ columns in the y direction. The value z_0 is the z coordinate of the base of the columns. The bounds of the (i,j)th column $(i = 1,\ldots,(N-2);\ j = 1,\ldots,(M-2))$ are x_i to x_{i+1} in x and y_j to y_{j+1} in y. This means that the last row and column of PXYZA are almost unused. These unused values are shown as dashes in the matrix. The $z_{i,j}$ values give the z coordinates of the tops of the columns and should not be smaller than z_0.

PTRN A real array of dimension (3,4) containing the perspective transformation.

CCA A real array containing the color control for the two-dimensional histogram. The value of CCA(1) selects one of three possibilities:

- CCA(1)=-1.0: This means each side of a column is to be colored in a distinct color. Additional data is supplied in the array as described below.
- CCA(1)=0.0: This means the color setting defined by the GKS-EZ subroutine GZSFAA is to be used. No additional data is supplied in the array.
- CCA(1)=1.0: This means that the columns are to be colored using the light source and reflection model. Additional data is supplied in the array as described earlier.

L An integer giving the length of the work array.

WA A real array of dimension L that will be used as a work array. L should be at least $6(M-2)(N-2)$.

This subroutine allows the special coloring scheme defined by a value of CCA(1) equal to minus one. The description of the CCA array in this case is as follows:

CCA(1) To specify this option, this value should contain a real value of minus one.

CCA(2) A real value which specifies the GKS color index to be used to draw the polygons. This value will be converted to an integer before it is used.

CCA(3) The red color value for the x_{min} or x_{max} sides of the columns.

CCA(4) The green color value for the x_{min} or x_{max} sides of the columns.

CCA(5) The blue color value for the x_{min} or x_{max} sides of the columns.

CCA(6) The red color value for the y_{min} or y_{max} sides of the columns.

CCA(7) The green color value for the y_{min} or y_{max} sides of the columns.

CCA(8) The blue color value for the y_{min} or y_{max} sides of the columns.

CCA(9) The red color value for the z_{min} or z_{max} sides of the columns.

CCA(10) The green color value for the z_{min} or z_{max} sides of the columns.

CCA(11) The blue color value for the z_{min} or z_{max} sides of the columns.

All of the color values must be between zero and one.

Figure (4.2) shows an example of a two-dimensional histogram. Like all of the examples in this chapter, it was drawn using hollow fill areas.

4.2. Mesh Surfaces

A mesh surface consists of a rectangular sheet positioned above a rectangular area in the x-y plane. The sheet is divided into smaller rectangular or triangular patches. The height of the corners of the patches of the sheet can be used to represent experimental data.

If the data suppled to the mesh surface subroutine is not relatively smooth, the resulting picture may be difficult to interpret. In this case, a two-dimensional histogram may be more appropriate.

4.2.1. Subroutine GZMESH: Draw a Mesh Surface

This subroutine may be used to draw a mesh surface. The mesh may be constructed by drawing rectangles or splitting each rectangle into a pair of triangles. Either the upper side, lower side, or both sides of the surface may be drawn. When only one side of the surface is drawn, a *skirt* is drawn around the base.

The polygons that constitute the surface may be drawn in one of two ways. In the first scheme, the normal GKS-EZ setting is used to draw the polygons. This should normally be done using hollow polygons. The second scheme provides a light source and reflection model to color the polygons.

The calling sequence is:

```
CALL GZMESH(M,N,PXYZA,SFLG,MFLG,PTRN,CCA,L,WA)
```

The input parameters are:

M An integer giving the first dimension of PXYZA.

N An integer giving the second dimension of PXYZA.

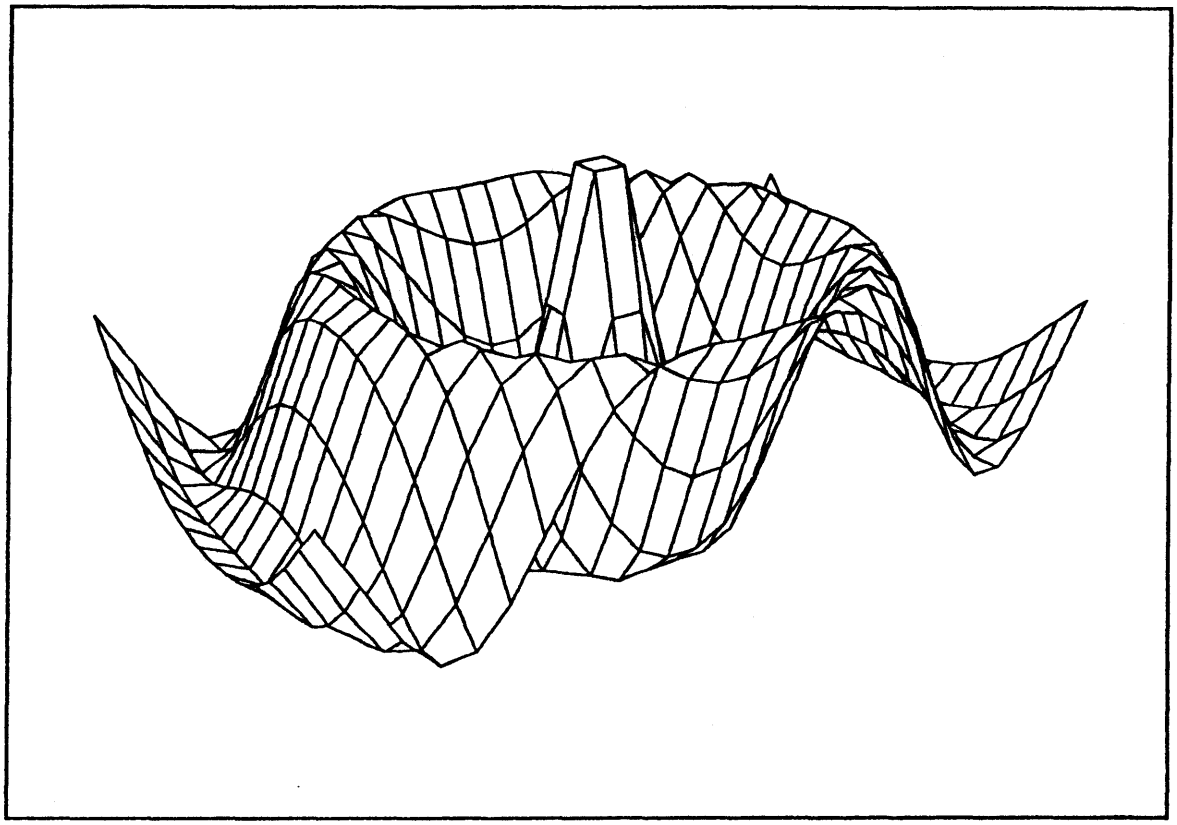

Figure 4.3. A mesh surface showing both upper and lower sides

PXYZA A real array of dimension (M,N) containing the x, y, and z coordinates of the mesh surface. The format of the array is

$$\begin{pmatrix} z_0 & x_1 & x_2 & \cdots & x_{N-2} & x_{N-1} \\ y_1 & z_{1,1} & z_{2,1} & \cdots & z_{N-2,1} & z_{N-1,1} \\ y_2 & z_{1,2} & z_{2,2} & \cdots & z_{N-2,2} & z_{N-1,2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ y_{M-2} & z_{1,M-2} & z_{2,M-2} & \cdots & z_{N-2,M-2} & z_{N-1,M-2} \\ y_{M-1} & z_{1,M-1} & z_{2,M-1} & \cdots & z_{N-2,M-1} & z_{N-1,M-1} \end{pmatrix}.$$

The sequences $(x_1, x_2, \ldots, x_{N-1})$ and $(y_1, y_2, \ldots, y_{M-1})$ must be monotonically increasing but do not have to be equally spaced. The mesh surface consists of $(N-2)$ surface elements in the x direction and $(M-2)$ surface elements in the y direction. The bounds of the (i,j)th surface element $(i = 1, \ldots, (N-2);\ j = 1, \ldots, (M-2))$ are x_i to x_{i+1} in x and y_j to y_{j+1} in y. The $z_{i,j}$ values give the z coordinates of the corners of the rectangular surface elements. The value z_0 is the z coordinate of the base of the structure and is only used if a skirt is being drawn. When a skirt is drawn for the upper side of the surface, z_0 must not be greater than any of the $z_{i,j}$ values. When a skirt is drawn for the lower side of the surface, z_0 must not be smaller than any of the $z_{i,j}$ values.

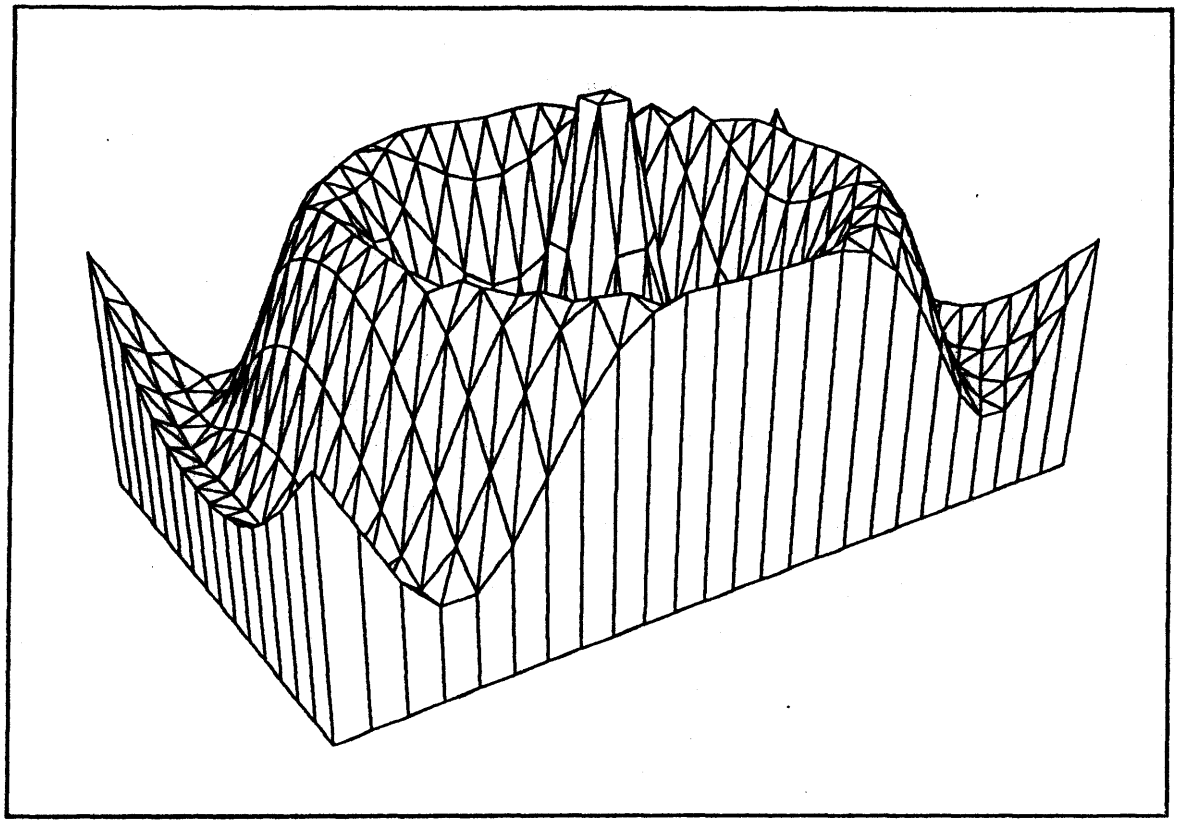

Figure 4.4. A mesh surface showing the upper side only

SFLG An integer specifying which side of the surface is to be drawn. A positive value means the upper side is to be drawn while a negative value means the lower side. A zero value means both sides are to be drawn.

MFLG An integer specifying the type of mesh to be drawn. A zero value means rectangles are to be drawn while nonzero values mean triangles are to be drawn. A positive value means the dividing line for the triangles will pass through $z_{1,1}$ and a negative value means it will not.

PTRN A real array of dimension (3,4) containing the perspective transformation.

CCA A real array containing the color control for the mesh surface. The value of CCA(1) selects one of two possibilities:

CCA(1)=0.0: This means the color setting defined by the GKS-EZ subroutine GZSFAA is to be used. No additional data is supplied in the array.

CCA(1)=1.0: This means that the surface elements are to be colored using the light source and reflection model. Additional data is supplied in the array as described earlier.

L An integer giving the length of the work array.

WA A real array of dimension L that will be used as a work array. The

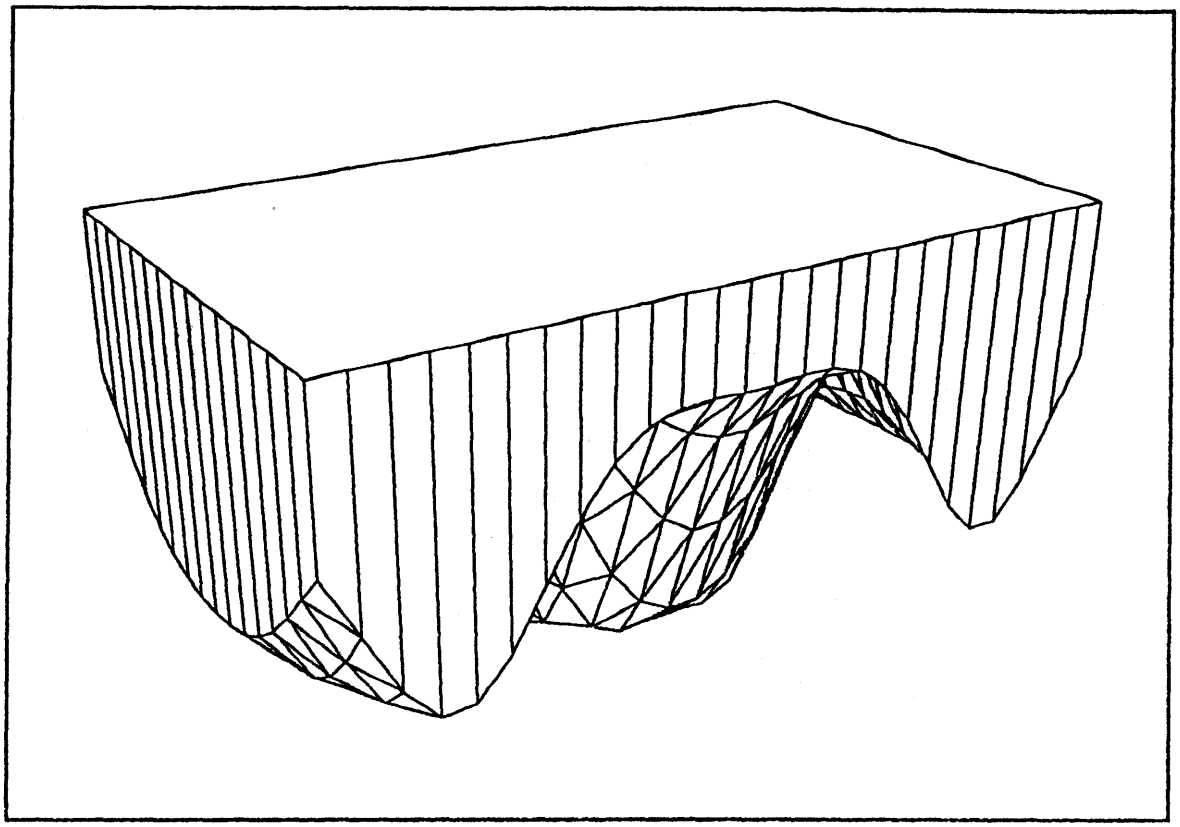

Figure 4.5. A mesh surface showing the lower side only

required size of WA is difficult to estimate. The maximum size is easier to compute. Start with $(M-2)(N-2)$. If triangles are being drawn, double that value. If both top and bottom are being drawn, double that value again. If only the top or bottom is being drawn, add $2(M-2)+(N-2)+1$. Finally, double that value. This value overestimates the number of words needed; the actual number will usually be about half this maximum number.

Figures (4.3), (4.4), and (4.5) all illustrate examples of mesh surfaces. In Figure (4.3) the mesh surface was broken down into rectangles, while the other two figures use triangles. In Figure (4.4) the triangular division goes through the $z_{1,1}$ point while in Figure (4.5) it does not.

4.3. Generalized Polyhedral Solids

This section describes a subroutine that can be used to draw any figure that can be broken down into planar polygons. Normally the polygons should be organized into solid polyhedra because, at most, only one side of each polygon will be drawn.

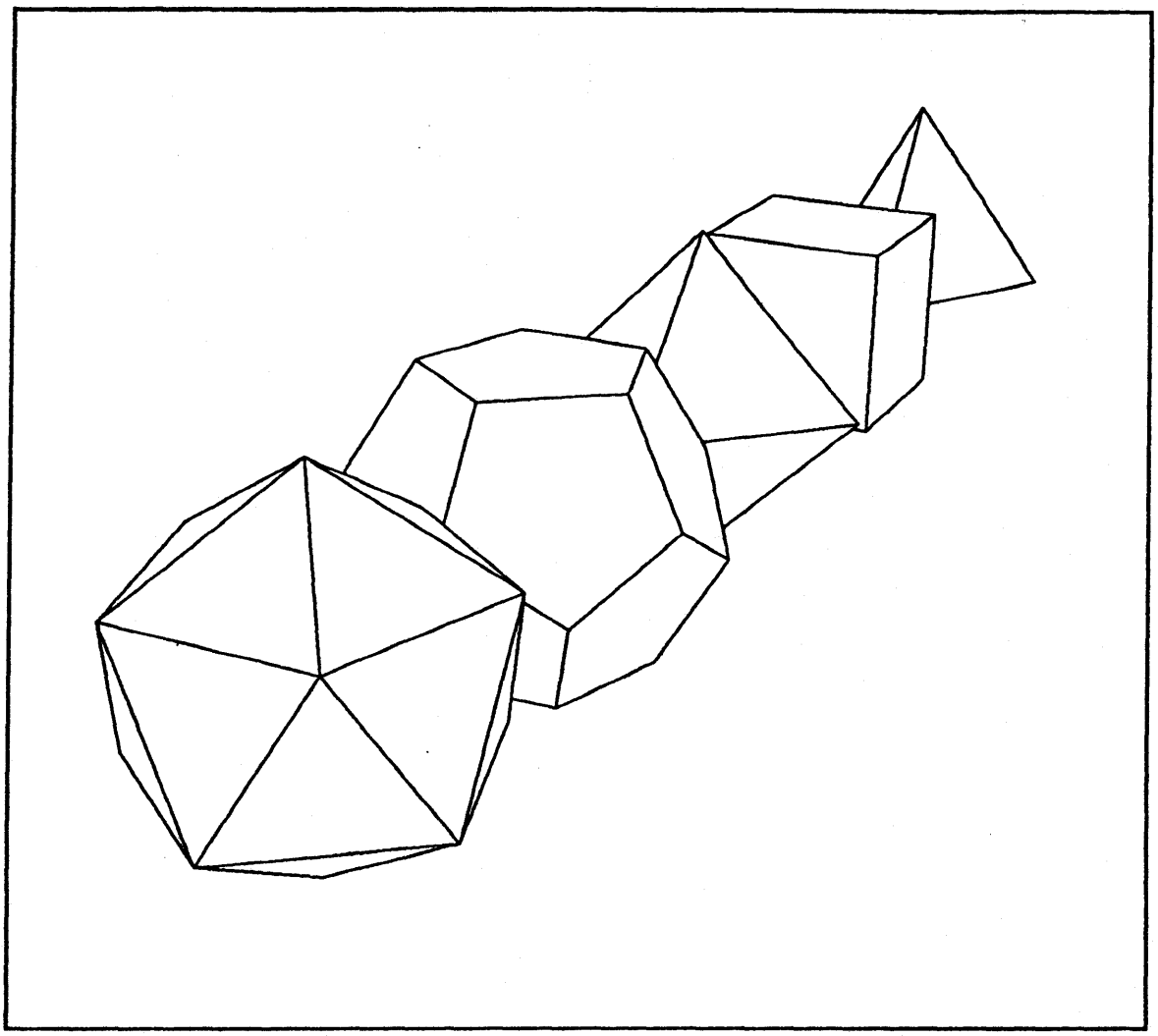

Figure 4.6. The five Platonic solids

4.3.1. Subroutine GZPOLY: Draw Generalized Polyhedra

This subroutine may be used to draw a group of polyhedra. A polyhedron consists of a solid body bounded by polygonal faces. The polygons should be planar or very nearly so. The polygons must also be nonintersecting. The points on the boundary should be ordered in such a manner that the polygon is to the left as one traverses the outside of the surface in the given order of the bounding points.

The polygons that constitute the polyhedra may be drawn in one of two ways. In the first scheme, the normal GKS-EZ setting is used to draw the polygons. This should normally be done using hollow polygons. The second scheme provides a light source and reflection model to color the polygons.

The calling sequence is:

```
CALL GZPOLY(PXA,PYA,PZA,M,NPA,IPA,IXA,PTRN,CCA,L,WA)
```

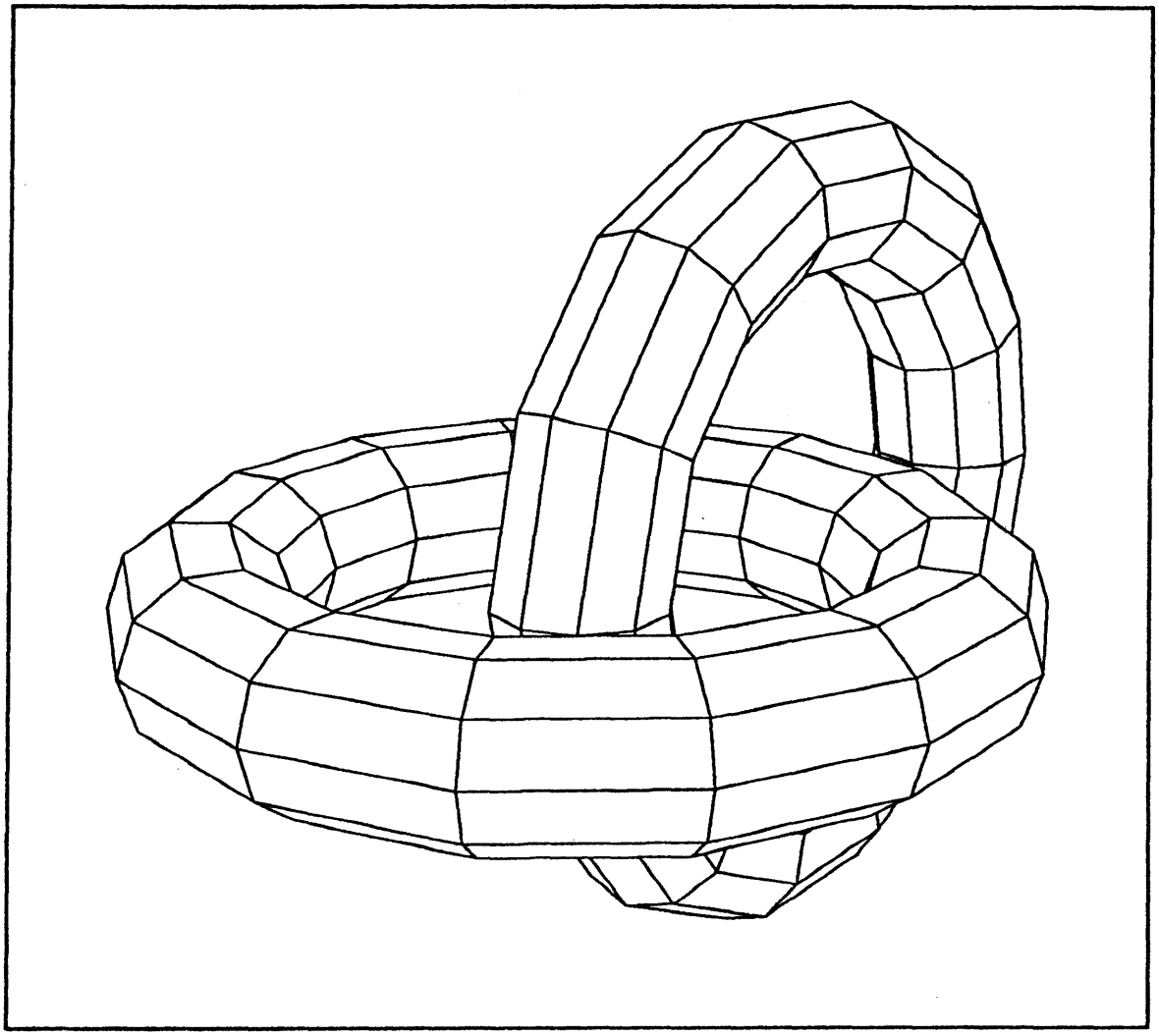

Figure 4.7. Two interlocked tori

The input parameters are:

PXA	A real array of containing the x coordinates of the points on the polyhedra.
PYA	A real array of containing the y coordinates of the points on the polyhedra.
PZA	A real array of containing the z coordinates of the points on the polyhedra.
M	An integer giving the number of polygons in the polyhedra.
NPA	An integer array of dimension M containing the number of points in each of the polygons. The maximum number of points allowed in each polygon is 16. However, there is no need to close the polygon; a triangular polygon may be defined by giving only three points.
IPA	An integer array of dimension M containing a pointer into the IXA array that gives the starting index of the indices pointing to the coordinates

of the points in the PXA, PYA, and PZA arrays bounding the polygon.

IXA — An integer array containing the indices of the points bounding the polygons.

PTRN — A real array of dimension (3,4) containing the perspective transformation.

CCA — A real array containing the color control for the polyhedra. The value of CCA(1) selects one of two possibilities:

- CCA(1)=0.0: This means the color setting defined by the GKS-EZ subroutine GZSFAA is to be used. No additional data is supplied in the array.
- CCA(1)=1.0: This means that the polygons are to be colored using the light source and reflection model. Additional data is supplied in the array as described earlier.

L — An integer giving the length of the work array.

WA — A real array of dimension L that will be used as a work array. The maximum size that will ever be required is 2M. The actual number needed will usually be about half this number.

Figures (4.6) and (4.7) illustrate two applications of this subroutine. Notice that Figure (4.6) could have been produced in two distinct ways. In the first case, a single call could be made to subroutine GZPOLY supplying it with all of the data necessary to draw all five solids. A second way that the figure could have been produced is to draw each of the five solids in turn starting with the farthest from the viewer; first the tetrahedron, then the cube, octahedron, dodecahedron, and finally the icosahedron. Since the farther solids do not hide the nearer ones, either method will produce exactly the same result. This second method will be slightly more efficient because the subroutine always has smaller files to sort. This shortcut will not work in producing Figure (4.7) because each of the tori hides part of the other; the entire figure must be produced in a single call to GZPOLY.

References

The following list contains more information about the books and reports that have been referenced in this document.

[ANS78] *American National Standard: Programming Language FORTRAN*, Document ANSI X3.9–1978, American National Standards Institute, Inc., New York, April 1978.

[ANS85a] *American National Standard for Information Systems: Computer Graphics – Graphical Kernel System (GKS) Functional Description*, Document ANSI X3.124–1985, American National Standards Institute, Inc., New York, June 1985.

[ANS85b] *American National Standard for Information Systems: Computer Graphics – Graphical Kernel System (GKS) FORTRAN Binding*, Document ANSI X3.124.1–1985, American National Standards Institute, Inc., New York, June 1985.

[Bea87] Robert C. Beach, *GKS-EZ Programming Manual for FORTRAN-77*, Stanford Linear Accelerator Center, CGTM Number 212, August 1987, Revised November 1988 and June 1991.

[Bea91] Robert C. Beach, *An Introduction to the Curves and Surfaces of Computer-Aided Design*, Van Nostrand Reinhold, New York, 1991.

[Rog85] David F. Rogers, *Procedural Elements for Computer Graphics* McGraw-Hill Book Company, New York, 1985.