Applied and Computational Mathematics Division

NISTIR 89-4210

Center for Computing and Applied Mathematics

An Algorithm and Computer Program for the Calculation of Envelope Curves

M. McClain, A. Feldman, D. Kahaner, and X.T. Ying

December 1989

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
Gaithersburg, MD 20899

INTRODUCTION

In many instances, experimental data consist of a series of oscillations bounded by upper and lower curves which are envelopes. Such behavior is frequently observed when interference effects are superimposed on a slowly varying trend. By knowing the envelope functions one may be able to deduce physical properties giving rise to the observed phenomenon. An example of this is the envelope method used by Manifacier et al.[1] to obtain the optical constants and thicknesses of weakly absorbing films from transmission spectra. It would therefore be useful if a simple automated procedure could be used to obtain the envelope functions, especially in numerical form. In this paper we describe an iterative method for calculating the envelope curves of a given set of oscillatory data. In an example, we analyze a set of transmission data on the basis of the paper by Manifacier et al.[1] and compare the results with a manual method.

THE PROCEDURE

The envelopes of a given oscillatory function $T(x)$ are two smooth curves that in some sense represent the maximum and minimum values of $T(x)$. By "smooth" it is meant here that the envelopes have very few inflection points compared with $T(x)$; often they have no inflection points at all. The envelopes are constrained to lie tangent to $T(x)$ and not to cross it. The top envelope $T_{max}(x)$ lies above the function, i.e., $T_{max}(x) \geq T(x)$, while the bottom envelope $T_{min}(x)$ lies below the function, i.e., $T_{min}(x) \leq T(x)$. Figure 1 shows a schematic representation of an oscillatory function T(x) and its envelopes $T_{max}(x)$ and $T_{min}(x)$. Note that the points of tangency are not the same as the critical points (i.e., the local extrema) of T(x). The tangent points often lie near the critical points, but this is not always the case; see for example the left-most tangent point in Figure 1.

The following steps outline a procedure for calculating envelope curves:

Step 1. Smooth the given data.

Step 2. Estimate the locations of the upper and lower tangent points.

Step 3. Interpolate a curve through the estimated upper tangent points and another through the estimated lower tangent points.

Step 4. If no points on the upper curve lie below the smoothed data and no points on the lower curve lie above the smoothed data, then stop. The upper curve is the top envelope T_{max}, and the lower curve is the bottom envelope T_{min}.

Step 5. Otherwise, improve the estimates of the tangent point locations and return to Step 3.

In what follows, the calculation of the top envelope is discussed in detail. The calculation of the bottom envelope is exactly analogous.

The only input required by the envelope algorithm is the set of data points (x,y) whose envelope is to be computed. Due to the likely presence of small errors in the data, the original set of data points is not used directly but is replaced by a smoother set (Step 1 above). The smoothed data points are obtained by evaluating a spline function $T(x)$ which is computed to be a least squares fit to the original data. $T(x)$ is required to lie within a specified error tolerance ϵ of each original data point; that is, for each data point (x,y), $|T(x)-y| \leq \epsilon$. Two separate smoothing functions are employed in the course of the envelope calculation: a rough initial smoothing $T_0(x)$ that is used in the first estimation of the tangent point locations and a more accurate final smoothing $T(x)$ that is used in determining the envelope curve.

The envelope algorithm starts by estimating the locations of the points where $T_{max}(x)$ is tangent to $T_0(x)$ (Step 2 above). The obvious choices for these tangent point estimates are the local maxima of $T_0(x)$. However, as noted above, not every tangent point is near a maximum point, so a more generally applicable estimation method is employed. This algorithm first finds the intervals where $T_0(x)$ opens downward, i.e., where $T_0''(x) < 0$. It then places a tangent point estimate halfway between the endpoints of each of these intervals, as shown in Figure 2. The endpoints are the inflection points of $T_0(x)$, i.e., points where $T_0''(x) = 0$. (The second derivative information needed for determining the inflection points and the direction of curvature is obtained during the smoothing process in Step 1.)

Downward-opening intervals containing the first or last data point ("endpoint intervals") require special treatment. In some cases, the data for these intervals might be incomplete; that is, if the data could be extended, $T_0(x)$ would continue to open downward for some interval beyond the original endpoint. The true tangent point for such an interval might then lie outside the range of the original data, as shown in Figure 3. The envelope algorithm does not attempt to estimate such tangent points. It will produce a tangent point estimate for an endpoint interval only if it contains a point where $T_0'(x) = 0$, and the tangent point estimate will be set to be this point, as shown in Figure 4(a). For an endpoint interval that does not contain such a point, no tangent point estimate will be produced and the resulting envelope curve will stop with the neighboring downward-opening interval, as shown in Figures 4(b) and 4(c). Recall again that not every tangent point is near a maximum point, so this method for handling endpoint intervals will occasionally miss valid tangent points, as in Figure 4(c), but it will not include invalid ones.

The initial estimation of the tangent points is the most critical part of the envelope calculation, and it depends heavily on the initial smoothing of the data. If the smoothed curve $T_0(x)$ does not match the data closely enough, there will be too few inflection points and hence too few tangent points. On the other hand, if $T_0(x)$ matches the data too closely there are likely to be many undesirable inflection points, and these will produce spurious tangent point estimates, as shown in Figure 5. It may be necessary to experiment with various error tolerances ϵ before finding a $T_0(x)$ that leads to appropriate tangent point estimates.

Next, a first approximation to the envelope curve is obtained by interpolating a smooth curve through the estimated tangent points (Step 3 above). The interpolated curve $I(x)$ is required to preserve monotonic behavior; that is, if the tangent points increase (or decrease) with x, the resulting envelope curve does also. If the tangent points are not monotonic, the interpolation method forces an extreme point in the envelope curve at any point where the direction of monotonicity changes.

The first approximation to the envelope curve will generally not be truly tangent to the smoothed data curve, unless the initial tangent point estimates happen to be exactly correct. In most cases, the interpolated curve will cut below the peaks of the data curve, as shown in Figure 6. Around each of the estimated tangent points, there will be an interval where $I(x) < T(x)$. (The estimated tangent point will actually be one of the endpoints of the interval.) The points of tangency of the true envelope $T_{max}(x)$ with $T(x)$ must lie somewhere on these intervals. The algorithm proceeds (Step 5 above) by taking the midpoints of each of these intervals as new estimates for the tangent points, as shown in Figure 7.

When checking for envelope points that lie below $T(x)$, it is generally desirable to use a smoothing fit that tightly matches the original data. This is especially necessary in the regions closely surrounding the tangent points, in order that the final envelope curve will be as accurate as possible. The initial smoothing function $T_0(x)$ used to obtain the first tangent point estimates may not have sufficient accuracy to serve as the final smoothing, so it is generally necessary to try another spline fit T(x) with a smaller error tolerance ϵ in order to produce the desired accuracy in the final envelope curve.

When the new tangent point estimates have been computed, a new curve I(x) is interpolated through them, which should lie closer to the true envelope. This process of finding new estimates for the tangent points and interpolating a curve through them is repeated until either no points on the interpolated curve lie beneath the data curve (Step 4 above) or until a maximum of twenty iterations have been performed. The resulting interpolated curve I(x) should lie very close to the true envelope $T_{max}(x)$.

We have implemented this envelope algorithm as a FORTRAN program running on an IBM PC or compatible computer. To smooth the data, we use the least squares spline-fitting subroutine EFC[2], augmented by a subroutine that automatically generates breakpoints for the spline (i.e., the points where the polynomial pieces of the spline are joined). To interpolate a monotonic function through the tangent estimates, we use the subroutine PCHIM[3]. Both of these subroutines are part of SLATEC[4], a public domain mathematical software library available from Argonne National Laboratory. The main program allows the user to request a data set and to enter the tolerance factors to be used for the data smoothing. The data and the resulting envelope curves are then plotted on the screen. The user has several options: (1) redo the envelope calculation with different error tolerances, (2) produce a high-quality hard copy of the plot, or (3) save the envelope data and tangent point locations in a file.

A typical wall-clock running time for the program on an IBM PC/AT is about 1.75 minutes for a data set containing 800 points. (This time does not include plotting and saving the results.) Approximately 1.25 minutes are spent on smoothing the data, while 0.50 minutes are spent on calculating the envelopes.

AN APPLICATION

Figure 8 shows the optical transmission of film composed of a mixture of yttria and silica deposited on a fused silica substrate. Shown superimposed on the data are the envelope functions computed by the procedure described above. The method of Manifacier et al.[1] has been used to obtain the refractive index, n, the absorption coefficient, α, and the thickness, t, of the film from the envelope functions.

The transmittance, T, of a weakly absorbing film on a transparent substrate can be represented by

$$T = \frac{n_o n^2 n_s A}{C_1^2 + C_2^2 A^2 + 2C_1 C_2 A \cos(4\pi n t/\lambda)} \qquad (1)$$

where $C_1 = (n+n_o)(n_s+n)$, $C_2 = (n-n_o)(n_s-n)$, $A = \exp(-4\pi k t/\lambda) = \exp(-\alpha t)$, n_o is the refractive index of the ambient (air), n_s is the refractive index of the substrate, k is the imaginary part of the refractive index of the film, and λ is the wavelength of the radiation. In general, each of the parameters except t are functions of λ. Equation (1) is an oscillatory function with envelope curves given by

$$T_{max} = \frac{n_o n^2 n_s A}{(C_1 + C_2 A)^2} \tag{2}$$

and

$$T_{min} = \frac{n_o n^2 n_s A}{(C_1 - C_2 A)^2} \quad , \tag{3}$$

as can be seen by taking the cosine to be +1 or -1. While there is no guarantee that the envelope algorithm described in this paper will produce the curves given by equations (2) and (3), it is believed to give a close approximation. The procedures for calculating n, α (or k) and t from the envelope functions and the points of tangency will not be repeated here, as they are discussed in Manifacier et al. The results are shown in Figure 9 where n is plotted as a function of λ and in Figure 10 where α is plotted as a function of λ. Figure 9 also shows the values obtained by an earlier manual procedure[5]. We see that the absorption coefficient data agree quite well. The refractive index data also agree reasonably well. (Note that the vertical scale is expanded and does not begin at zero.) The value obtained for thickness was 0.58 μm which agrees reasonably well with a mechanical thickness measurement of 0.60 μm and a value from the manual data analysis procedure of 0.56 μm.

SUMMARY

A procedure has been developed to numerically calculate the envelope functions of an oscillatory curve. The method has been shown to be applicable to optical transmission data, but it is general enough to be used for many other data sets. The computer program is available on request.

REFERENCES

1. J. C. Manifacier, J. Gasiot, and J. P. Fillard, "A simple method for the determination of the optical constants n, k and the thickness of a weakly absorbing film," J. Phys. E.: Sci. Instr. **9**, 1002 (1976).

2. R. J. Hanson, Constrained Least Squares Curve Fitting to Discrete Data Using B-Splines, a User's Guide, Sandia Laboratories Technical Report SAND-78-1291, 1978.

3. F. N. Fritsch and R. E. Carlson, "Monotone Piecewise Cubic Interpolation," SIAM J. Numer. Anal. **17**, 238 (1980).

4. B. L. Buzbee, "The SLATEC Common Mathematical Library," in Sources and Development of Mathematical Software, W. R. Cowell, ed. (Prentice-Hall, Englewood Cliffs, New Jersey, 1984).

5. A. Feldman, X. T. Ying, and E. N. Farabaugh, "Optical Properties of Mixed Yttria-Silica Films," Applied Optics, in press.

FIGURE CAPTIONS

Figure 1. Schematic representation of an oscillatory function and its envelopes. The dots represent the points of tangency of the function with its envelopes.

Figure 2. A method for obtaining initial estimates of tangent points. A tangent estimate is set at the midpoint of each downward-opening interval bounded by two inflection points.

Figure 3. A problem with estimating tangent points near the ends of the data. In this figure, the right-most downward-opening interval of the data curve is incomplete. If it could be extended (as the dotted curve indicates), the true tangent point would be seen to lie beyond the end of the original data. Using any point in the existing data as a tangent point estimate would produce an "envelope curve" very different from the true envelope curve.

Figure 4. Examples of a method for estimating tangent points for endpoint intervals. In (a), both endpoint intervals contain points where the derivative is zero, implying that the tangent points for these intervals probably lie nearby, within the range of the data. Tangent point estimates are produced for both intervals. In (b), the righthand endpoint interval does not contain a point where the derivative is zero, so there is no guarantee that the tangent point lies within the range of the data. No tangent point estimate is produced for this interval. In (c), neither endpoint interval contains a point where the derivative is zero, so no tangent point estimate is produced for either end. In this picture, the method misses two valid tangent points.

Figure 5. A spurious tangent point, caused by a small wiggle in the smoothed data curve. This leads to an unacceptable envelope curve.

Figure 6. The first approximation to the envelope curve usually lies below the true envelope.

Figure 7. Obtaining an improved tangent point estimate. The first approximation to the envelope curve cuts off the peaks of the data curve. A new tangent point estimate is set at the midpoint of each cut-off interval.

Figure 8. Transmittance vs. wavelength of a mixed yttria-silica film after subtraction of reflection from the air-silica surface of the substrate. This figure, showing experimental data, smoothed data, and fitted envelope curves, was produced directly by the envelope program.

Figure 9. Refractive index vs. wavelength of a mixed yttria-silica film. The squares represent results obtained by a prior manual analysis, the solid curve represents results calculated from the fitted envelope curves, and the dashed curve represents a fit to the manual data.

Figure 10. Absorption coefficient vs. wavelength of a mixed yttria-silica film. The squares represent results obtained by a prior manual analysis and the solid curve represents results calculated from the fitted envelope curves.

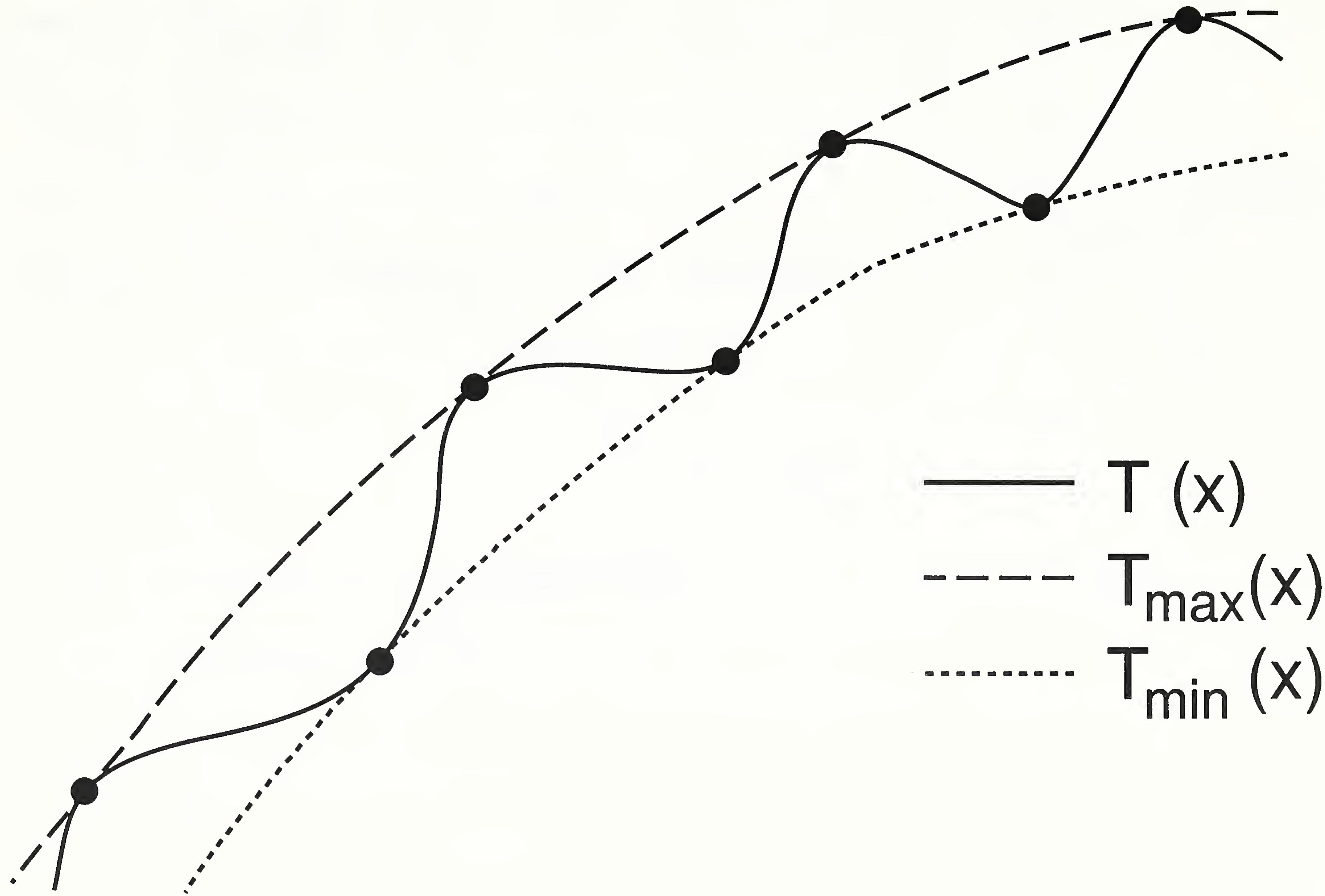


FIGURE 1

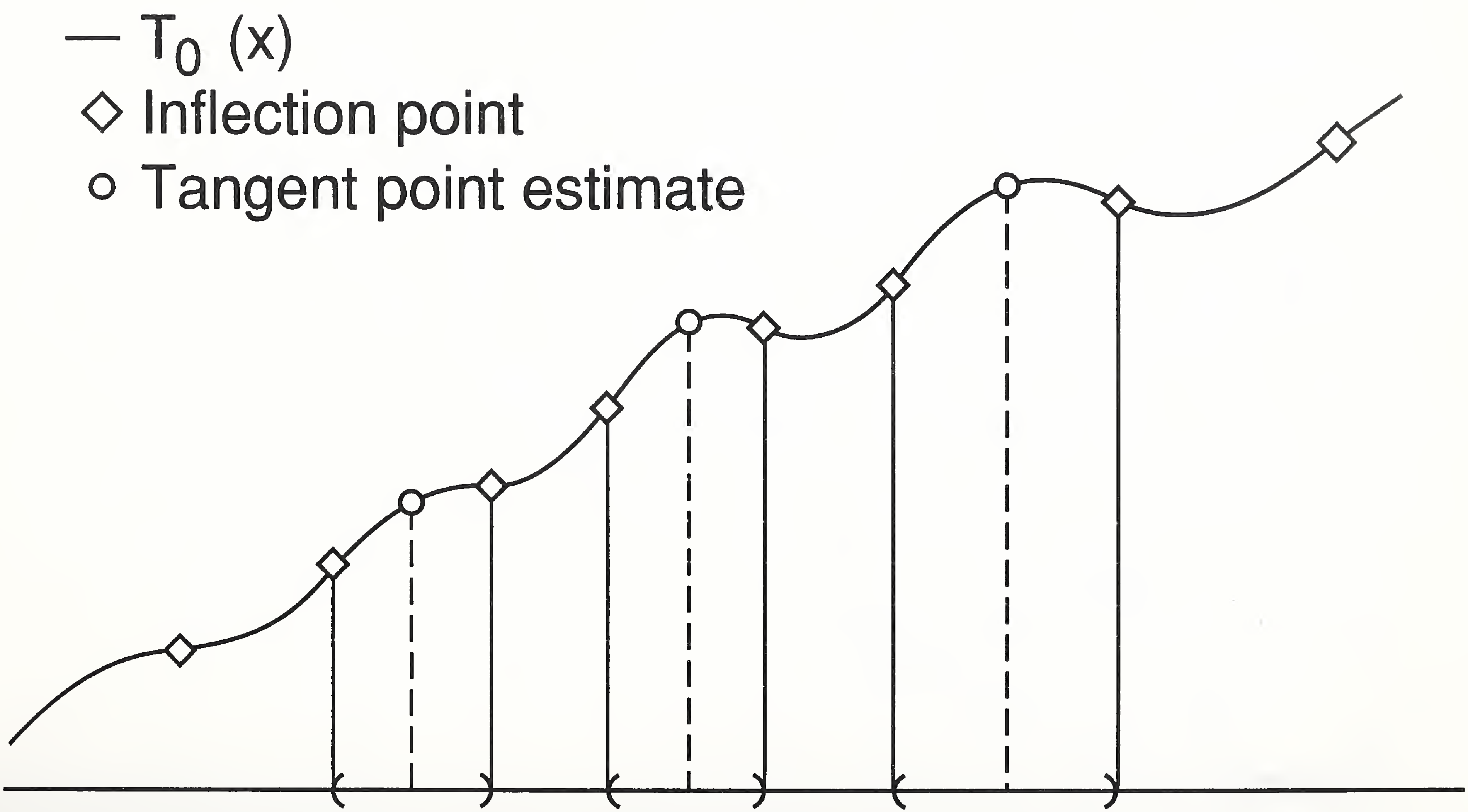


FIGURE 2

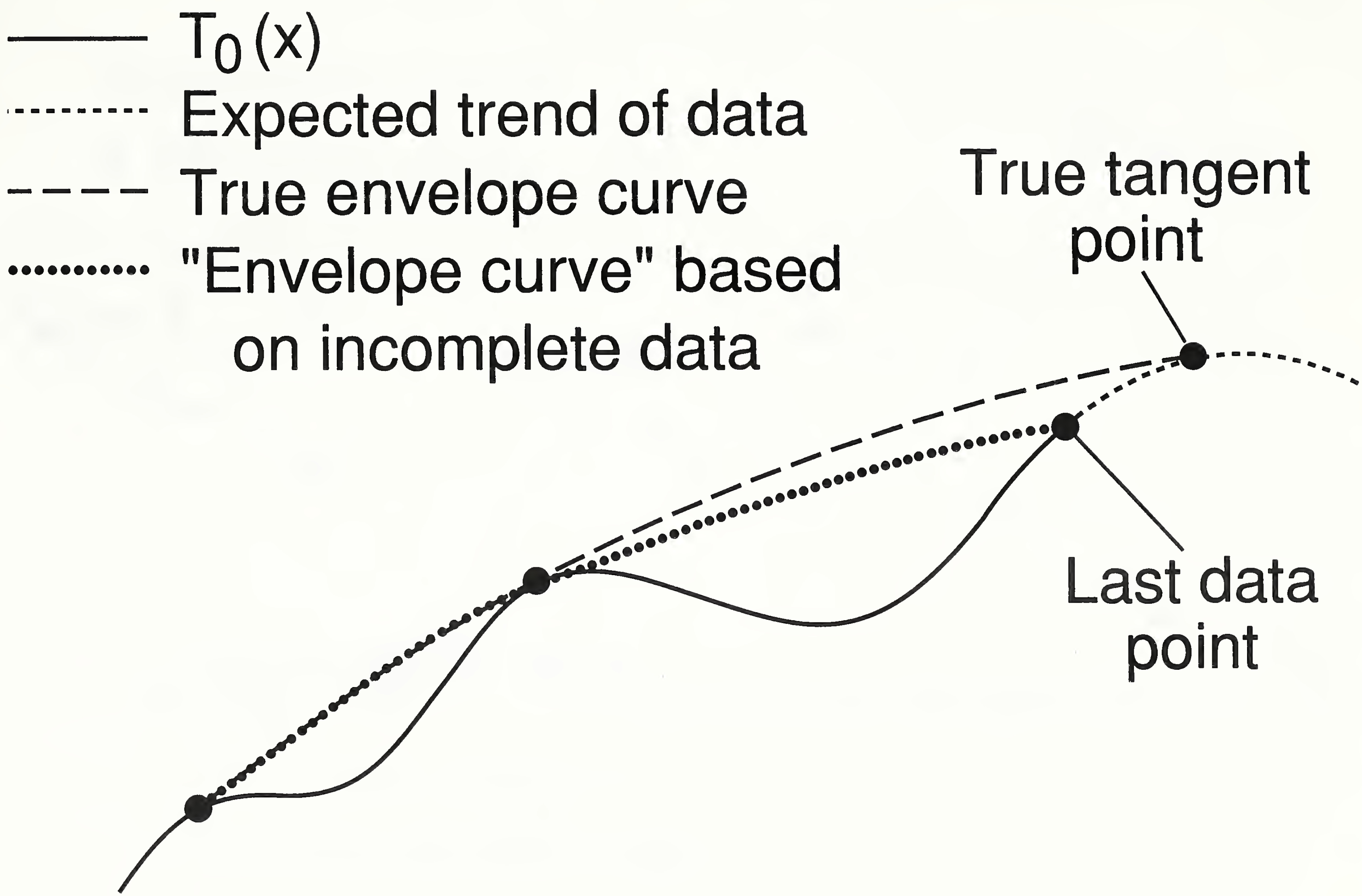
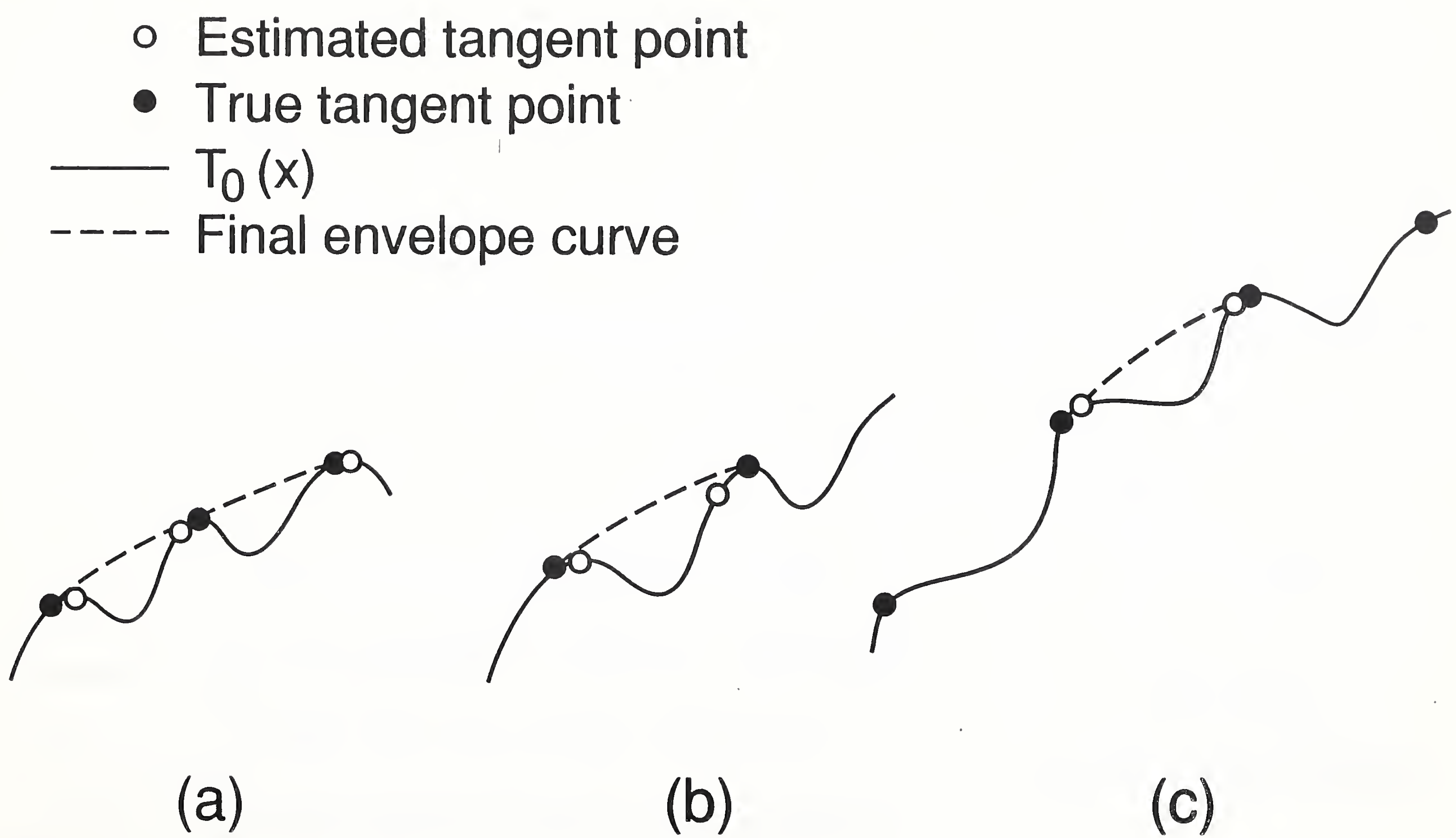

$T_0(x)$
Expected trend of data
True envelope curve
"Envelope curve" based on incomplete data
True tangent point
Last data point

FIGURE 3



Estimated tangent point
True tangent point
T0 (x)
Final envelope curve
(a)
(b)
(c)

FIGURE 4

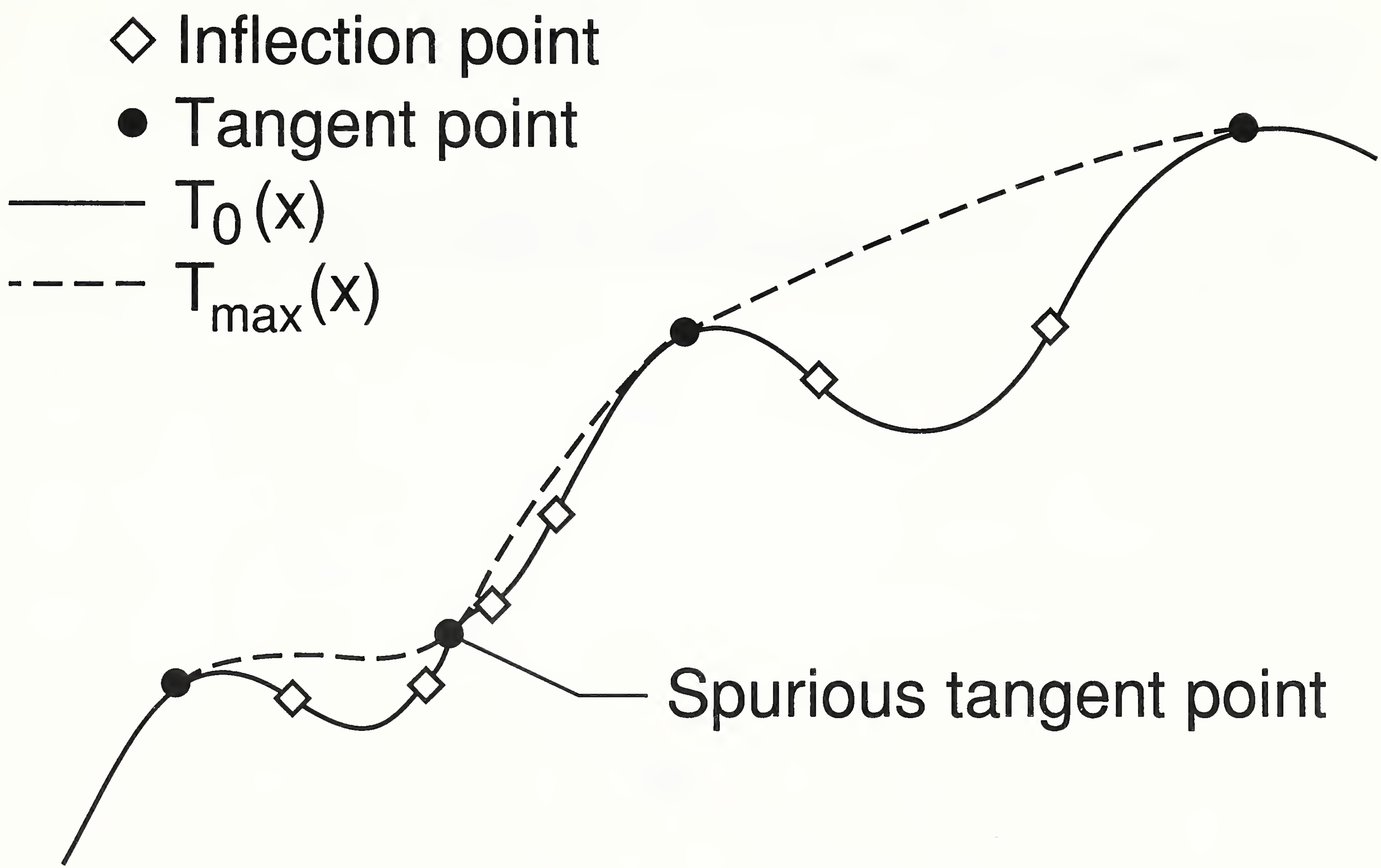

◇ Inflection point
● Tangent point
$T_0(x)$
$T_{max}(x)$
Spurious tangent point

FIGURE 5

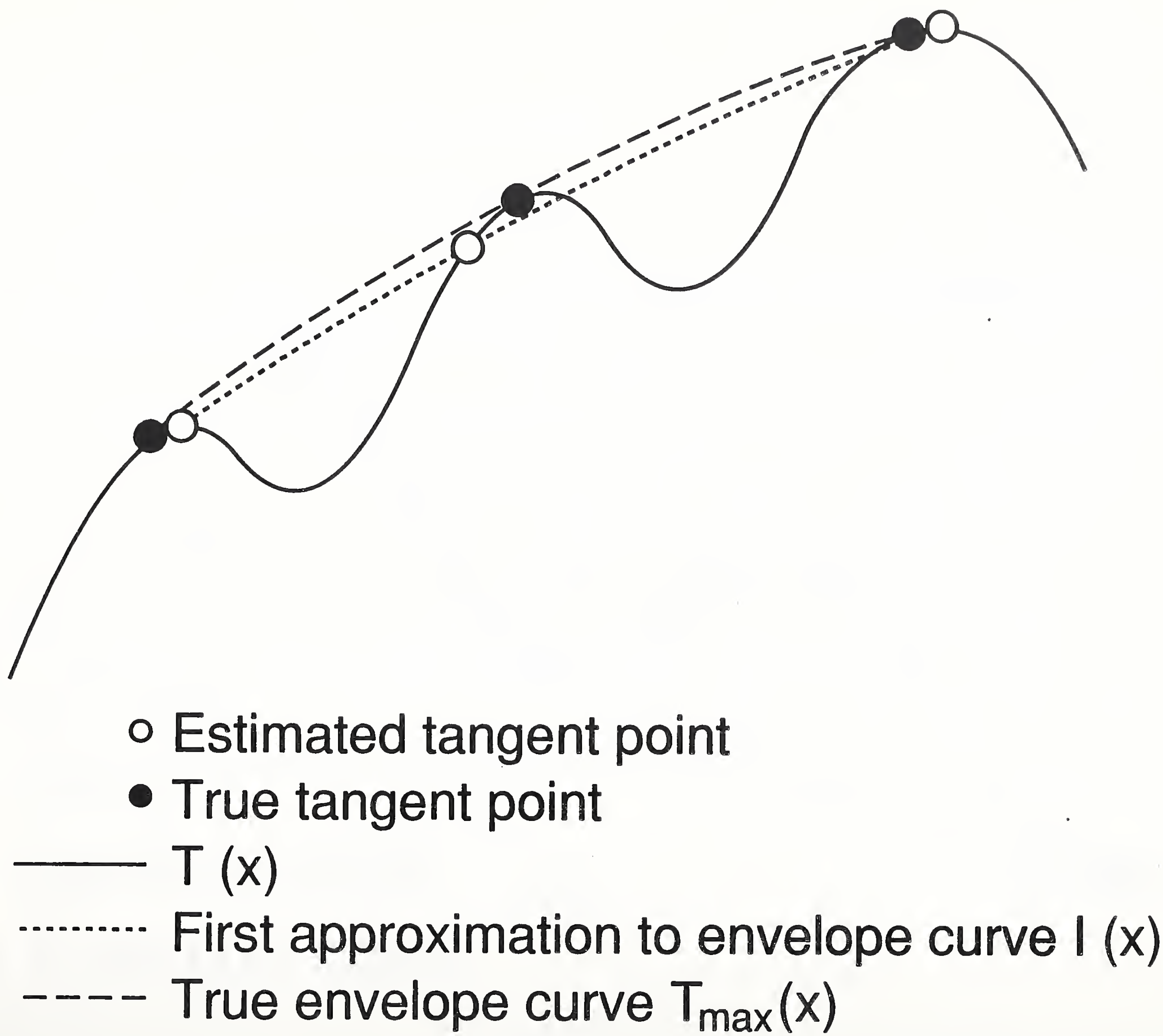

○ Estimated tangent point
● True tangent point
T (x)
First approximation to envelope curve I (x)
True envelope curve $T_{max}(x)$

FIGURE 6

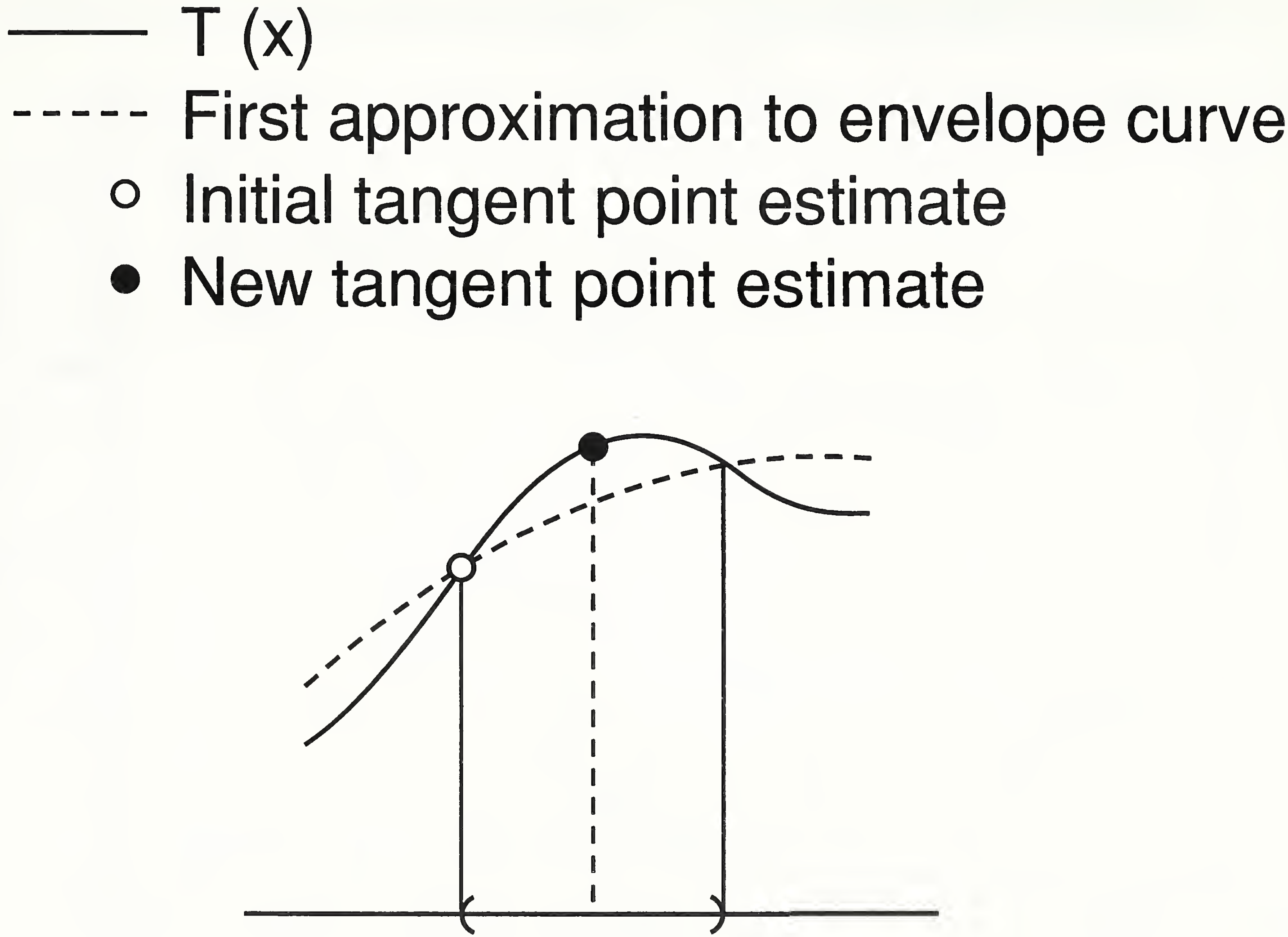

T (x)
First approximation to envelope curve
Initial tangent point estimate
New tangent point estimate

FIGURE 7

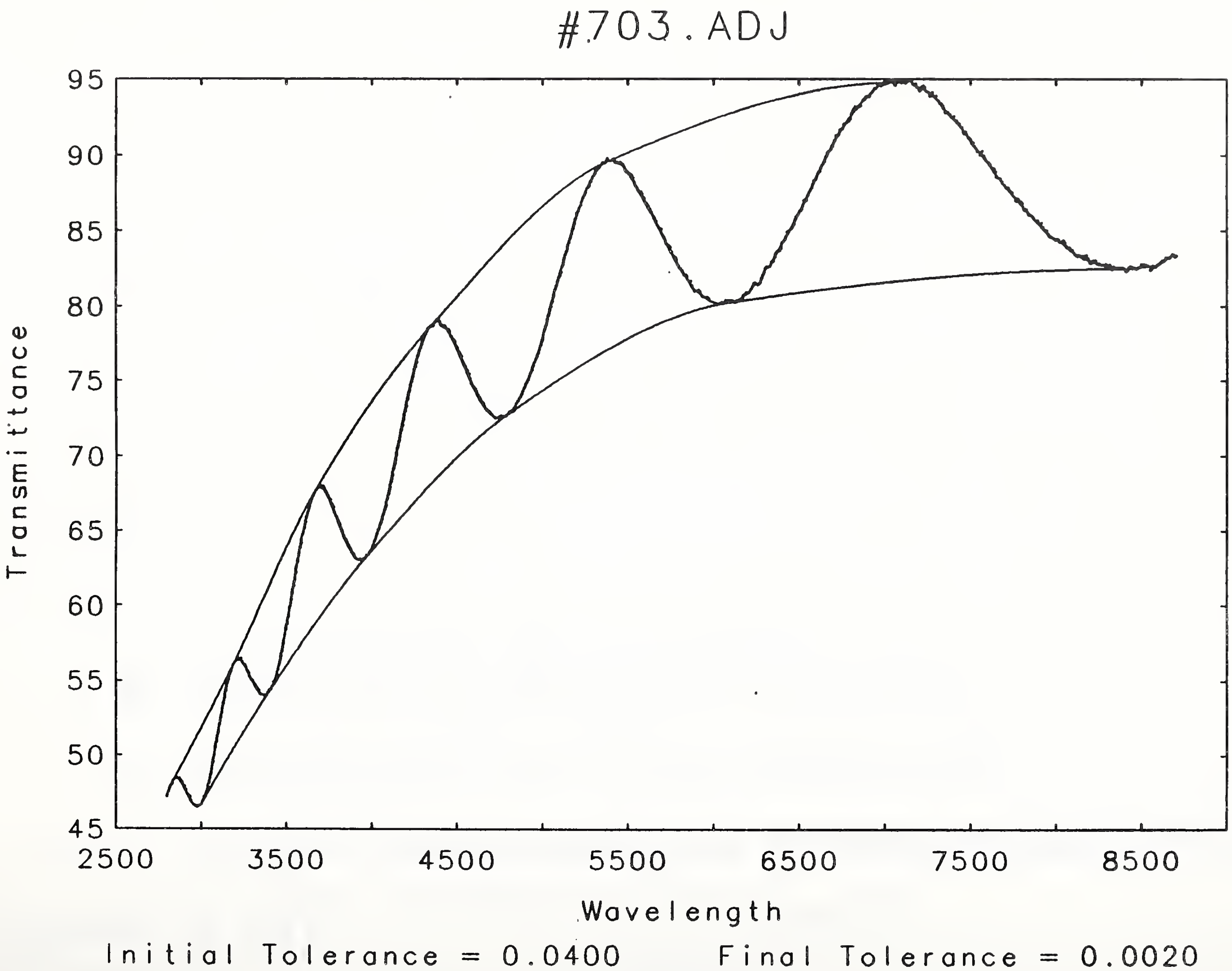
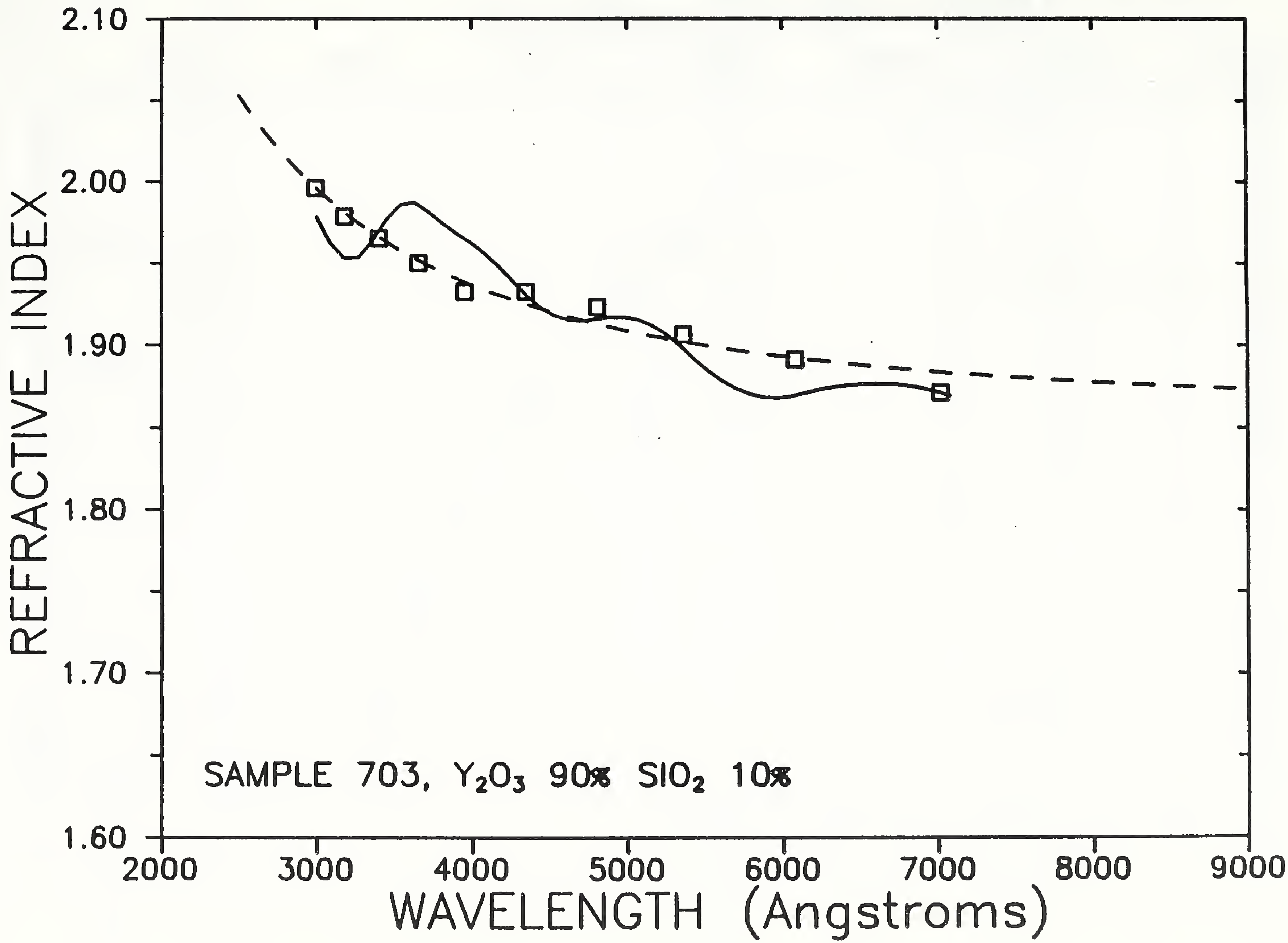

#703.ADJ
95
90
85
80
75
70
65
60
55
50
45
Transmittance
2500
3500
4500
5500
6500
7500
8500
Wavelength
Initial Tolerance = 0.0400   Final Tolerance = 0.0020

FIGURE 8



2.10
2.00
1.90
1.80
1.70
1.60
REFRACTIVE INDEX
2000
3000
4000
5000
6000
7000
8000
9000
WAVELENGTH (Angstroms)
SAMPLE 703, Y_2O_3 90% SIO_2 10%

FIGURE 9

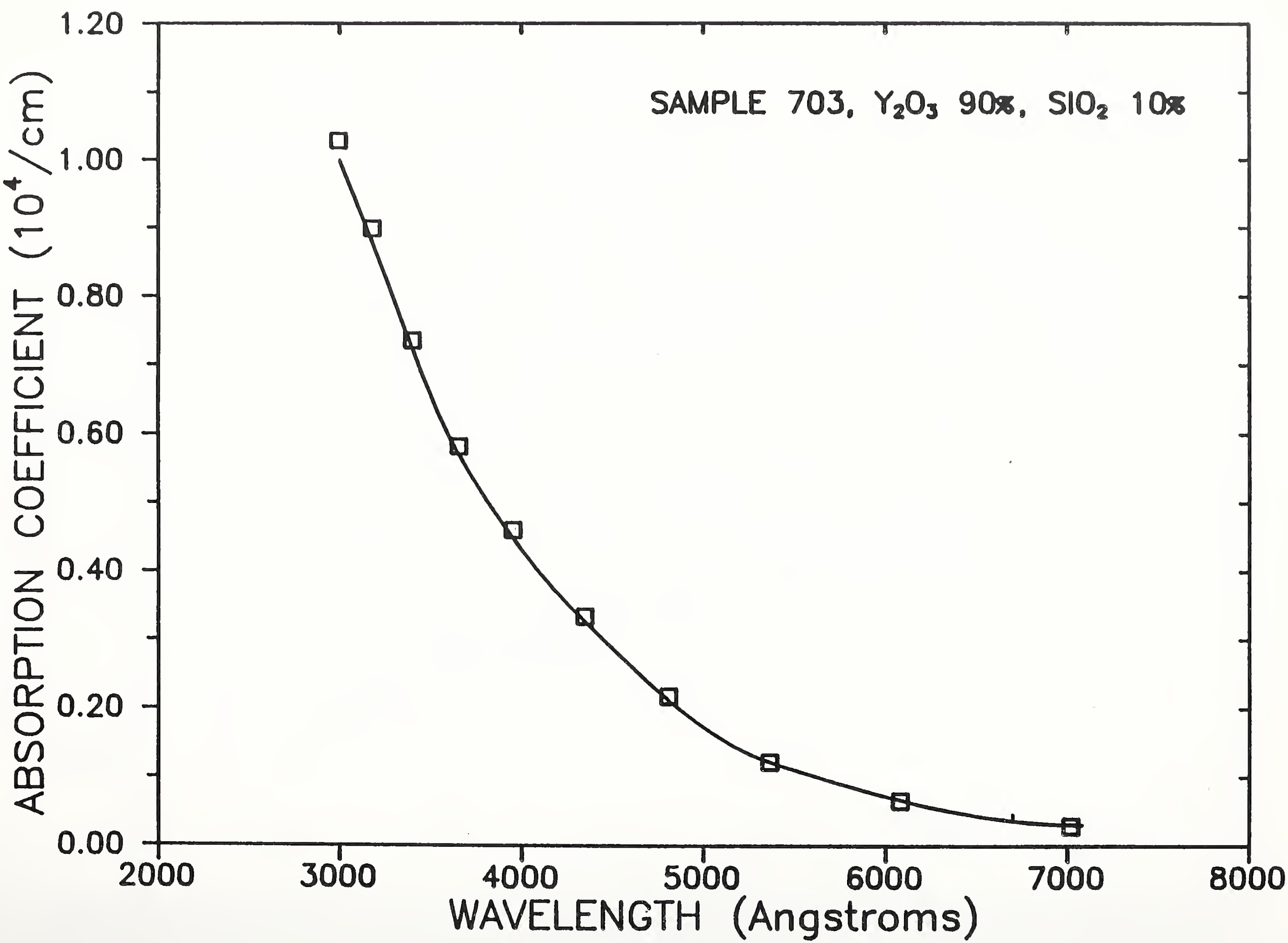

SAMPLE 703, Y2O3 90%, SIO2 10%
ABSORPTION COEFFICIENT (10⁴/cm)
1.20
1.00
0.80
0.60
0.40
0.20
0.00
2000
3000
4000
5000
6000
7000
8000
WAVELENGTH (Angstroms)

FIGURE 10

NIST-114A (REV. 3-89)	U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY	1. PUBLICATION OR REPORT NUMBER NISTIR 89-4210
		2. PERFORMING ORGANIZATION REPORT NUMBER
	BIBLIOGRAPHIC DATA SHEET	3. PUBLICATION DATE DECEMBER 1989

4. TITLE AND SUBTITLE

An Algorithm and Computer Program for the Calculation of Envelope Curves

5. AUTHOR(S)

M. McClain, A. Feldman, D. Kahaner, X. T. Ying

6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS)

U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
GAITHERSBURG, MD 20899

7. CONTRACT/GRANT NUMBER

8. TYPE OF REPORT AND PERIOD COVERED

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)

10. SUPPLEMENTARY NOTES

☐ DOCUMENT DESCRIBES A COMPUTER PROGRAM; SF-185, FIPS SOFTWARE SUMMARY, IS ATTACHED.

11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.)

A procedure has been developed to numerically calculate the envelope functions of an oscillatory curve. The method has been shown to be applicable to optical transmission data, but it is general enough to be used for many other data sets. The program is available on request.

12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)

absorption coefficient; curve fitting; envelope curves; interpolation; optical transmittance; oscillatory data; refractive index

13. AVAILABILITY

- [X] UNLIMITED
- [ ] FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS).
- [ ] ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402.
- [X] ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.

14. NUMBER OF PRINTED PAGES

25

15. PRICE

A02

ELECTRONIC FORM