U. S. DEPARTMENT OF COMMERCE
Maurice H. Stans, Secretary
ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION
Robert M. White, Administrator
RESEARCH LABORATORIES
George S. Benton, Director

ESSA TECHNICAL REPORT ERL 101-ITS 73

A Method of Smooth Curve Fitting

HIROSHI AKIMA

Ionospheric Telecommunications Laboratory

INSTITUTE FOR TELECOMMUNICATION SCIENCES
BOULDER, COLORADO
January 1969

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402
Price 40 cents.

ESSA RESEARCH LABORATORIES

The mission of the Research Laboratories is to study the oceans, inland waters, the lower and upper atmosphere, the space environment, and the earth, in search of the understanding needed to provide more useful services in improving man's prospects for survival as influenced by the physical environment. Laboratories contributing to these studies are:

Earth Sciences Laboratories: Geomagnetism, seismology, geodesy, and related earth sciences; earthquake processes, internal structure and accurate figure of the Earth, and distribution of the Earth's mass.

Atlantic Oceanographic Laboratories and Pacific Oceanographic Laboratories: Oceanography, with emphasis on ocean basins and borders, and oceanic processes; sea-air interactions; and land-sea interactions. (Miami, Florida)

Atmospheric Physics and Chemistry Laboratory: Cloud physics and precipitation; chemical composition and nucleating substances in the lower atmosphere; and laboratory and field experiments toward developing feasible methods of weather modification.

Air Resources Laboratories: Diffusion, transport, and dissipation of atmospheric contaminants; development of methods for prediction and control of atmospheric pollution. (Silver Spring, Maryland)

Geophysical Fluid Dynamics Laboratory: Dynamics and physics of geophysical fluid systems; development of a theoretical basis, through mathematical modeling and computer simulation, for the behavior and properties of the atmosphere and the oceans. (Princeton, New Jersey)

National Hurricane Research Laboratory: Hurricanes and other tropical weather phenomena by observational, analytical, and theoretical means; hurricane modification experiments to improve understanding of tropical storms and prediction of their movement and severity. (Miami, Florida)

National Severe Storms Laboratory: Tornadoes, squall lines, thunderstorms, and other severe local convective phenomena toward achieving improved methods of forecasting, detecting, and providing advance warnings. (Norman, Oklahoma)

Space Disturbances Laboratory: Nature, behavior, and mechanisms of space disturbances; development and use of techniques for continuous monitoring and early detection and reporting of important disturbances.

Aeronomy Laboratory: Theoretical, laboratory, rocket, and satellite studies of the physical and chemical processes controlling the ionosphere and exosphere of the earth and other planets.

Wave Propagation Laboratory: Development of new methods for remote sensing of the geophysical environment; special emphasis on propagation of sound waves, and electromagnetic waves at millimeter, infrared, and optical frequencies.

Institute for Telecommunication Sciences: Central federal agency for research and services in propagation of radio waves, radio properties of the earth and its atmosphere, nature of radio noise and interference, information transmission and antennas, and methods for the more effective use of the radio spectrum for telecommunications.

Research Flight Facility: Outfits and operates aircraft specially instrumented for research; and meets needs of ESSA and other groups for environmental measurements for aircraft. (Miami, Florida)

ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION
BOULDER, COLORADO

TABLE OF CONTENTS

	Page
ABSTRACT	1
1. INTRODUCTION	1
2. DISCUSSION OF SOME EXISTING METHODS	2
3. NEW METHOD	7
3.1. Outline of the Method	7
3.2. Direction of the Tangent	8
3.3. Interpolation Between a Pair of Points	11
3.4. Extrapolation of the Curve at an End Point	13
4. EXAMPLES	14
5. CONCLUDING REMARKS	17
6. ACKNOWLEDGMENTS	18
7. REFERENCES	18
APPENDIX A. Analytical Expression of the Condition for Determining the Direction of the Tangent	19
APPENDIX B. Computer Subroutine CRVFIT	23

A METHOD OF SMOOTH CURVE FITTING

Hiroshi Akima

A new mathematical method of fitting a smooth curve to a set of given points in a plane is developed, and a computer subroutine is programmed to implement the method. This method is devised in such a way that the resultant curve will pass through all the given points and will look smooth and natural. The interpolation between the given points is performed locally, and no assumption of the functional form is made for the whole curve. ( )

Key words: Direction of tangent, interpolation, polynomial, smooth curve fitting.

1. INTRODUCTION

When we try to determine a relation between two variables, we either perform computations or make measurements. The result is given as a set of discrete points in a plane. Knowing that the relation can be represented by a smooth curve, we next try to fit a smooth curve to the set of points so that the resultant curve will pass through all the given points. Manual drawing is the most primitive method for this purpose and results in a reasonable curve if it is done by a well-trained scientist or engineer. But, since this method is very tedious and time consuming, we wish to let a computer draw a curve. To do so, we must provide the computer with necessary instructions for mathematically interpolating points between the given points.

There are several mathematical methods of interpolating the value of a function from a given set of values (Milne, 1949; Hildebrand, 1956; Ralston and Wilf, 1967), but the application of one of these methods to curve fitting sometimes results in a curve that is very different from one drawn manually. In other words, the resultant curve sometimes appears strange and unnatural. In this report, we present a new method of interpolation and smooth curve fitting that is devised so that the resultant curve will look smooth and natural.

2. DISCUSSION OF SOME EXISTING METHODS

A simple example, taken from a study of FM distortion being conducted by the author, will serve to illustrate difficulties encountered by existing mathematical methods of interpolation. Assume that the values of x and y at 11 points are as follows:

x	0	1	2	3	4	5	6	7	8	9	10
y = y(x)	10	10	10	10	10	10	10.5	15	50	60	85

Knowing from the physical nature of the phenomena that y(x) is a single-valued smooth function of x, we try to interpolate the value of y(x) and to fit a smooth curve to the given set of points.

First we apply the method of interpolation based on polynomials (Milne, 1949; Hildebrand, 1956, ch. 2, 3, 4). This method is perhaps the one most often used because, as stated by Milne (1949), "polynomials are simple in form, can be calculated by elementary operations, are free from singular points, are unrestricted as to range of values, may be differentiated or integrated without difficulty, and the coefficients to be determined enter linearly." There are several variants of this method, known by the names of Newton-Cotes, Lagrange, Aitken, and Neville. Each has its own advantages and disadvantages, but they are all based on the common assumption that y(x) can be closely approximated by a polynomial of x of order n - 1, where n is the number of given points. They should give the same result, because the uniqueness of the polynomial of order n - 1 that agrees with given values of y(x) at the given n points has been proved (Hildebrand, 1956, p. 44). The result obtained by applying the 10th order (Lagrangian) polynomial is shown in figure 1A where the given data points are encircled.

Next we try to apply the method based on a ratio of two polynomials or a rational function (Hildebrand, 1956, sec. 9.9 - 9.12). This method is not so commonly used as the polynomial method. Perhaps its most serious disadvantage is that the desired function does not always exist; in our example, it does not. Another disadvantage is that the nonsingularity of the function is not guaranteed. If we omit the first point (0, 10), however, the function exists. The result thus obtained is shown in figure 1B.

The third one is a well-known method based on the Fourier series (Hildebrand, 1956, sec. 9.3), which exemplifies the so-called orthogonal functions. In applying this method to our data, we assume that the whole range of x from 0 to 10 corresponds to one-half of the fundamental period from 0 to π and apply a series of cosine functions up to the 10th order harmonic terms. The result is shown in figure 1C.

Finally, we apply the method based on a spline function (Ralston and Wilf, 1967). The spline function of degree m is a piecewise function composed of a set of polynomials, each of order at most m and applicable to successive intervals of the given data points. It has the characteristic that the function and its derivatives of order 1, 2, ..., m - 1 are continuous in the whole range of x. This function includes the (Lagrangian) polynomial as a special case when $m = n - 1$, where n is the number of given points. The result of applying the spline function of degree three is shown in figure 1D.

In addition to these results of four rather representative mathematical methods, the curve obtained manually is shown in figure 1E.

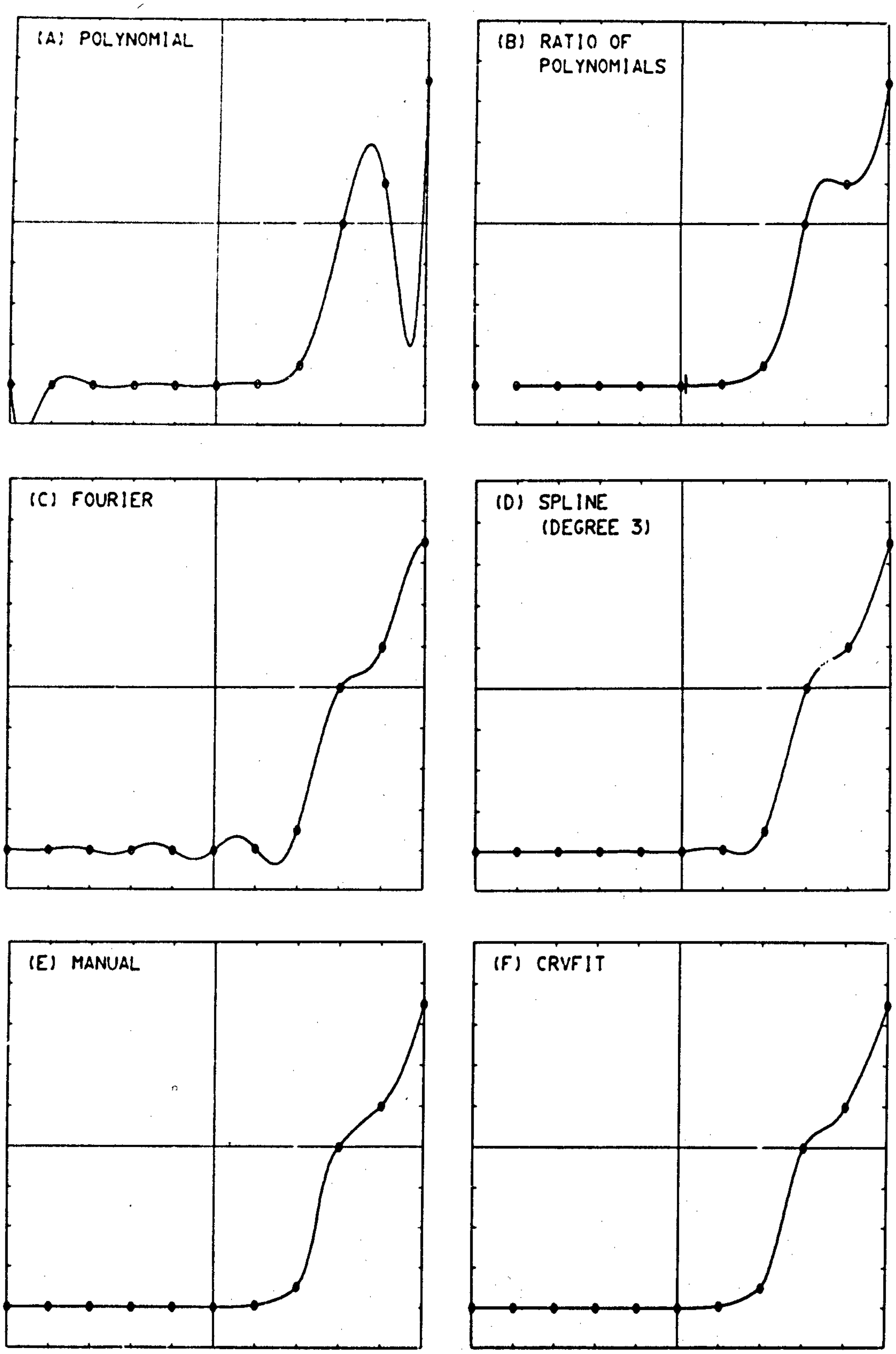


Figure 1. Comparison of several methods of smooth curve fitting.

It is the average of curves drawn manually by several scientists and engineers. A comparison of the curves in figures 1A - 1D with 1E indicates that the first three mathematical methods are not suitable for the example given. It also indicates that the curve obtained by the spline function and shown in figure 1D resembles the one in figure 1E, but we are proposing an alternate method, discussed in the sections that follow. As we see, the curve obtained by the new method, shown in figure 1F, is closer than the other curves to the manually drawn curve in figure 1E.

The common difficulty encountered by the existing mathematical methods is that the resultant curve shows unnatural wiggles. This seems inevitable if we make any assumption concerning the functional form for the whole set of given points other than the continuity and the smoothness of the curve. When such an assumption is not justified, the resultant curve is very likely to behave strangely, as in figures 1A - 1D. The functional form is what we try to derive and not what we assume.

When we try to fit a smooth curve manually, we do not assume any functional form for the whole curve. We draw a portion of the curve based on a relatively small number of points, without taking into account the whole set of points. This local procedure is a very important feature of manual curve fitting and the basis for our new method. Note that, although the spline function is a piecewise function composed of a set of polynomials, all polynomials are determined simultaneously on the basis of the assumption of continuities of the function and its derivatives in the whole range, and no individual polynomial can be determined locally.

It is not easy to develop a mathematical method of smooth curve fitting based on the local procedure. If one of the existing mathematical methods is applied locally or piecewise without any special consideration, the continuity of the function or its first order derivative is not generally

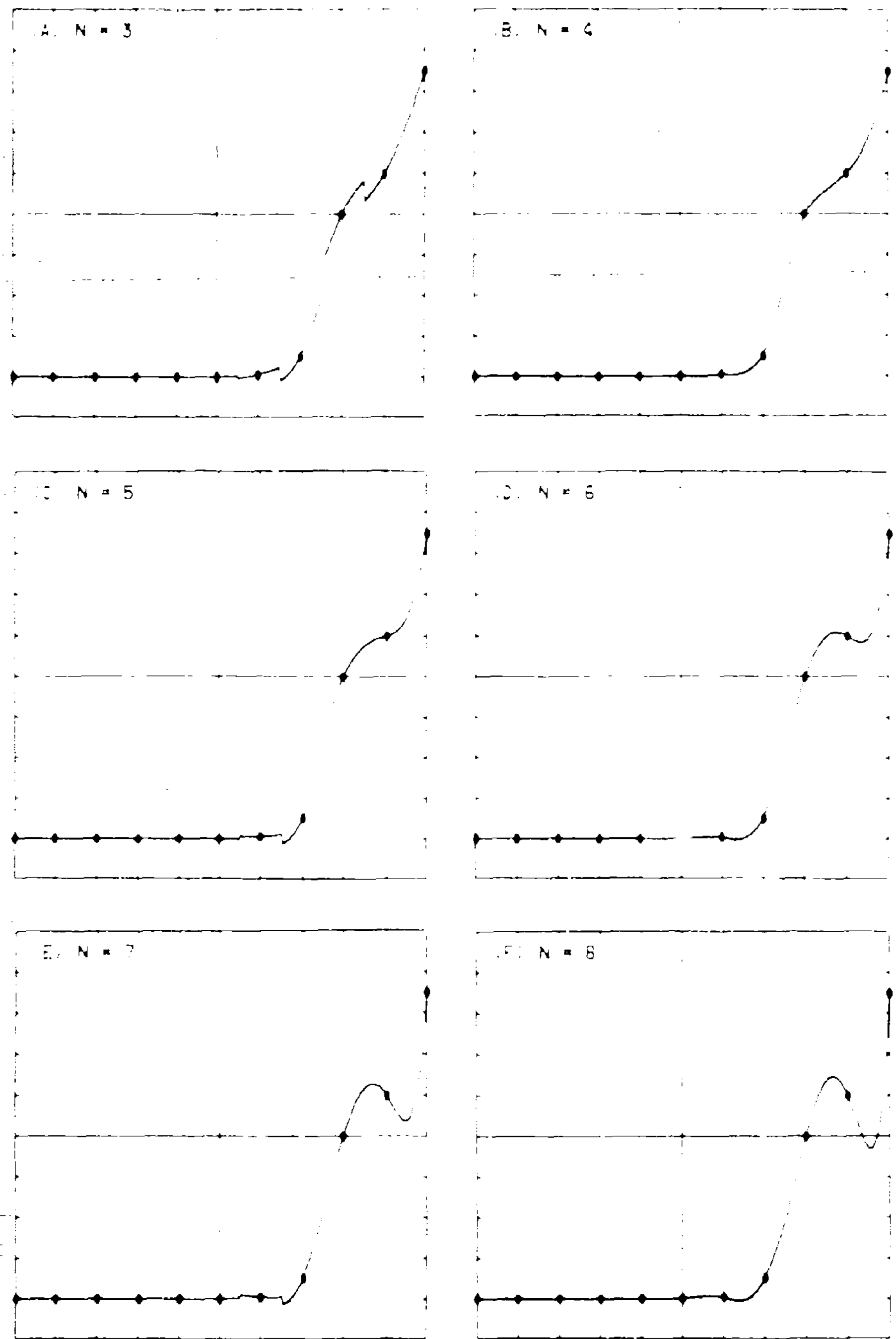


Figure 2. Examples of simple piecewise application of the polynomial method. (A polynomial of order N - 1 is applied to each set of N successive points.)

guaranteed. This situation is illustrated in figure 2, where the same data as in figure 1 are used and, for each N, a polynomial of order N - 1 is applied to each set of N successive points to interpolate the part of the curve in a unit interval around the center of the N points. When N is an odd number, the resultant curves are discontinuous; when it is an even number, the slopes of the resultant curves are not continuous, and thus the curves are not smooth.

3. NEW METHOD

3.1. Outline of the Method

Our method is devised so as to work in two different ways, i.e., one for a single-valued function and the other for a multiple-valued function, to correspond to the two ways in which a smooth curve is fitted manually, depending on whether we know that the given data points represent a single-valued function or not.

Our method is based on a piecewise function given by a polynomial of the third order in each interval for the case of a single-valued function, and by a pair of polynomials of the third order for the case of a multiple-valued function. For a single-valued function, our method is somewhat similar to the spline function of degree three. As in the spline function, the continuity of the function itself and of its first-order derivative (the direction of the tangent to the curve or the slope of the curve) are assumed. But, instead of assuming the continuity of the second-order derivative as in the third-degree spline function, we determine the direction of the tangent locally under certain assumptions. By doing so, we can fit a curve piecewise to the given set of data points without having discontinuities in the curve and its slope.

We assume that the direction of the tangent to the curve (or the slope of the curve) at a given point P_i is determined by the coordinates of five points, P_{i-2}, P_{i-1}, P_i, P_{i+1}, and P_{i+2}. In other words, the points more than two intervals away are assumed not to affect the determination of the slope. This is discussed in more detail in the next section.

The portion of the curve between a pair of points is assumed to be determined only by the coordinates of and the slopes at the two points. This interpolation procedure is described in section 3. 3. Since the slope of the curve should be determined also at the end points of the curve, estimation of two more points is necessary at each end point. This extrapolation procedure is described in section 3. 4.

3. 2. Direction of the Tangent

Consider five points, 1, 2, 3, 4, and 5, as shown in figure 3A. Let the point of intersection of the two straight lines extended from line segments $\overline{12}$ and $\overline{34}$ be denoted by A and the same point corresponding to line segments $\overline{23}$ and $\overline{45}$ by B. We seek a reasonable condition for determining the direction of the tangent $\overline{CD}$ at point 3.

It seems appropriate to assume that the direction of $\overline{CD}$ should approach that of $\overline{23}$ when the direction of $\overline{12}$ approaches that of $\overline{23}$, and that angle $\angle 23C$ (the angle between $\overline{32}$ and $\overline{3C}$) should be equal to $\angle D34$ when $\angle 123$ is equal to $\angle 345$. With these rather intuitive reasonings as a guideline, the condition of determining the direction of $\overline{CD}$ is still not unique. For simplicity we assume that the tangent $\overline{CD}$ is determined by the condition

$$\frac{\overline{2C}}{\overline{CA}} = \frac{\overline{4D}}{\overline{DB}} .$$

This condition, however, does not exist for certain configurations of five points, such as the one shown in figure 3B. In this case the alternate condition,

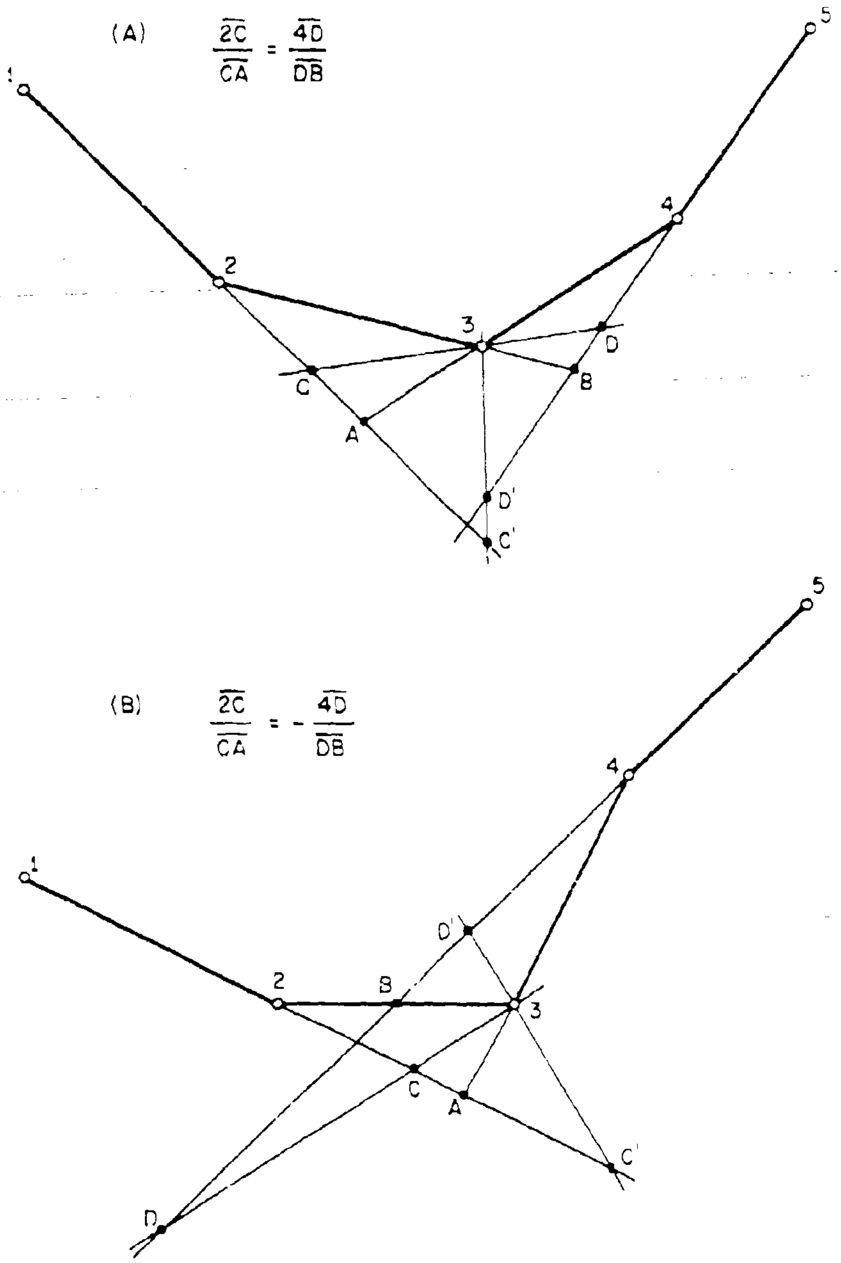


Figure 3. Determination of the direction of the tangent to the curve.

$$\frac{\overline{2C}}{\overline{CA}} = -\frac{\overline{4D}}{\overline{DB}} ,$$

does exist, as shown in the figure, and we shall use this condition.

Summarizing the above observations, we assume that the direction of the tangent $\overline{CD}$ is determined by the condition

$$\frac{\overline{2C}}{\overline{CA}} = \pm\frac{\overline{4D}}{\overline{DB}} . \tag{1}$$

The double sign depends on the configuration of the five points, and the one for which the condition exists will be selected.

Analytically, as shown in appendix A, condition (1) results in a quadratic equation with respect to $\tan\theta$, where θ is the angle of $\overline{CD}$ measured from the x axis. For a given configuration of the points, the discriminant of the quadratic equation is positive or zero for one sign and negative for the other. The former should be selected as a matter of course, so that the equation has real roots. Note that there is no ambiguity in the selection of the sign.

When one line segment is parallel to another one, the direction of the tangent $\overline{CD}$ may sometimes be indefinite under condition (1). Analytically this corresponds to the case where the three coefficients of the quadratic equation are all zero. To avoid this uncertainty, we assume the following:

(a) When line segment $\overline{23}$ is parallel to $\overline{34}$, the tangent is parallel to $\overline{23}$. If the direction of $\overline{23}$ is opposite that of $\overline{34}$, the tangent changes its direction by 180° at point 3.

(b) When $\overline{12}$ is parallel to $\overline{23}$ and $\overline{34}$ is parallel to $\overline{45}$, the tangent is parallel to $\overline{24}$.

(c) When $\overline{12}$ is parallel to $\overline{34}$ and $\overline{23}$ is parallel to $\overline{45}$, the tangent is parallel to $\overline{24}$.

Two roots or solutions, $\overline{CD}$ and $\overline{C'D'}$, satisfy condition (1). The one, $\overline{CD}$, should be selected so that the two points, 2 and 4, lie on the same side of the straight line through C and D. This can be done by adding a second condition,

$$\sin(\underline{/C32}) \cdot \sin(\underline{/C34}) > 0 \ . \tag{2}$$

When this method is applied to a multiple-valued function, the sense of the direction of the tangent should also be determined. In other words, we have to determine both $\cos\theta$ and $\sin\theta$, not only $\tan\theta$. The sense should be selected so that both the positive direction of the tangent and line segment $\overline{34}$ lie on the same side of the straight line through 2 and 3. This can be done by adding a third condition,

$$\sin(\theta - \theta_{\overline{23}}) \cdot \sin(\theta_{\overline{34}} - \theta_{\overline{23}}) > 0, \tag{3}$$

where $\theta_{\overline{23}}$ and $\theta_{\overline{34}}$ are the angles of line segments $\overline{23}$ and $\overline{34}$ measured from the x axis, respectively.

Note that the procedure for determining the slope is a geometrical one and is independent of the coordinate system.

3. 3. Interpolation Between a Pair of Points

3. 3. 1. Single-Valued Function

For a single-valued function $y = y(x)$, we assume that the curve between a pair of points can be expressed by

$$y = p_0 + p_1 x + p_2 x^2 + p_3 x^3, \tag{4}$$

where the p's are constants. Since the coordinates of the two points, say (x_1, y_1) and (x_2, y_2), as well as the directions of the curve $\tan\theta_1$

and $\tan\theta_2$ at the points, are given, we further assume that x and y satisfy the conditions

$$y = y_1 \text{ and } \frac{dy}{dx} = \tan\theta_1 \text{ at } x = x_1,$$

$$y = y_2 \text{ and } \frac{dy}{dx} = \tan\theta_2 \text{ at } x = x_2.$$

From these conditions we can uniquely determine the p constants.

Note that the interpolated curve is independent of the scalings of the x and y axes.

3.3.2. Multiple-Valued Function

For a multiple-valued function, we assume that the curve between a pair of points (x_1, y_1) and (x_2, y_2) can be expressed by

$$x = p_0 + p_1 z + p_2 z^2 + p_3 z^3,$$
$$y = q_0 + q_1 z + q_2 z^2 + q_3 z^3, \qquad (5)$$

where the p's and q's are constants and z is a parameter that varies from 0 to 1 as the curve is traversed from (x_1, y_1) to (x_2, y_2). Since the coordinates of the two points (x_1, y_1) and (x_2, y_2), as well as the directions of the curve $(\cos\theta_1, \sin\theta_1)$ and $(\cos\theta_2, \sin\theta_2)$ at the points, are given, we further assume that x and y satisfy the conditions

$$x = x_1,\ y = y_1,\ \frac{dx}{dz} = r\cos\theta_1, \text{ and } \frac{dy}{dz} = r\sin\theta_1 \text{ at } z = 0,$$

$$x = x_2,\ y = y_2,\ \frac{dx}{dz} = r\cos\theta_2, \text{ and } \frac{dy}{dz} = r\sin\theta_2 \text{ at } z = 1,$$

where

$$r = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} .$$

From these conditions we can uniquely determine the p and q constants.

Note that in this case the interpolated curve depends on the scalings of the x and y axes.

3.4. Extrapolation of the Curve at an End Point

Except for the case of a closed curve, estimation of two more points from the given points is required at each end of the curve.

3.4.1. Single-Valued Function

For a single-valued function $y = y(x)$, we assume that the curve near the end can be expressed by

$$y = g_0 + g_1 x + g_2 x^2, \tag{6}$$

where the g's are constants. The constants can be determined from the coordinates of three given points (x_1, y_1), (x_2, y_2), and (x_3, y_3). Assuming that

$$x_5 - x_4 = x_4 - x_3 = x_3 - x_2 ,$$

we can determine the ordinates y_4 and y_5 corresponding to x_4 and x_5, respectively, from (6).

Note that the extrapolated points are independent of the scalings of the x and y axes.

3.4.2. Multiple-Valued Function (Nonclosed Curve)

For a multiple-valued function, we assume that the curve near the end can be expressed by

$$x = g_0 + g_1 z + g_2 z^2 ,$$
$$y = h_0 + h_1 z + h_2 z^2 , \tag{7}$$

where the g's and h's are constants and z is a parameter. We further assume that

$$x = x_i \text{ and } y = y_i \text{ at } z = i \ (i = 1, 2, 3, 4, 5).$$

From the coordinates of three given points (x_1, y_1), (x_2, y_2), and (x_3, y_3), the g and h constants can be determined, and consequently, also the coordinates (x_4, y_4) and (x_5, y_5).

Note that the extrapolated points are independent of the scalings of the x and y axes.

4. EXAMPLES

An example of the application of this method is shown in figure 1F where the curve is very close to the one in 1E determined manually.

Figure 4 gives some examples of the application of the method in different modes. Two curves are drawn for single-valued functions $y = y(x)$ (MODE = 1) and $x = x(y)$ (MODE = 2) in A and B, respectively. Two examples for nonclosed-curve, multiple-valued function (MODE = 3) are given in C and D. A circle and an ellipse are drawn in E and F, respectively, as examples for the case of closed curve (MODE = 4).

Figure 5 shows artificial examples for simple configurations of given points that are designed to supplement the description of the method, especially of the direction of the tangent (see sec. 3.2). This figure also illustrates how strangely curves may sometimes behave for adverse configurations of given data points. Application of the third-

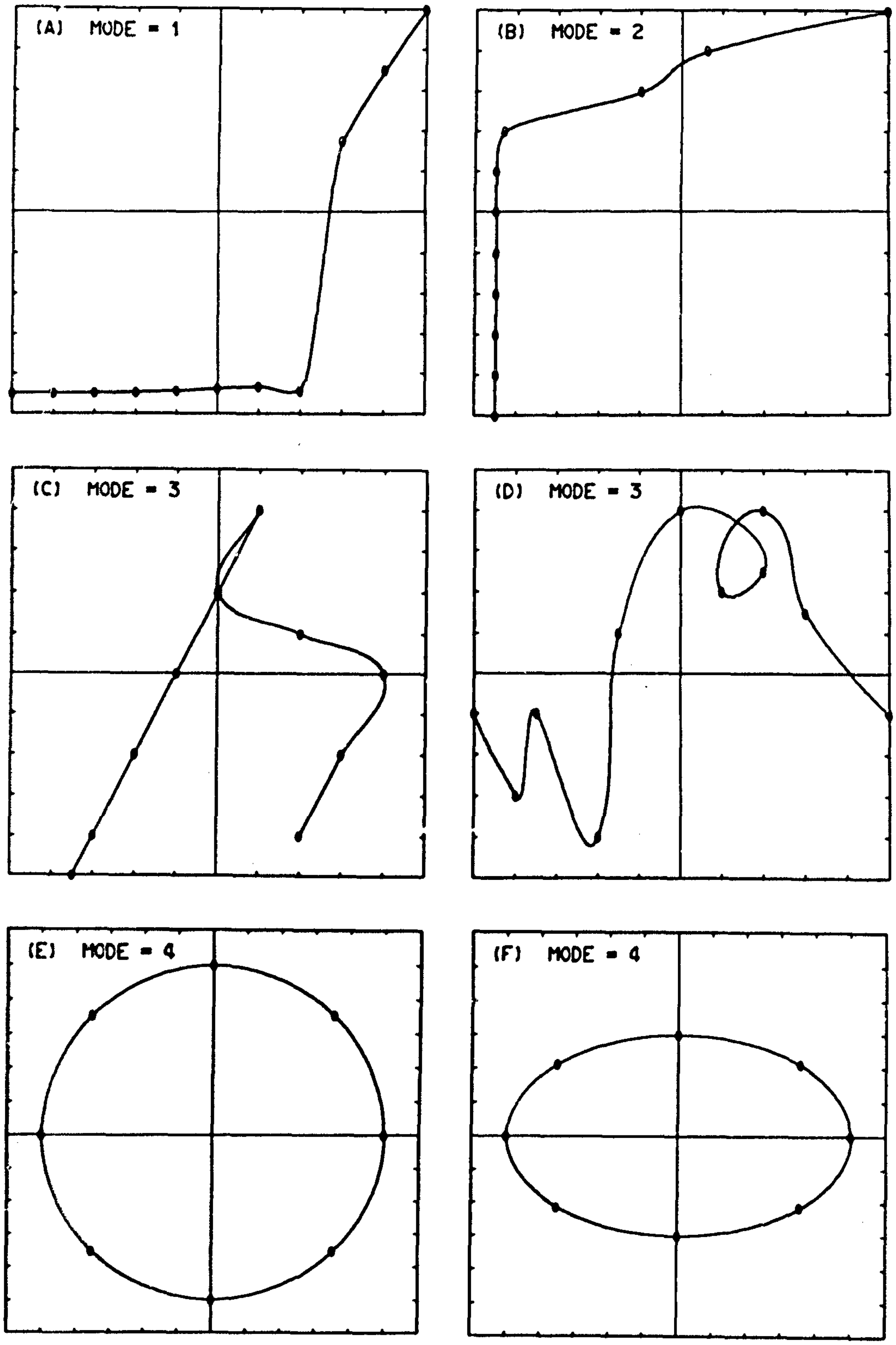


Figure 4. Examples of application of the new method.

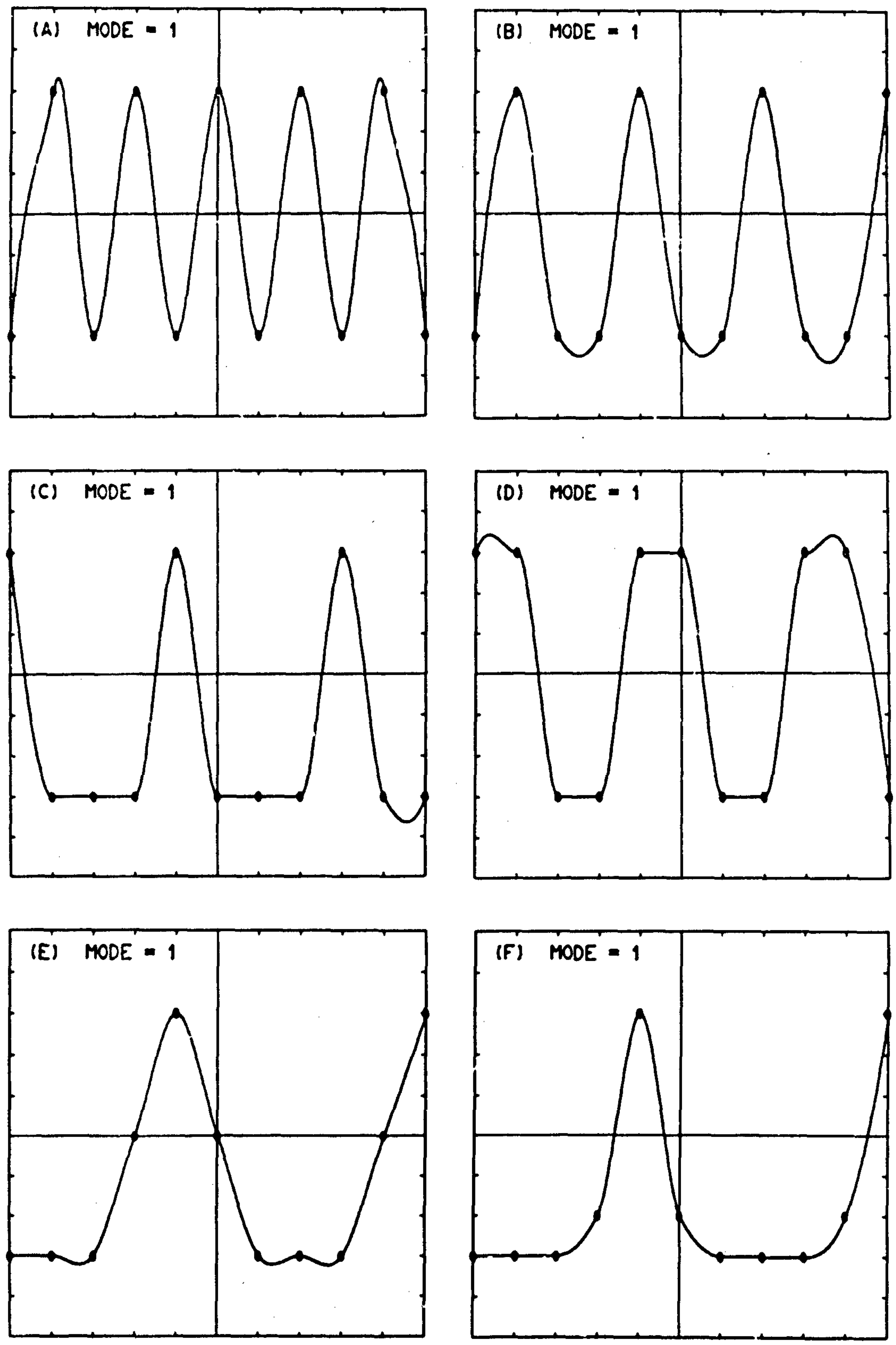


Figure 5. Further examples of application of the new method.

degree spline function to the same data results in slightly better, but still not completely satisfactory, curves. All the examples in this figure are periodic functions, but neither the new method nor the spline function method is devised so as to detect or take advantage of regularity, such as periodicity. Because periodicity is neglected in both these methods, the resultant curves sometimes appear strange, especially near their end points.

5. CONCLUDING REMARKS

We have described a new method of smooth curve fitting. For proper application of the new method, the following remarks seem pertinent:

(1) The curve obtained by this method passes through all the given points. Therefore, the method is applicable only to the case where the precise values of the coordinates of the points are given. It should be recognized that all experimental data have some errors in them, and unless the errors are negligible it is more appropriate to smooth the data, i.e., to fit a curve approximating the data appropriately, than to fit a curve passing through all the points.

(2) Use of this method is not recommended when given data points manifest apparent regularity or when we have a priori knowledge on the regularity of the data.

(3) As is true for any method of interpolation, no guarantee can be given of the accuracy of the interpolation, unless the method in question has been checked in advance against precise values or a functional form.

(4) The method yields a smooth and natural curve and is

therefore useful in cases where manual, but tedious, curve fitting will do in principle.

(5) For a single-valued function, the resultant curve is invariant under a linear-scale transformation of the coordinate system. In other words, different scalings of the coordinates result in a same curve.

(6) For a multiple-valued function, the resultant curve is variant under a linear-scale transformation of the coordinate system. The scalings of the coordinates should be coincident with the actual size of the graph.

A computer subroutine, named CRVFIT, has been programmed to implement the method reported on in this paper. It is described in detail in appendix B.

6. ACKNOWLEDGMENTS

The author expresses his deep appreciation to E. L. Crow, R. K. Rosich, and J. S. Washburn of the Institute for Telecommunication Sciences, ESSA Research Laboratories, for their helpful discussions.

7. REFERENCES

Hildebrand, F. B. (1956), "Introduction to Numerical Analysis," ch. 2, 3, 4, and 9 (McGraw-Hill Book Co., New York, N.Y.).

Milne, W. E. (1949), "Numerical Calculus," ch. III (Princeton University Press, Princeton, N.J.).

Ralston, A., and H. S. Wilf (1967), "Mathematical Methods for Digital Computers," vol. II, ch. 8 (John Wiley & Sons, New York, N.Y.).

APPENDIX A

Analytical Expression of the Condition for Determining the Direction of the Tangent

Let the coordinates of points, 1, 2, 3, 4, 5, A, B, C, and D in figure 3 in the text be denoted by (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5), (x_a, y_a), (x_b, y_b), (x_c, y_c), and (x_d, y_d), respectively, and the tangent of the angle of $\overline{CD}$ measured from the x axis by t. Then, the following equations should hold:

$$\frac{y_a - y_2}{x_a - x_2} = \frac{y_c - y_2}{x_c - x_2} = \frac{y_2 - y_1}{x_2 - x_1} , \tag{8}$$

$$\frac{y_4 - y_b}{x_4 - x_b} = \frac{y_4 - y_d}{x_4 - x_d} = \frac{y_5 - y_4}{x_5 - x_4} , \tag{9}$$

$$\frac{y_3 - y_a}{x_3 - x_a} = \frac{y_4 - y_3}{x_4 - x_3} , \tag{10}$$

$$\frac{y_b - y_3}{x_b - x_3} = \frac{y_3 - y_2}{x_3 - x_2} , \tag{11}$$

$$\frac{y_3 - y_c}{x_3 - x_c} = \frac{y_d - y_3}{x_d - x_3} = t . \tag{12}$$

Condition (1) in the text can be expressed by

$$\frac{x_2 - x_c}{x_c - x_a} = \pm \frac{x_4 - x_d}{x_d - x_b} . \tag{13}$$

If we introduce new constants defined by

$$a_i = x_{i+1} - x_i , \quad (i = 1, 2, 3, 4) , \tag{14}$$

$$b_i = y_{i+1} - y_i , \quad (i = 1, 2, 3, 4) , \tag{15}$$

$$S_{ij} = a_i b_j - a_j b_i , \quad (i = 1, 2, 3; \; j = 2, 3, 4; \; i < j) , \tag{16}$$

and eliminate x's and y's from (8) to (13), our condition reduces to a quadratic equation of the form

$$At^2 - 2Bt + C = 0, \tag{17}$$

where

$$A = S_{12} S_{24} a_3^2 \mp S_{13} S_{34} a_2^2 , \tag{18}$$

$$B = S_{12} S_{24} a_3 b_3 \mp S_{13} S_{34} a_2 b_2 , \tag{19}$$

$$C = S_{12} S_{24} b_3^2 \mp S_{13} S_{34} b_2^2 . \tag{20}$$

The discriminant of (17) is expressed by

$$D = B^2 - AC$$

$$= \pm S_{23}^2 S_{12} S_{24} S_{13} S_{34} . \tag{21}$$

We will take the upper sign in (1) and (13) when the product $S_{12} S_{24} S_{13} S_{34}$ is positive or zero, and the lower sign when it is negative. By doing so, we can always make the discriminant nonnegative.

If the coefficients A, B, and C in (17) are all zero, the solution of the equation is indefinite. This occurs when

(a) $S_{23} = 0$,

(b) $S_{12} = 0$ and $S_{34} = 0$,

(c) $S_{13} = 0$ and $S_{24} = 0$,

(d) $S_{12} = 0$ and $S_{13} = 0$,

(e) $S_{24} = 0$ and $S_{34} = 0$.

For the case (a), we assume the provision that the tangent t is equal to $b_2 / a_2 = b_3 / a_3$. For cases (b) and (c), we assume that the tangent t is equal to $(b_2 + b_3)/(a_2 + a_3)$. Cases (d) and (e) are special cases of (a), and the provision for (a) applies.

Equation (17) has two roots, i.e.,

$$t = \frac{B \pm \sqrt{D}}{A}. \tag{22}$$

Of the two, the desired one will be selected so that it satisfies

$$S_{20} S_{03} > 0, \tag{23}$$

where

$$S_{20} = a_2 b_0 - a_0 b_2, \tag{24}$$

$$S_{03} = a_0 b_3 - a_3 b_0, \tag{25}$$

$$a_0 = 1/\sqrt{1 + t^2}, \tag{26}$$

$$b_0 = t/\sqrt{1 + t^2}. \tag{27}$$

Condition (23) can equivalently be written as (2) in the text.

If S_{20} satisfies

$$S_{20} S_{23} > 0, \tag{28}$$

$\cos\theta$ and $\sin\theta$ are equal to a_0 and b_0, respectively. Otherwise, $\cos\theta$ and $\sin\theta$ are equal to $-a_0$ and $-b_0$, respectively. Condition (28) is equivalent to (3) in the text.

APPENDIX B

Computer Subroutine CRVFIT

The CRVFIT subroutine is a FORTRAN subroutine programmed for the CDC-3800 computer to implement the method of smooth curve fitting reported on in the text. Necessary information for the user of the subroutine is given on pages 25 and 26 in a format compatible with the library function manual. A FORTRAN listing of the subroutine is given on pages 27 to 30. A binary deck of the subroutine is available from the library, Computer Division, ESSA Research Laboratories, Boulder, Colorado, under the name of E2-IERB-CRVFIT.

Although the CRVFIT subroutine has been programmed for the CDC-3800 computer, it can be modified without difficulty for other digital computers that accept a FORTRAN language.

The CRVFIT subroutine is devised so as to compute coordinates of closely spaced points on a smooth curve determined by a set of given points, and not to compute a value of ordinate for a specific value of abscissa. But there is no difficulty in modifying the subroutine for this purpose.

PURPOSE: To fit a smooth curve to a set of given points in a plane; i.e., to compute coordinates of a new set of points (output points) that are located on a smooth curve determined by a set of given points (input points) in a plane and are more closely spaced than the set of input points on the curve. (This subroutine interpolates points between each pair of input points, and the output points consist of both the input points and the interpolated points.)

FORTRAN CALLING SEQUENCE:

```
CALL CRVFIT(MD,L,X,Y,M,N,U,V)
```

where

MD = mode of the curve (input parameter),

= 1 for a single-valued function Y = Y(X),
= 2 for a single-valued function X = X(Y),
= 3 for a multiple-valued function, nonclosed curve,
= 4 for a multiple-valued function, closed curve,

L = number of input point (input parameter),

X, Y = arrays containing the abscissas and ordinates of L input points (input parameters),

M = number of divisions between each pair of input points (input parameter),

N = number of output points (output parameter), and

U, V = arrays where the abscissas and ordinates of N output points are to be displayed (output parameters).

ERROR MESSAGE: When $L \leq 0$ or $M \leq 0$, the error message

```
***     L = 0/NEG OR M = 0/NEG.
        A = (value of L)     Q = (value of M)
ERROR DETECTED IN ROUTINE  CRVFIT
```

will be printed on the standard output unit, and the job will be aborted.

STORAGE: 826 locations.

TIMING: (55M + 710) L microseconds for MD = 1,
(68M + 720) L microseconds for MD = 2,
(70M + 1000) L microseconds for MD = 3 or 4.

METHOD:

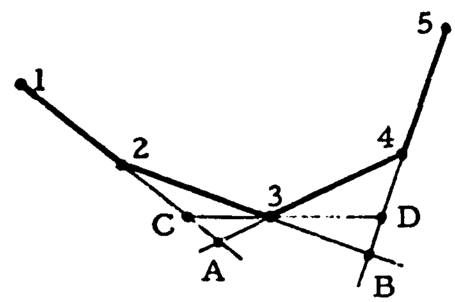
Step 1. Except for MD = 4, two more points are estimated at each end point of the curve by assuming that the curve near the end point can be expressed by

$$y = g_0 + g_1 x + g_2 x^2 \qquad \text{for MD} = 1$$

and

$$x = g_0 + g_1 z + g_2 z^2$$
$$y = h_0 + h_1 z + h_2 z^2 \qquad \text{for MD} = 3.$$

Step 2. The direction of the tangent to the curve at point 3 is determined by the configuration of 5 points, 1, 2, 3, 4, and 5, shown in the figure, by



$$\left|\frac{\overline{2C}}{\overline{CA}}\right| = \left|\frac{\overline{4D}}{\overline{DB}}\right| .$$

Step 3. The curve between a pair of points is computed by

$$x = p_0 + p_1 z$$
$$y = q_0 + q_1 z + q_2 z^2 + q_3 z^3 \qquad \text{for MD} = 1$$

and

$$x = p_0 + p_1 z + p_2 z^2 + p_3 z^3$$
$$y = q_0 + q_1 z + q_2 z^2 + q_3 z^3 \qquad \text{for MD} = 3,\ 4.$$

Computation for MD = 2 is made by interchanging the X and Y arrays, applying the method for MD = 1, and finally interchanging the U and V arrays.

```
      SUBROUTINE  CRVFIT(MODE,LO,X,Y,MO,NO,U,V)
C     SMOOTH CURVE FITTING
C     PROGRAMMED BY HIROSHI AKIMA, ESSA-ITS

C     MODE = 1  FOR  Y = Y(X)
C     MODE = 2  FOR  X = X(Y)
C     MODE = 3  FOR  X = X(S)  AND  Y = Y(S),  NONCLOSED CURVE
C     MODE = 4  FOR  X = X(S)  AND  Y = Y(S),  CLOSED CURVE
C     LO = NO. OF INPUT POINTS,   X AND Y ARE INPUT POINTS
C     NO = NO. OF OUTPUT POINTS,  U AND V ARE OUTPUT POINTS
C     MO = NO. OF DIVISIONS

C     DECLARATION STATEMENTS

      DIMENSION      X(10),Y(10),U(100),V(100)
      DIMENSION      AO(2),BO(2)
      EQUIVALENCE    (A,AO(1)),(B,AO(2),BO(1)),(C,BO(2)),
     1               (PO,X2),(QO,Y2),(DX,A2),(DY,B2)
      EQUIVALENCE    (FLM,TS,Z),(JP,JS),
     1               (DU,DA,D,X1),(DV,DB,R,Y1),
     2               (S2,S20,A1),(S3,S03,B1),
     3               (P1,S12),(P2,C12),(P3,R12),
     4               (Q1,S13),(Q2,C13),(Q3,R13)
      DIMENSION      ER1(2),ER2(2),MSG(3)
      TYPE DOUBLE    DER1,DER2
      EQUIVALENCE    (L,ER1(1)),(M,ER1(2)),(DER1,ER1),
     1               (LM1,ER2(1)),(MM1,ER2(2)),(DER2,ER2)
      DATA           (MSG=24HL = 0/NEG OR M = 0/NEG. )

C     STATEMENT FUNCTION

      SCR(SIJ,CIJ)=ABSF(SIJ)-ABSF(CIJ)*1.0E-8

C     PRELIMINARY PROCESSING

      MD=MODE                 $   L=LPS=LO              $   M=MO
      IF(L.LE.0.OR.M.LE.0)        GO TO 900
      KP1=L*M+1               $   IP=L+1
      DO 10  JP=1,L           $   KP1=KP1-M             $   IP=IP-1
      U(KP1)=X(IP)            $   V(KP1)=Y(IP)
   10 CONTINUE                $   KP2=KP3=1
      DO 20  I=2,L            $   KP2=KP2+M
      IF(U(KP2).EQ.U(KP3).AND.V(KP2).EQ.V(KP3))         GO TO  20
      KP3=KP3+M               $   U(KP3)=U(KP2)         $   V(KP3)=V(KP2)
   20 CONTINUE                $   L=KP3/M+1             $   N=KP3
      IF(N.EQ.1)                  GO TO 890
      IF(MD.NE.2)                 GO TO  50
   30 DO 40  KP4=1,N,M
      TS=U(KP4)               $   U(KP4)=V(KP4)         $   V(KP4)=TS
   40 CONTINUE
   50 MM1=M-1                 $   FLM=M                 $   DZ=1.0/FLM
      IF(L.EQ.2)                  GO TO 100
      LM1=L-1                 $   GO TO 200
```

```
C       SMOOTH CURVE FITTING FOR L = 2

  100 DU=(U(N)-U(1))*DZ         $     DV=(V(N)-V(1))*DZ
      DO 110 KS=1,MM1
      U(KS+1)=U(KS)+DU          $     V(KS+1)=V(KS)+DV
  110 CONTINUE                  $     GO TO 800

C       SMOOTH CURVE FITTING FOR L GREATER THAN 2

  200 X3=U(1)                   $     Y3=V(1)
      X4=U(M+1)                 $     Y4=V(M+1)
      K5=1+M+M                  $     X5=U(K5)              $     Y5=V(K5)
      A3=X4-X3                  $     B3=Y4-Y3
      A4=X5-X4                  $     B4=Y5-Y4
      S34=A3*B4-A4*B3           $     C34=A3*A4+B3*B4       $     R34=SCR(S34,C34)
      IF(MD.LE.3)                     GO TO 400
      K1=N-M-M                  $     X1=U(K1)              $     Y1=V(K1)
      K2=K1+M                   $     X2=U(K2)              $     Y2=V(K2)
      A1=X2-X1                  $     B1=Y2-Y1
      A2=X3-X2                  $     B2=Y3-Y2
  230 S12=A1*B2-A2*B1           $     C12=A1*A2+B1*B2       $     R12=SCR(S12,C12)
      S13=A1*B3-A3*B1           $     C13=A1*A3+B1*B3       $     R13=SCR(S13,C13)
      S23=A2*B3-A3*B2           $     C23=A2*A3+B2*B3       $     R23=SCR(S23,C23)
      S24=A2*B4-A4*B2           $     C24=A2*A4+B2*B4       $     R24=SCR(S24,C24)
      ASSIGN 240 TO LBL         $     GO TO 500
  240 DO 290 I=2,L              $     K5=K5+M
      X2=X3                     $     Y2=Y3
      X3=X4                     $     Y3=Y4
      X4=X5                     $     Y4=Y5
      A2=A3                     $     B2=B3
      A3=A4                     $     B3=B4
      S12=S23                   $     R12=R23
      S13=S24                   $     R13=R24
      S23=S34                   $     C23=C34               $     R23=R34
      COS2=SGN*COS3             $     SIN2=SGN*SIN3
      IF(I.LT.LM1)                    GO TO 270
      IF(MD.EQ.4)                     GO TO 260
      IF(I.EQ.LM1)                    GO TO 450
      A4=A5                     $     B4=B5                 $     GO TO 280
  260 IF(K5.GT.N)                     K5=1+M
  270 X5=U(K5)                  $     Y5=V(K5)
      A4=X5-X4                  $     B4=Y5-Y4
  280 S24=A2*B4-A4*B2           $     C24=A2*A4+B2*B4       $     R24=SCR(S24,C24)
      S34=A3*B4-A4*B3           $     C34=A3*A4+B3*B4       $     R34=SCR(S34,C34)
      ASSIGN 600 TO LBL         $     GO TO 500
  290 CONTINUE                  $     GO TO 800
```

```
C       EXTRAPOLATION AT THE BEGINNING

  400 IF(R34.LE.0.0)                GO TO 430
      IF(MD.EQ.3)                   GO TO 420
  410 A1=A2=A3             $        DB=2*A3*S34/(A4*(A3+A4))
      B2=B3-DB             $        B1=B2-DB             $      GO TO 230
  420 DA=A4-A3             $        DB=B4-B3
      A2=A3-DA             $        B2=B3-DB
      A1=A2-DA             $        B1=B2-DB             $      GO TO 230
  430 A1=A2=A3             $        B1=B2=B3             $      GO TO 230

C       EXTRAPOLATION AT THE END

  450 IF(R23.LE.0.0)                GO TO 480
      IF(MD.EQ.3)                   GO TO 470
  460 A5=A4=A3             $        DB=2*A3*S23/(A2*(A2+A3))
      B4=B3+DB             $        B5=B4+DB             $      GO TO 280
  470 DA=A3-A2             $        DB=B3-B2
      A4=A3+DA             $        B4=B3+DB
      A5=A4+DA             $        B5=B4+DB             $      GO TO 280
  480 A5=A4=A3             $        B5=B4=B3             $      GO TO 280

C       DETERMINATION OF THE DIRECTION

  500 SGN=1.0
      IF(R23.LE.0.0)                                            GO TO 550
      IF(R12.LE.0.0.AND.R34.LE.0.0)                             GO TO 580
      IF(R13.LE.0.0.AND.R24.LE.0.0)                             GO TO 580
      IF(R12.LE.0.0.OR. R24.LE.0.0)                             GO TO 560
      IF(R13.LE.0.0.OR. R34.LE.0.0)                             GO TO 570
      S2=S12*S24           $        S3=S13*S34
      IF(S2*S3.LT.0.0)              S3=-S3
      A=S2*A3*A3-S3*A2*A2
      B=S2*A3*B3-S3*A2*B2
      C=S2*B3*B3-S3*B2*B2
      D=S23*SQRTF(S2*S3)
      IF(B*D.LT.0.0)                D=-D                 $      B=B+D
      S20=A2*B0(1)-A0(1)*B2
      S03=A0(1)*B3-A3*B0(1)
      IF(S20*S03.LE.0.0)                                        GO TO 510
      COS3=A0(1)           $        SIN3=B0(1)           $      GO TO 520
  510 S20=A2*B0(2)-A0(2)*B2
      COS3=A0(2)           $        SIN3=B0(2)
  520 IF(S20*S23.GT.0.0)                                        GO TO 590
      COS3=-COS3           $        SIN3=-SIN3           $      GO TO 590
  550 IF(C23.LT.0.0)                SGN=-1.0
  560 COS3=A2              $        SIN3=B2              $      GO TO 590
  570 COS3=A3              $        SIN3=B3              $      GO TO 590
  580 COS3=A2+A3           $        SIN3=B2+B3
  590 IF(MD.LE.2)                   GO TO LBL, (240,600)
      R=SQRTF(COS3*COS3+SIN3*SIN3)
      COS3=COS3/R          $        SIN3=SIN3/R
      GO TO LBL, (240,600)
```

```
C     INTERPOLATION IN A SECTION

  600 KS0=(I-2)*M+1          $     Z=0.0
      IF(MD.GT.2)                  GO TO 660
  610 KS1=KS0
      P1=DX
      Q1=P1*SIN2/COS2
      Q2=3*DY-2*Q1-P1*SIN3/COS3
      Q3=DY-Q1-Q2
      DO 620 JS=1,MM1
      KS1=KS1+1              $     Z=Z+DZ
      U(KS1)=P0+Z*P1
  620 V(KS1)=Q0+Z*(Q1+Z*(Q2+Z*Q3))                  $     GO TO 290
  660 KS2=KS0
      R=SQRTF(DX*DX+DY*DY)
      P1=R*COS2
      P2=3*DX-R*(2*COS2+COS3)
      P3=DX-P1-P2
      Q1=R*SIN2
      Q2=3*DY-R*(2*SIN2+SIN3)
      Q3=DY-Q1-Q2
      DO 670 JS=1,MM1
      KS2=KS2+1              $     Z=Z+DZ
      U(KS2)=P0+Z*(P1+Z*(P2+Z*P3))
  670 V(KS2)=Q0+Z*(Q1+Z*(Q2+Z*Q3))                  $     GO TO 290

C     NORMAL RETURN

  800 IF(MD.NE.2)                  GO TO 890
      DO 810 KR=1,N
      TS=U(KR)               $     U(KR)=V(KR)      $     V(KR)=TS
  810 CONTINUE
  890 L0=LPS                 $     N0=N             $     RETURN

C     ERROR EXIT

  900 DER2=DER1              $     CALL Q8QERROR(0,MSG)
      END
```


GPO 849-390