NASA Technical Memorandum 72854

AN EFFICIENT ALGORITHM FOR CHOOSING THE DEGREE OF A POLYNOMIAL TO APPROXIMATE DISCRETE NONOSCILLATORY DATA

David R. Hedgley

April 1978

Center-printed Technical Memorandums are issued to provide rapid transmittal of technical information from the researcher to the user. As such, they are not subject to the usual NASA review process.

NASA

NASA Technical Memorandum 72854

AN EFFICIENT ALGORITHM FOR CHOOSING THE DEGREE OF A POLYNOMIAL TO APPROXIMATE DISCRETE NONOSCILLATORY DATA

David R. Hedgley
Dryden Flight Research Center
Edwards, California

National Aeronautics and
Space Administration

Scientific and Technical
Information Office

1978

AN EFFICIENT ALGORITHM FOR CHOOSING THE DEGREE OF A POLYNOMIAL TO APPROXIMATE DISCRETE NONOSCILLATORY DATA

David R. Hedgley
Dryden Flight Research Center

INTRODUCTION

The least-squares method for curve fitting, or defining a curve that best approximates a data set, is well known and is used in almost every technical discipline. However, there has always been the recurrent problem of how to efficiently choose the degree of the polynomial to obtain a fit that is at least moderately good without any prior information concerning the nature of the data. Although different types of data are approached by various methods, oscillatory (high-frequency) data are usually smoothed, while nonoscillatory (low-frequency) data are usually approximated.

The algorithm discussed in this paper deals primarily with the efficiency of choosing the degree of the polynomial as it relates to nonoscillatory data but does not neglect the smoothing of high-frequency data.

SYMBOLS

a_j	coefficient of polynomial
$F(a_0, \ldots, a_m)$	function of m variables
I, K	arbitrary constants
L	degree of approximating polynomial
NP	total number of peaks
$P(x)$	polynomial of any degree
$P'(x)$	first derivative of $P(x)$
$P''(x)$	second derivative of $P(x)$

TP	total number of data points
w	weighting function
x_i, y_i	discrete data points
ε_0, ε_1	arbitrary tolerance values
σ	residual

DISCUSSION

The least-squares approach to curve fitting using a polynomial as a model is based on the minimization of the sum of the squares of the differences between the data and a polynomial of degree m evaluated at corresponding given observations. (Note that the use of least squares assumes the errors in the data to be normally distributed with a mean of zero.) The function to be minimized is

$$F(a_0, a_1, \ldots, a_m) = \sum_{i=1}^{TP} w(x_i)\left(y_i - \sum_{j=0}^{m} a_j x_i^j\right)^2$$

where m is the degree of the model, x_i and y_i are discrete data points, a_j is the coefficient to be determined, $w(x_i)$ is the weighting factor, and TP is the number of data points (ref. 1). If the data can be accurately represented by a polynomial, then $\sigma_m = F(a_0, a_1, \ldots, a_m)$ tends to zero as m approaches L, where L is the degree of the approximating polynomial (ref. 1). However, in practice, low order polynomials are preferred for curve fitting, and thus, it is sufficient to notice the magnitude of change in σ_m as m increases. In other words, if

$$\left|(\sigma_m/\sigma_{m+1}) - 1\right| < \varepsilon_0 < 1$$

a polynomial of degree m + 1 is considered sufficient (ref. 1). (An additional restraint,

$$R = \left|\left(y_i - w(x_i) \sum_{j=0}^{m} a_j x_i^j\right)/y_i\right| < \varepsilon_1$$

where R represents the maximum percentage of error incurred over the entire collection of points, would be desirable as it would provide an alternate criterion for evaluating the quality of the fit.) One of the least desirable features of this conventional approach is the need to examine σ_m for $1 \leqslant m \leqslant L$ when the data are

nonoscillatory. This procedure results in a needless waste of computer time in converging to L if $L \geqslant 2$. This problem is addressed in the following discussion.

Let K be a constant such that $0 < K < 1$. The data are considered to be oscillatory if one of the following conditions is true: (1) $NP/TP > K$ where TP represents the total number of data points and NP is the total number of peaks, or (2) $NP > I$ where I is a constant integer (for example, $I = 10$). A peak is considered to have occurred if

$$(y_i - y_{i+1})(y_{i+1} - y_{i+2}) < 0$$

where $x_i < x_{i+1} < x_{i+2}$. If a polynomial, $P(x)$, is sought to smooth data that is considered oscillatory based on condition (1) or (2) above, then by observing σ_m as m increases and by applying the above criterion for quality of fit, it can be observed that $\left|(\sigma_m/\sigma_{m+1}) - 1\right|$ approaches zero for small values of m almost without exception. Of course by choosing a large enough value of m, one could match the data closely, but the smoothing properties of the curve would be sacrificed. Moreover, the degree of the polynomial would be sufficiently large to make the computational time prohibitive. Therefore, this technique selects polynomials of small degree for oscillatory data and, hence, tends to smooth the data.

If, on the other hand, the data are determined to be nonoscillatory and are assumed to have no wild points, the following analysis applies.

Let NP be the total number of peaks and $P(x)$ be the desired polynomial that approximates the data. From the definition of a peak and because a polynomial and its derivatives are continuous, there exists a point (x_0, y_0), where $x_i \leqslant x_0 \leqslant x_{i+1}$, such that $P'(x_0) = 0$. That is, $P(x_0)$ is a local extremum or peak for $P(x)$. Hence, if there are NP local extrema, there are at least NP real zeros for $P'(x)$. However, this implies that the degree of $P'(x)$ must be at least NP, which implies that the degree of $P(x)$ must be at least $NP + 1$.

This argument can be extended to second derivatives. That is,

$$\left[(y_{i+2} - y_{i+1})/(x_{i+2} - x_{i+1}) - (y_{i+1} - y_i)/(x_{i+1} - x_i)\right]\left[(y_{i+1} - y_i)/(x_{i+1} - x_i) - (y_i - y_{i-1})/(x_i - x_{i-1})\right] < 0$$

implies that there exists a point x_0, where $x_i \leqslant x_0 \leqslant x_{i+1}$, such that $P''(x_0) = 0$, where $P(x)$ is the approximating polynomial. Let Nx be the sum of all such occurrences. It is then sufficient to examine the value $\left|1 - (\sigma_m/\sigma_{m+1})\right|$ for $1 \leqslant NC \leqslant m \leqslant L$ where L is the degree of the approximating polynomial and $NC = \text{maximum}\ [(NP + 1), (Nx + 2)]$. This approach is an improvement over the conventional methods, and the improvement should increase as the degree of the polynomial required increases. Finally, an additional advantage of this technique is the increased probability of obtaining a good fit. When using the conventional method, the ratio σ_m/σ_{m+1} may approach 1 with m small and σ_m large, which would cause the iteration to terminate too soon. This obviously would produce

a rather poor approximating polynomial when the data are nonoscillatory but have several extrema.

These observations indicate that both accuracy and efficiency can be improved with the proposed method. Illustrative examples are given in the appendix.

CONCLUDING REMARKS

While the approach presented in this paper offers nothing new for oscillatory data, it defines a criterion that discriminates between oscillatory and nonoscillatory data and attempts to handle both without loss of generality. The approach eliminates the need to examine the residuals of polynomials with degrees from 1 to the degree of the approximating polynomial for nonoscillatory data without sacrificing the performance of the least-squares method as it relates to oscillatory data. It also increases the probability of selecting a good approximating polynomial for nonoscillatory data.

NASA Dryden Flight Research Center
Edwards, Calif., November 9, 1977

APPENDIX—ILLUSTRATIVE EXAMPLES

Figures 1 and 2 are examples of oscillatory and nonoscillatory data that are smoothed and approximated using the method presented in the body of this report. These examples show the generality of the proposed method. Figure 3 shows the weakness of the conventional method. From these examples, it is clear that the proposed method is an improvement over the conventional technique for automatically choosing the degree of a polynomial for curve fitting.

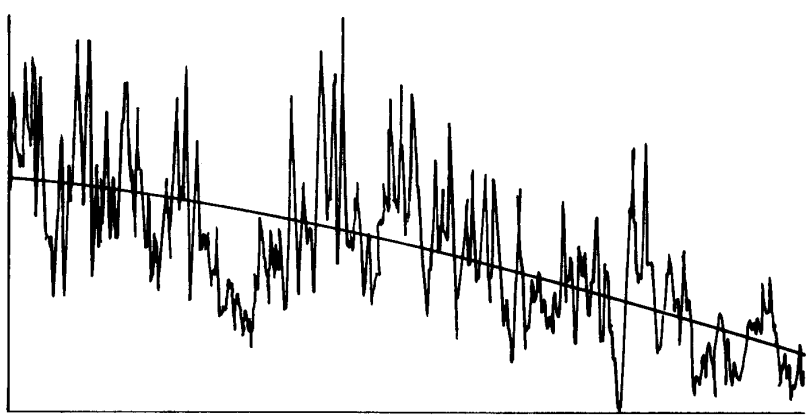

Figure 1. Oscillatory data smoothed with proposed method.

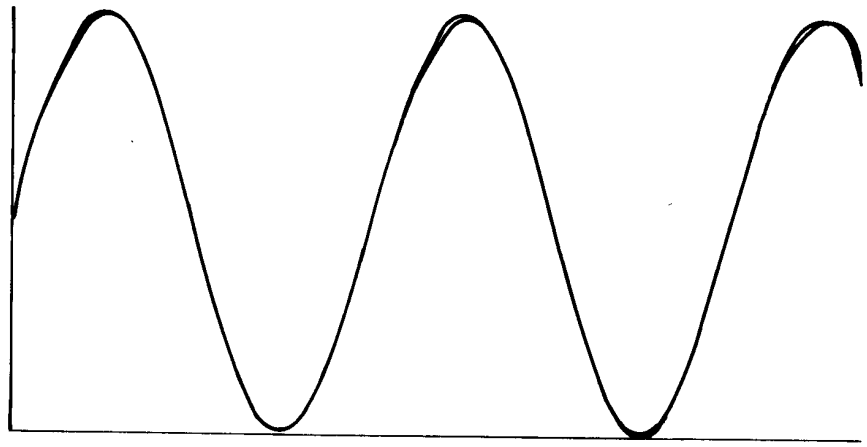

Figure 2. Nonoscillatory data smoothed with proposed method.

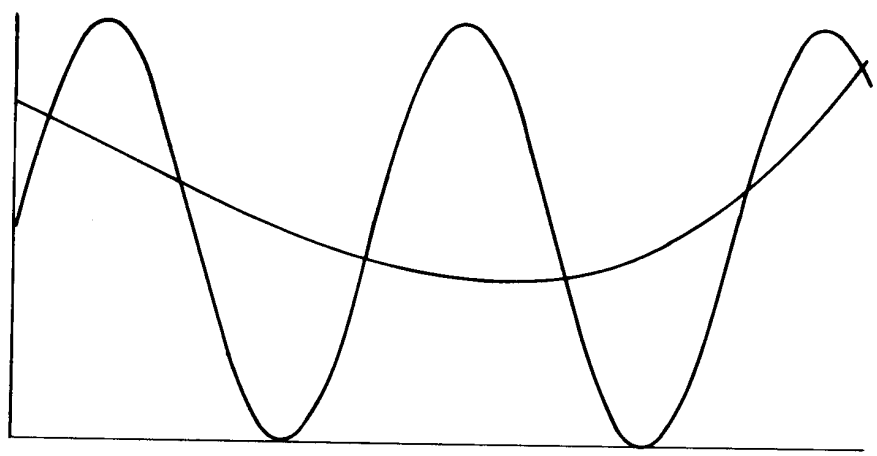

Figure 3. Nonoscillatory data approximated with conventional method.

REFERENCE

1. Ralston, Anthony: A First Course in Numerical Analysis. International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., Inc., c.1965, pp. 228-265.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM-72854		

4. Title and Subtitle	5. Report Date
AN EFFICIENT ALGORITHM FOR CHOOSING THE DEGREE OF A POLYNOMIAL TO APPROXIMATE DISCRETE NONOSCILLATORY DATA	April 1978
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
David R. Hedgley	H-1010
9. Performing Organization Name and Address	10. Work Unit No.
Dryden Flight Research Center P.O. Box 273 Edwards, California 93523	312-04-02
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Aeronautics and Space Administration Washington, D.C. 20546	Technical Memorandum
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

An efficient algorithm for selecting the degree of a polynomial that defines a curve that best approximates a data set is presented. This algorithm is applied to both oscillatory and nonoscillatory data without loss of generality.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Polynomials Computer programs Curve fitting	Unclassified - Unlimited Category: 64

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	9	$3.25

*For sale by the National Technical Information Service, Springfield, Virginia 22161