ROBUST NONLINEAR LEAST SQUARES ESTIMATION USING THE CHOW-YORKE HOMOTOPY METHOD

J. P. Dunyak*, J. L. Junkins†, and L. T. Watson**
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

ROBUST NONLINEAR LEAST SQUARES ESTIMATION USING THE CHOW-YORKE HOMOTOPY METHOD

J. P. Dunyak*, J. L. Junkins†, and L. T. Watson**
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

INTRODUCTION

Homotopy and continuation methods have recently found widespread adoption as approaches for developing exceptionally robust algorithms for many multidimensional problems in nonlinear solid and fluid mechanics and optimal control [1-8]. To a lesser extent, these ideas have been applied to nonlinear estimation and system identification problems, Refs. [9,10] are recent examples. It is important to remark that merely adopting a homotopy or a continuation approach guarantees neither efficiency nor robustness of the resulting algorithm. A marginal increase in the domain of numerical convergence in exchange for a large decrease in computational efficiency is a dubious justification for a homotopy or continuation method in comparison to, say, a Gauss-Newton method.

In the present paper, we demonstrate a recently developed homotopy algorithm [3], and find an order of magnitude increase in the domain of convergence in comparison to the embedded Gauss-Newton algorithm of Kirszenblat and Chetrit [10]. Since Kirszenblat and Chetrit's algorithm was originally shown to be considerably more robust than the classical

---

*Graduate Research Assistant, Department of Engineering Science and Mechanics.
†Professor of Engineering Science and Mechanics, Associate Fellow of AIAA.
**Associate Professor of Computer Science.

Gauss-Newton algorithm, we establish that the present Chow-Yorke algorithm is robust indeed.

THE NONLINEAR LEAST SQUARE PROBLEM

We consider the problem of determining the best estimate of a parameter vector $\underline{x} = (x_1, x_2 \ldots x_n)^T$ which results in a given nonlinear function

$$y = f(t, \underline{x}) \tag{1}$$

being a best fit of a given set of measured y-values

$$\{t_1, \tilde{y}_1;\ t_2, \tilde{y}_2; \ldots; t_m, \tilde{y}_m\} \quad , \quad m > n \tag{2}$$

where the m measurement times t_j are assumed perfectly known. Adopting the simple least squares penalty function, we seek the estimate of $\underline{x}$ which minimizes

$$J(\underline{x}) = \sum_{j=1}^{m} [\tilde{y}_j - f(t_j, \underline{x})]^2 \tag{3}$$

The classical Gauss-Newton algorithm [1] for obtaining successive coorections $\Delta\underline{x}(i)$ to a sequence of trial vectors $\underline{x}(i)$ is based upon taking $\underline{x} = \underline{x}(i) + \Delta\underline{x}(i)$ and linearizing $f(t_j, \underline{x})$ about $\underline{x}(i)$; upon substituting this linear approximation for $f(t_j, \underline{x})$ into Eq. (3), the resulting quadratic (in $\Delta\underline{x}$) approximation of J can be minimized with respect to $\Delta\underline{x}$ to obtain the normal equations:

$$\Delta\underline{x}(i) = (A^T A)^{-1} A^T \Delta\underline{y} \quad ; \quad i = 0, 1, 2 \ldots \tag{4}$$

where

$$[A] = \begin{bmatrix} \left.\frac{\partial f(t_1, \underline{x})}{\partial x_1}\right|_{\underline{x}(i)} & \cdots & \left.\frac{\partial f(t_1, \underline{x})}{\partial x_n}\right|_{\underline{x}(i)} \\ \vdots & & \vdots \\ \left.\frac{\partial f(t_m, \underline{x})}{\partial x_1}\right|_{\underline{x}(i)} & \cdots & \left.\frac{\partial f(t_m, \underline{x})}{\partial x_n}\right|_{\underline{x}(i)} \end{bmatrix} \tag{5}$$

is the locally evaluated Jacobian matrix and

$$\Delta\underline{y} = \{[\tilde{y}_1 - f(t_1,\underline{x}(i))] \quad , \quad \ldots[\tilde{y}_m - f(t_m,\underline{x}(i))]\}^T \tag{6}$$

is the residual vector.

THE CONTINUATION METHOD OF KIRSZENBLAT AND CHETRIT

Motivated by the desire to enlarge the domain of convergence of the Gauss-Newton algorithm [successive corrections using Eq. (4)], Kirszenblat and Chetrit [10] imbedded the Gauss-Newton algorithm into a continuation process; they introduced the one-parameter family of penalty functions

$$J(\underline{x},\lambda) = \lambda J(\underline{x}) + (1 - \lambda) \sum_{j=1}^{m} [a_j(\lambda) - f(t_j,\underline{x})]^2 \tag{7}$$

where

$$a_j(\lambda) = \lambda\tilde{y}_j + (1 - \lambda)f(t_j,\underline{x}(0)). \tag{8}$$

Careful inspection of Eqs. (7) and (8) reveals that

$$J(\underline{x},1) \equiv J(\underline{x}) \tag{9}$$

and

$$J(\underline{x},0) = 0, \text{ at } \underline{x} = \underline{x}(0). \tag{10}$$

Thus sweeping λ from zero to one defines a family of neighboring least squares problems; if a converged solution at $\lambda = 1$ can be found, then we have a minimum of the original penalty function. The algorithm of Kirszenblatt and Chetrit involved introducing a sequence of λ-values and for each λ_k linearizing the local behavior of $f(t,\underline{x})$; an algorithm identical in structure to (4) is obtained, see Ref. [10] for the details. In essence, the classical Gauss-Newton algorithm has been imbedded into a process defining a one-parameter family of neighboring least squares problems. By sweeping λ at small increments, we can

remain close to converged neighboring solutions which can hopefully be used to obtain good starting iterates. Thus one can structure the algorithm to make the linearization implicit in Eq. (4) more nearly satisfied for each of a sequence of problems; as is evident below, this usually results in an increased domain of convergence.

A number of difficulties may be encountered, however, in applications of this continuation approach. The Jacobian matrix implicit in this quadratic approximation of $J(\underline{x} + \Delta\underline{x},\lambda)$ from (7) is not always well-conditioned. Also turning points and bifurcation points can be encountered for some λ-value less than one. The numerical results presented below support the conclusion that our implementation [13] of the Chow-Yorke algorithm [12] is vastly superior vis-a-vis the resulting domain of convergence.

THE CHOW-YORKE HOMOTOPY ALGORITHM

The Chow-Yorke algorithm is a scheme for solving nonlinear systems of equations based on homotopy maps of the form

$$\underline{F}(\underline{x},\lambda,\underline{a}) = \underline{0}, \tag{11}$$

where $\underline{x}$, $\underline{F}$, $\underline{a}$ are n-vectors and $\lambda\varepsilon[0,1]$ is a scalar continuation or homotopy parameter. Under fairly weak assumptions on $\underline{F}$, for almost all (fixed) vectors $\underline{a}$, the zero set of (11) is a smooth curve γ in the $(n + 1)$-dimensional $(\underline{x},\lambda)$ space. The smooth curve γ can be further parameterized by the arc length s along the $(n + 1)$ dimensional space curve γ, thus

$$\underline{x} = \underline{x}(s) \text{ and } \lambda = \lambda(s) \quad , \quad 0 \leq \lambda \leq 1. \tag{12}$$

Eq. (11) then becomes

$$\underline{F}(\underline{x}(s),\lambda(s),\underline{a}) = \underline{0}. \tag{13}$$

Considering both $\underline{x}$ and λ to be dependent functions of s has many

theoretical and computational advantages. Whereas $\underline{x}$ is often not a single-valued function of λ, $\underline{x}$ is (with probability one) a single-valued function of s. We now develop a set of simultaneous differential equations whose numerical solution gives $\underline{x}(s)$, $\lambda(s)$. Taking the s-derivative of Eq. (13) gives the homogeneous equation

$$[D] \begin{Bmatrix} \frac{d\underline{x}}{ds} \\ \frac{d\lambda}{ds} \end{Bmatrix} = 0 \tag{14}$$

and, since ds is the differential arc length, we have

$$\{\frac{d\underline{x}}{ds}\}^T\{\frac{d\underline{x}}{ds}\} + (\frac{d\lambda}{ds})^2 = 1 \tag{15}$$

where the n x (n + 1) Jacobian matrix is

$$D(\underline{x}(s),\lambda(s)) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} & \frac{\partial F_1}{\partial \lambda} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n} & \frac{\partial F_n}{\partial \lambda} \end{bmatrix}$$

The initial conditions are

$$\underline{x}(0) = \underline{x}_0 \quad , \quad \lambda(0) = 0. \tag{16}$$

Equations (14) and (15) determine the derivative vector $(\frac{d\underline{z}}{ds}) = (\frac{d\underline{x}}{dx}, \frac{d\lambda}{ds})^T$ only implicitly, but as is shown in Ref. [3], the solution (for D of rank n) is unique if one imposes a continuity assumption (the current derivative $(\frac{d\underline{z}}{ds})$ must always make an acute angle with the previous value of this vector). In Ref. [3] a novel solution process is

given which uses a robust Householder matrix reduction method [14] to calculate $\frac{d\underline{z}}{ds} = (\frac{d\underline{x}}{ds}, \frac{d\lambda}{ds})^T$ on each step of a variable step, variable order Adams PECE differential equation solver. The Adams PECE differential equation solver is implemented in L. F. Shampine's subroutine STEP [15].

For the nonlinear least squares problem, we introduce the following homotopy:

$$\underline{F}(\underline{x}(s),\lambda(s),\underline{a}) = (1 - \lambda)(\underline{x} - \underline{a}) + \lambda \begin{Bmatrix} \frac{\partial J}{\partial x_1}\Big|_{\underline{x}(s)} \\ \vdots \\ \frac{\partial J}{\partial x_n}\Big|_{\underline{x}(s)} \end{Bmatrix} = \underline{0} \qquad (17)$$

It is evident that $\lambda = 0$ has the trivial root $\underline{x}(0) = \underline{a}$ and a solution $\underline{x}(1)$ at $\lambda = 1$ requires that the gradient of the original residual penalty function [Eq. (3)] vanish.

In the example below, we consider the domain of numerical convergence resulting from applying the three algorithms:

(1) The classical Gauss-Newton algorithm [11].

(2) The continuation algorithm of Kirszenblat and Chetrit [10].

(3) The Chow-Yorke homotopy algorithm [13] which follows the zeroes of Eq. (17) from $\lambda = 0$ to the condition $\lambda = 1$.

The example we use is the nonlinear least squares example introduced by Kirszenblat and Chetrit [10]. Other examples using the Chow-Yorke algorithm are given in [16].

NUMERICAL SOLUTIONS OF KIRSZENBLAT AND CHETRIT'S PROBLEM

Following Ref. [10], we consider the function of four parameters and time:

$$y = x_1 e^{x_2 t} \cos(x_3 t + x_4) \tag{18}$$

Data was simulated by taking

$$\underline{x} = [x_1 \quad x_2 \quad x_3 \quad x_4]^T = [1.00 \quad -0.70 \quad 2.00 \quad 0.00] \tag{19}$$

Thirty time samples were taken, starting at t = 0 and using a uniform increment $\Delta t = 0.2$ sec between successive measurements. Both perfectly calcuatled y-values and y-values corrupted by zero mean Gaussian random numbers ($\sigma = 0.02$) were used as measurements. The addition of this small random noise had negligible impact upon any of our major results; we therefore will not discuss further these noisy data solutions.

Following the pattern of Kirszenblat and Chetrit, we vary the starting iterates for the four elements of $\underline{x}$ away from their true values (which obviously minimize Eq. (3) with a minimum value of zero), to study the domain of numerical convergence of their algorithm. Since graphical presentations are difficult in a four dimensional space, we will discuss here a sub-space of starting iterates for which x_3 and x_4 are intially assigned their true-values, but starting iterates for x_1 and x_2 are varied over a broad region. After the first correction, of course, all four elements of $\underline{x}$ are typically displaced from their true values until convergence is achieved. Using the sub-space of x_1, x_2 over which convergence is achieved as a graphical means to study convergence is introduced in Ref. [10]; we follow this pattern so homogeneous convergence comparisons can be made.

Figure 1 displays a portion of the (x_1, x_2) starting iterates over which successful numerical convergence ensues using (i) the classical

Gauss-Newton algorithm, and (ii) the continuation algorithm of Kirszenblat and Chetrit. A typical Kirszenblat-Chetrit continuation path is also shown in Figure 1. The Gauss-Newton algorithm converges (in three to six iterations) below the dashed curve whereas Kirszenblat and Chetrit's algorithm converges below the solid line if ten or more continuation steps are taken. Thus Kirszenblat and Chetrit's algorithm does significantly expand the domain of convergence of the Gauss-Newton algorithm for this example. The convergence domain of Kirszenblat and Chetrit's continuation process does not increase significantly, however, even if fifty continuation steps are taken. The failure to converge is a consequence of all elements of the Jacobian matrix (which must be inverted) tending to zero for large negative values of x_2 (due to the $e^{x_2 t}$ terms). The results we computed in Figure 1 are essentially identical to those originally reported by Kirszenblat and Chetrit [10]. Figure 2 displays the domain of convergence and a typical homotopy path for the Chow-Yorke homotopy algorithm. Notice that the region of convergence is over an order of magnitude larger in the x_2 direction. Figure 3 displays $\underline{x}(s)$ along the same trajectory shown in Figure 2. Considering arc length as the independent variable avoids the numerical difficulty at turning points if λ is chosen as the independent variable; this is an inherent limitation of the Kirszenblat-Chetrit approach.

CONCLUDING REMARKS

The Chow-Yorke homotopy algorithm was found to be much more robust than the Kirszenblat-Chetrit continuation method, when applied to their example nonlinear least squares problem. The Chow-Yorke algorithm is broadly applicable for achieving very large numerical convergence

domains. However, as reported in Ref. [1], the homotopy methods as a group (and the Chow-Yorke algorithm in particular) are relatively expensive vis-a-vis computer run time compared to, for example, a successful solution using a Gauss-Newton algorithm. Of course, an expensive solution is vastly prefered to no solution at all. Robustness measures "stability with respect to initial ignorance," a high degree of such stability is most important when seeking iterative solutions of problems having both high dimension and nonlinearity.

REFERENCES

1. Kamat, M. P., Watson, L. T., and Venkayya, V. B., "A Quasi-Newton Versus a Homotopy Method for Nonlinear Structural Analysis," Computers and Structures, to appear.

2. Kamat, M. P., Watson, L. T., and Junkins, J. L., "A Robust and Efficient Hybrid Method for Finding Multiple Equilibrium Solutions," 3rd International Symposium on Numerical Methods in Engineering, Paris, Mar. 14-17, 1983.

3. Watson, L. T., "A Globally Convergent Algorithm for Computing Fixed Points of C^2 Maps," Appl. Math. Comput., Vol. 5, 1979, pp. 297-311.

4. Junkins, J. L., and Turner, J. D., "Optimal Continuous Torque Attitude Maneuvers," AIAA J. of Guidance and Control, Vol. 3, No. 3, May 1980, pp. 210-217.

5. Turner, J. D., and Junkins, J. L., "Optimal Large Angle Rotational Maneuvers for Flexible Spacecraft," AIAA J. of Guidance and Control, Vol. 3, No. 6, Nov. 1980, pp. 578-585.

6. Turner, J. D., Chun, and Junkins, J. L., "Optimal Large-Angle Maneuvers with Vibration Suppression," presented to the Large Space Structures Symposium, Williamsburg, Va., published in NASA Conference Publication #2258, May 13, 1982.

7. Wasserstrom, E., "Solving Boundary Value Problems by Imbedding," J. Assoc. Comp. Mech., Vol. 18, 1971, pp. 594-602.

8. Roberts, S. M., and Shipman, J. S., Two-Point Boundary-Value Problems: Shooting Methods, American Elsevier, New York, 1972, p. 153.

9. Wasserstrom, E., "Numerical Solutions by the Continuation Method," SIAM Review, Vol. 15, No. 1, Jan. 1973.

10. Kirszenblat, A. and Chetrit, M., "Least Squares Nonlinear Parameter Estimation by the Iterative Continuation Methods," AIAA Journal, Vol. 12, No. 12, 1974, pp. 1751-1752.

11. Junkins, J. L., Optimal Estimation of Dynamical Systems, Sijthoff-Noordhoff, Leyden, The Netherlands, 1978, p. 33.

12. Chow, S. N., Mallet-Paret, J., and Yorke, J. A., "Finding Zeros of Maps: Homotopy Methods that are Constructive with Probability One," Math. Comp., Vol. 32, 1978, pp. 887-889.

13. Watson, L. T., and Fenner, D., "Chow-Yorke Algorithm for Fixed Points or Zeros of C^2 Maps," ACM Trans. Math. Software, Vol. 6, 1980, pp. 252-260.

14. Businger, P., and Golub, G. H., "Linear Least Squares Solutions by Householder Transformations," Numer. Math., Vol. 7, 1965, pp. 269-276.

15. Shampine, L. F., and Gordon, M. K., Computer Solution of Ordinary Differential Equations: The Initial Value Problem, W. H. Freeman, San Francisco, 1975.

16. Watson, L. T., "Engineering Applications of the Chow-Yorke Algorithm," Appl. Math. Comput., Vol. 9, 1981, pp. 111-133.

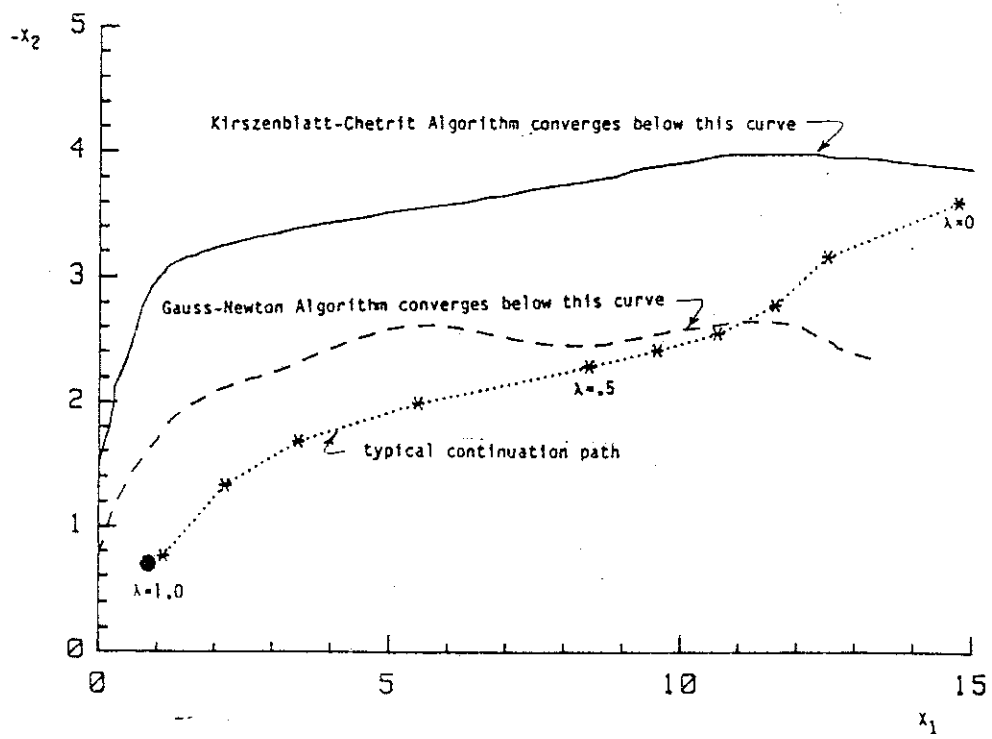

Figure 1 Gauss-Newton and Kirzenblatt-Chetrit Convergence Domains

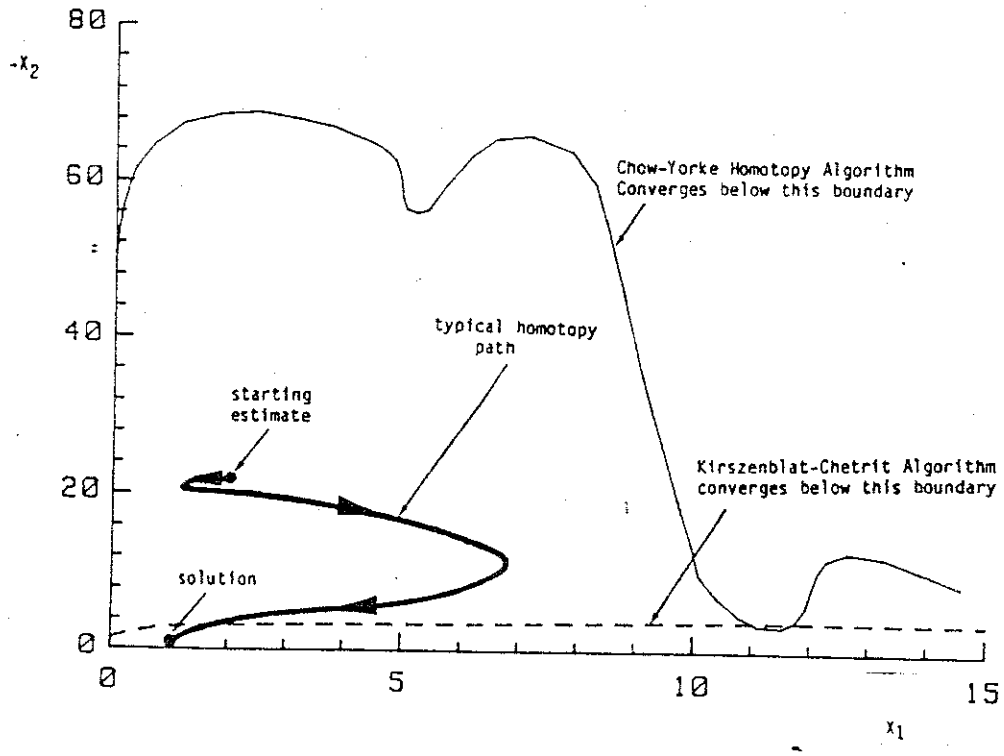

Figure 2 Chow-Yorke Homotopy Algorithm Convergence Domain

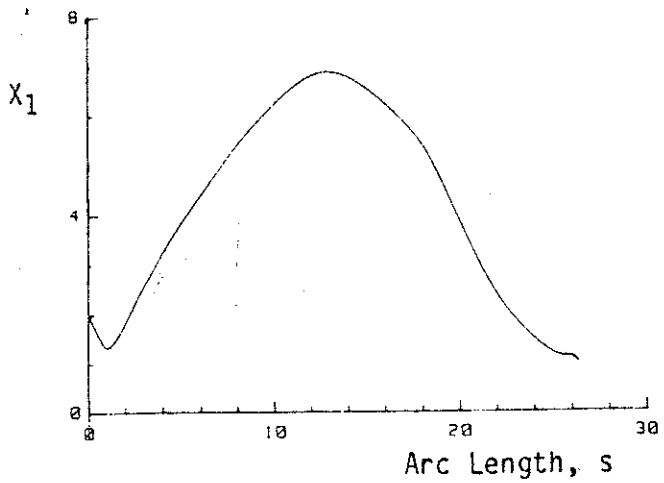

Figure 3a $X_1(s)$

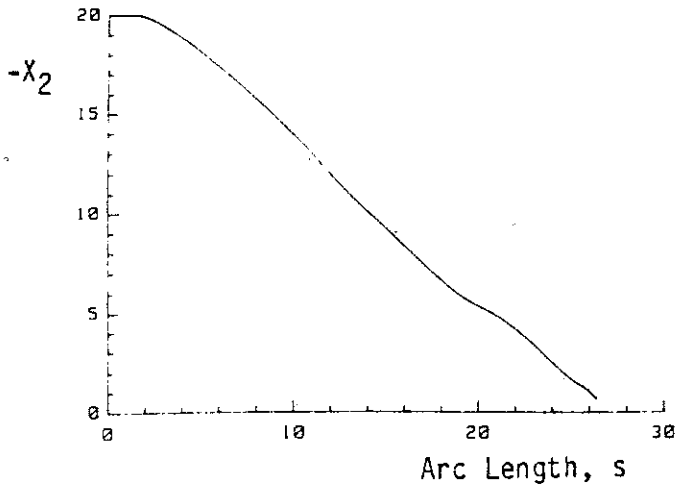

Figure 3b $X_2(s)$

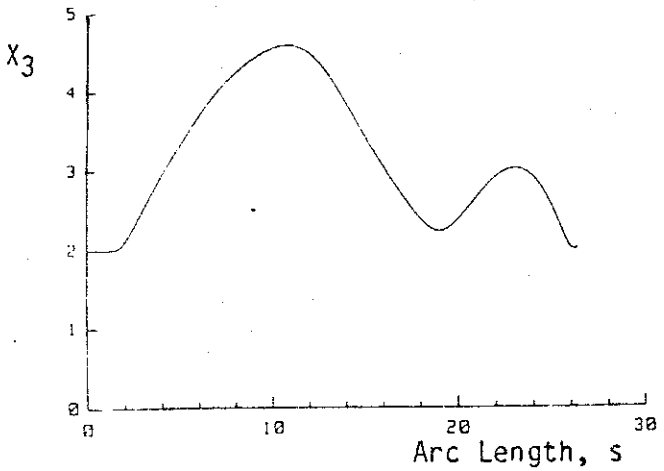

Figure 3c $X_3(s)$

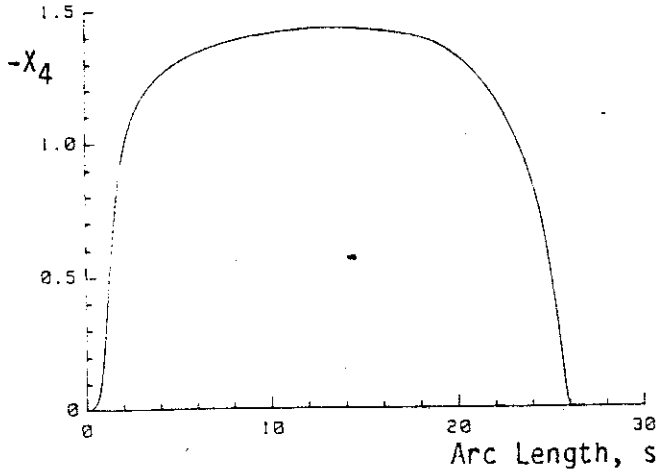

Figure 3d $X_4(s)$

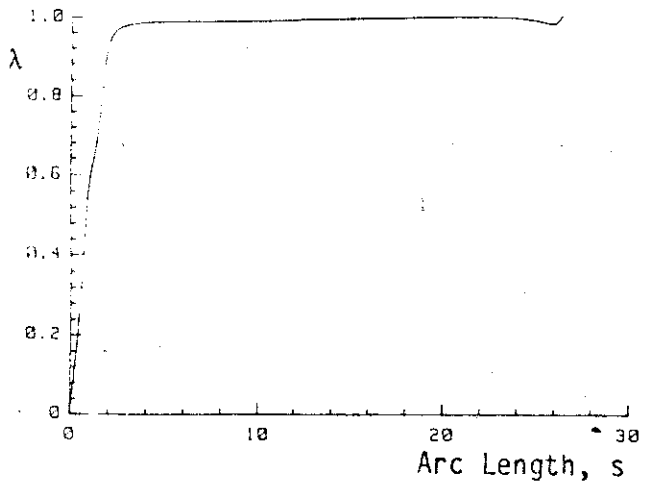

Figure 3e $\lambda(s)$

Figure 3 A Typical Chow-Yorke Homotopy Trajectory $\underline{X}(s)$, $\lambda(s)$