WADC TECHNICAL REPORT 54-250
PART 14
ASTIA DOCUMENT No. AD 97315

DYNAMIC SYSTEM STUDIES: ERROR ANALYSIS FOR DIFFERENTIAL ANALYZERS

K. S. MILLER
NEW YORK UNIVERSITY

F. J. MURRAY
COLUMBIA UNIVERSITY

SEPTEMBER 1956

AERONAUTICAL RESEARCH LABORATORY
PROJECT 7060
ADVISORY BOARD ON SIMULATION
THE UNIVERSITY OF CHICAGO
CONTRACT AF 33(038)-15068, SUPPLEMENTS 2 AND 11

WRIGHT AIR DEVELOPMENT CENTER
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

Carpenter Litho & Prtg. Co., Springfield, O.
700 - November 1956

FOREWORD

The Advisory Board on Simulation has concluded a three-year research program in air weapon system dynamics sponsored by Wright Air Development Center, with P. W. Nosker/WCRR as project engineer. This volume is one of the following 16 comprising the final report, WADC TR 54-250, entitled Dynamic System Studies:

Part No.	Subtitle	Editing Agency
1	Conclusion and Recommendations	University of Chicago
2	The Design of a Facility	" " "
3	The Mission of a Facility (Confidential)	" " "
4	Technical Staff Requirements	" " "
5	Analog Computation	Naval Ordnance Lab.
6	Operation & Maintenance Procedures for Analog Computers	University of Chicago
7	Digital Computers	" " "
8	Recorders	" " "
9	Flight Tables (Confidential)	" " "
10	Performance Requirements for Flight Tables	Mass. Inst. of Tech.
11	Load Simulators (Confidential)	Cook Research Lab.
12	Guidance Simulation (Secret)	Naval Ordnance Lab.
13	Error Studies	University of Chicago
14	Error Analysis for Differential Analyzers (written by F. J. Murray, Columbia U., and K. S. Miller, N.Y.U.)	" " "
15	Air Vehicle Characteristics (Secret)	" " "
16	Aerodynamic Studies (written by M. Z. Krzywoblocki, U. of Ill.)	" " "

The history of the project and a complete bibliography may be found in Part 1. All reports may be obtained through the project engineer.

This report represents the culmination of the assignment to determine the proper mission, equipmentation, operating procedures, and personnel for an engineering facility in the field of air weapon systems dynamics. The subdivision of the report correspond to these four basic objectives and the subsidiary work in their support, and reflect the role of simulation as a dominant technique. The functions of each part and the relations among them are indicated in the technical summary, Part 2.

The following organizations have participated directly in the program:

Organization	Contract No.	Time of Performance
University of Chicago	AF33(038)-15068 Supplements 2 and 11	1 Feb. '51-31 Aug. '54
J. B. Rea Company	AF33(038)-15068 Subcontract 2	1 Feb. '51-31 Oct. '52
Cook Research Laboratories	AF33(038)-15068 Subcontracts 3 and 9	1 Feb. '51-31 May '54
RCA Laboratories	AF33(038)-15068 Subcontract 4	1 Feb. '51-1 Mar. '53
Armour Res. Foundation of Ill. Inst. of Technology	AF33(038)-15068 Subcontract 5	1 Feb. '52-30 Nov. '52
Northwestern University, Aerial Meas. Lab.	AF33(038)-15068 Subcontract 8	17 July '52-22 Aug. '52
Mass. Inst. of Technology, Flight Control Lab.	AF33(038)-15068 Purchase Order A2086	20 Apr '54-31 Aug. '54
Mass. Inst. of Technology, Dynamic Analysis & Control Laboratory	AF339038)-15068 Purchase Order A23883	22 July '53-30 Nov. '53
Mass. Inst. of Tech., D.A.C.L.	AF33(616)-2263 Task Statement 2	1 Dec. '53-30 Sept. '54
Nat. Bur. of Standards Corona, which became	(33-038)-51-4345-E	25 Feb. '51-Sept. '53
Naval Ordnance Lab., Corona	MIPR(33-616)54-154	20 Nov. '53-31 Dec. '55

This is a record of formal participation only; the program was aided immeasurably by the splendid cooperation of all governmental, industrial and educational organizations (particularly the simulation laboratories) contacted. Although it is impractical to mention them all here, the extent of their assistance is evident throughout the reports and is hereby gratefully acknowledged. Details of these affiliations, including statements of work, may be found throughout the 21 Bimonthly Progress Reports issued by the University of Chicago during the course of the work. (All formal participation in the program is recorded above; missing supplement and subcontract numbers do not pertain to this project.)

The University of Chicago was assigned prime responsibility for integration of the program. This has been effected by a full time staff at the University, and by periodic meetings of the following advisory committee, selected by the Air Force:

Dean Walter Bartky, Chairman	University of Chicago	1 Feb.'51-31 Aug.'54
Prof. C. S. Draper	Mass. Inst. of Tech.	1 Feb.'51-28 Feb.'53
Mr. Donald McDonald	Cook Research Lab.	1 Feb.'51-31 Aug.'54
Prof. F. J. Murray	Columbia University	1 Apr.'52-31 Aug.'54
Dr. J. B. Rea	J.B.Rea Company	1 Feb.'51-28 Feb.'53
Prof. R. C. Seamans, Jr.	Mass. Inst. of Tech.	1 Sept.'53-31 Aug.'54
Mr. R. J. Shank	Hughes Aircraft Co.	1 July'51-31 Aug.'54
Dr. H. K. Skramstad	NBS-NOLC	1 Feb.'51-31 Aug.'54
Mr. A. W. Vance	RCA Laboratories	1 Feb.'51-31 Aug.'54
ex officio:		
Mr. P. W. Nosker, Project Eng.	WADC	1 Feb.'51-31 Aug.'54
Dr. B. E. Howard, Secretary	University of Chicago	1 Feb.'51-31 Aug.154

The meetings have been recorded in the Bimonthly Progress Reports previously mentioned. Except for Dr. Skramstad, who has participated through direct arrangement between NBS-NOLC and WADC, members of the advisory committee who are not connected directly with the University have participated in the program through consulting agreements with the University of Chicago. In addition, similar consulting agreements with the University have provided for the participation of:

Dr. R. R. Bennett	Hughes Aircraft Co.	1 Jan.'52-31 Jan.'54
Mr. J. P. Corbett	Libertyville, Ill (formerly with the University	11 May'54-31 Aug.'54
Dr. Thornton Page	John Hopkins Univ. (formerly with the University, and Secretary to the Board until 1 Aug.'51)	7 Aug.'51-1 Mar.'53
Prof. M. Z. Krzywoblocki	Univ. of Illinois	15 Jan.'52-31 Aug.'54
Prof. K. S. Miller	New York Univ.	2 Nov.'53-31 Aug.'54
Dr. J. Winson	Riverside, N.Y. (formerly consultant to Project Cyclone)	1 Mar.'53-30 June'54

Many others have contributed significantly to the progress of the work. Among those from other organizations in regular attendance at most of the meetings of the committee have been Mr. Charles F. West, Air Force Missile Test Center; Prof. L. L. Rauch, University of Michigan, representing Arnold Engineering Development Center; Col. A. I. Lingard, WADC; and Dr. F. W. Bubb, WADC.

Coordination of the program and administration of the prime contract at the University of Chicago has been under the charge of Dr. Walter Bartky, Dean of the Division of Physical Sciences and Director of the Institute for Air Weapons Research; Dr. B. E. Howard, Assistant to the Director; and Messrs. William R. Allen and William J. Riordan, Group Leaders. The work at the cooperating institutions has been directed by the appropriate member of the advisory committee and his assistants: Dr. H. K. Skramstad and Mr. Gerald L. Landsman at the National Bureau of Standards-Naval Ordnance Laboratory, Corona; Messrs. Donald McDonald and Jay Warshawsky at Cook Research Laboratories; Messrs. A. W. Vance, J. Lehman, and Dr. E. C. Hutter at RCA Laboratories; Dr. J. B. Rea at J. B. Rea Company; Prof. R. C. Seamans at the Flight Control Laboratory and Dr. W. W. Seifert and Mr. H. E. Blanton at the Dynamic Analysis and Control Laboratory, Mass. Inst. of Technology. V. H. Disney, S. Hori, and G. F. Warnke at Armour Research Foundation and J. C. MacAnulty and George Geolz at Northwestern University, Aerial Measurements Laboratory have directed the contributory studies at their respective organizations. More explicit credit is found in appropriate places throughout the reports; biographical sketches are in Part 1. Space does not allow full credit that is due to all the workers on the combined project, but special mention is certainly due the project engineer for his conception of the project and for his cooperation during its execution.

The original investigation upon which the present volume is based was developed by K. S. Miller and F. J. Murray, as a regular academic research project assisted by the Office of Naval Research under ONR contract 266-06. The present treatise, prepared for publication by Mr. E. R. Spangler of the University of Chicago, is fundamental to the error investigations of the basic program. An adequate error theory is a necessary adjunct to this program as an aid in (1) specifying the problems which can be handles, (2) determining allowable tolerance on the design specifications of equipment, and (3) providing methods for determining bounds on the accuracy of solutions of practical problems.

This volume contains a complete treatment of the error theory as developed by the authors to date. Introductory summaries have appeared in the Advisory Board on Simulation's Summary Progress Report for the Year Ending 1 February 1953, Volume V, Error Studies, edited by G. Weiss and R. Farrell, and in the report Project Cyclone Symposium II on Simulation and Computing Techniques, Reeves Instrument Corporation, under sponsorship of the U. S. Navy Bureau of Aeronautics, April 28 - May 2, 1952, New York, pp. 139-146. Certain mathematical aspects of the subject are treated in the paper by Miller and Murray, "A Mathematical Basis for the Error Analysis of Differential Analyzers," Journal of Mathematics and Physics, xxxii (July-Oct. 1953), 136-163.

ABSTRACT

An analog computer does not realize exactly the equations whose solution is desired. Rather it realizes a different system whose solutions are to be used as approximations to the solutions of the first system. Since, in general, the new system will be of higher order (the "λ error" effect), novel methods are developed to justify the assumptions made concerning the solutions of the two systems and to give a theoretical basis for stability analysis of machine setup. The basic theory also includes the "linearization" process and the treatment of inaccuracies and perturbations (β and α errors).

PUBLICATION REVIEW

The publication of this report does not constitute approval by the Air Force of the findings or the conclusions contained therein. It is published only for the exchange and stimulation of ideas.

FOR THE COMMANDER:

Aldro Lingard
ALDRO LINGARD
Colonel, USAF
Chief, Aeronautical Research Laboratory
Directorate of Research

TABLE OF CONTENTS

1. PRELIMINARY DISCUSSION

1.1 Purpose and Literature

The purpose of this paper is to present a mathematical basis for a general error analysis of the solution of systems of ordinary differential equations by machine methods. Specifically, we shall concern ourselves with the effect of errors on the machine solutions. We show that the study of these effects for general non-linear systems can be referred to the solution of linear systems of ordinary differential equations; but we do this without "linearizing" or simplifying the given system.

Since we permit perturbations in the solution as given by the machine, our discussion is applicable to both continuous computers and digital machines using step by step methods. However, stability discussions for such a digital process are most conveniently given in terms of difference equations rather than differential equations and are not given here.

There is an extensive literature on digital solutions dealing with truncation and rounding errors: Brock and Murray (1), Chadaja (8), Duncan (11), (12), (13), Forsythe (14), Fricke (15), Gill (17), Huskey (19), Kirby (20), Loud (27), Miller (30), Murray (31), Papoulis (33), Rademacher (34), Todd (37), Turton (39). But an intensive analysis of this work is not appropriate here. One of our major objectives is to avoid the "linearization" which appears in these; in this sense our results can be regarded as supplementing this work.

The effect of errors on solutions obtained by means of continuous computers has been studied in the case in which the given problem involves a system of linear equations with constant coefficients, notably by Raymond (35) and Macnee (28), (29).

Our discussion is based to a certain extent on well known theories for the dependence of systems of ordinary differential equations on parameters. However, it was necessary to extend this theory in order to properly consider those errors which affect the order of the system. Order variations in systems of equations have been considered from other points of view by Coddington and Levinson (9), Friedrichs and Wasow (16), Gradstein (18), and Levinson (26).

We have also listed in our bibliography certain papers in Russian of which we have considered only reviews: Bruevič (3), Byhovskiǐ (5), (6), (7), Tihonov (38), Vasil´eva (41), (42). These papers may contain material relevant to our present discussion.

Manuscript released by the author in May 1954 for publication as a WADC Technical Report.

BIBLIOGRAPHY

(1) Brock, P., and F. J. Murray. Planning and error analysis for the numerical solution of a test system of differential equations on the IBM sequence calculator, Project Cyclone, Reeves Instrument Corp., N. Y., 2 Oct. 1950.

(2) Brock, P., and F. J. Murray, "The use of exponential sums in step-by-step integration", Math. Tables and Other Aids to Comp., VI (1952), 63-78, 138-150.

(3) Bruevič, N. G., "On the accuracy of the fundamental formula of the theory of errors of a mechanism", Bull. Acad. Sci. USSR, Cl. Sci. Tech. (Izvestiya Akad. Nauk SSSR), 1944, 545-558. (Russian).

(4) Bush, V., "The differential analyzer. A new machine for solving differential equations", J. Franklin Inst., CCXII (1931), 447-488.

(5) Byhovskiĭ, M. L., "The accuracy of mechanisms controlled by differential equations", Izvestiya Akad. Nauk SSSR, Otd. Tehn. Mauk, 1947, 1455-1512. (Russian).

(6) Byhovskiĭ, M. L., "The accuracy of electric networks intended for the solution of Laplace's equation", Izvestiya Akad. Nauk SSSR. Otd. Tehn. Nauk, 1950, 489-526. (Russian).

(7) Byhovskiĭ, M. L., "The accuracy of electrical circuits for calculation", Izvestiya Akad. Nauk SSSR. Otd. Tehn. Nauk, 1948, 1239-1278. (Russian).

(8) Chadaja, F. G., "On the error in the numerical integration of ordinary differential equations by the method of finite differences", Trav. Inst. Math. Tbilissi (Trudy Tbiliss. Mat. Inst.) XI (1942), 97-108.

(9) Coddington, E. A., and N. Levinson, "A boundary value problem for a non-linear differential equation with small parameter", Proc., A.M.S., III (1952), 73-81.

(10) Collatz, L., and R. Zurmühl, "Beiträge zu den Interpolationsverfahren der numerischen Integration von Differentialgleichungen 1. und 2. Ordnung", Z. Ang. Math. Mech., XXII (1942), 42-55.

(11) Duncan, W. J., Assessment of errors in approximate solutions of differential equations, Coll. Aeronaut. Cranfield, Rep. XIII (1947), 9 pp.

(12) Duncan, W. J., "Assessment of errors in approximate solutions of differential equations", Quart. J. Mech. Appl. Math., I (1948), 470-476.

(13) Duncan, W. J., "Technique of the step-by-step integration of ordinary differential equations", Phil. Mag. (7), XXXIX (1948), 493-509.

(14) Forsythe, G. E., "Note on rounding-off errors", National Bureau of Standards, Los Angeles, Calif., 1950, 3 pp.

(15) Fricke, A., "Über die Fehlerabschätzung des Adamsschen Verfahrens zur Integration gewöhnlicher Differentialgleichungen 1. Ordung", Z. Ang. Math. Mech., XXIX (1949), 165-178.

(16) Friedrichs, K. O., and W. R. Wasow, "Singular perturbations of non-linear oscillations", Duke Math. J., XIII (1946), 367-381.

(17) Gill, S., "A process for the step-by-step integration of differential equations in an automatic digital computing machine", Proc. Cambridge Phil. Soc., XLVII (1951), 96-108.

(18) Gradšteĭn, I. S., "Linear equations with variable coefficients and small parameters in the highest derivatives", Mat. Sbornik, N. S. XXVII (1950), 47-68.

(19) Huskey, H. D., "On the precision of a certain procedure of numerical integration", J. Research, Nat. Bur. Standards, XLII (1949), 57-62.

(20) Kirby, S., The relative accuracy of quadrature formulae of the Cotes' closed type, Coll. Aeronaut. Cranfield., Rep. XVII (1948), 6 pp.

(21) Kobrinskiĭ, N. E., and L. A. Lyusternik, "Mathematical Technics", Uspehi Matem. Nauk, N.S. I (1946), 3-26.

(22) Korn, G. A., "The difference analyzer: A simple differential equation solver", Math. Tables and Other Aids to Comput., VI (1952), 1-8.

(23) Korn, G. A., "Elements of D. C. analog computers", Electronics, XXI (1948), 124-127.

(24) Korn, G. A., "Design of D. C. electronic integrators", Electronics, XXI (1948), 124-126.

(25) Lahaye, E., "Une méthode de résolution des équations différentielles", Bull. Acad. Roy. Belgique Cl. Sci. (5) XXXIV (1948), 851-862.

(26) Levinson, N., "Perturbations of discontinuous solutions of non-linear systems of differential equations", Acta Mathematica, LXXXII (1947), 71-76.

(27) Loud, W. S., "On the long-run error in the numerical solution of certain differential equations", J. Math. Phys., XXVIII (1949), 45-49.

(28) Macnee, A. B., "An electronic differential analyzer", Proc. I.R.E., XXXVII (1949), 1315-1324.

(29) Macnee, A. B., "Some limitations on the accuracy of electronic differential analyzers", Proc., I.R.E., XL (1952), 303-308.

(30) Miller, J. C. P., "Checking by differences I", Math. Tables and Other Aids to Computation, IV (1950), 3-11.

(31) Murray, F. J., "Planning and error considerations for the numerical solution of a system of differential equations on a sequence calculator", Math. Tables and Other Aids to Comput., IV (1950), 133-144.

(32) Murray, F. J., "Error analysis for mathematical machines", Trans., N. Y. Acad. Sci., XIII (1951), 168-174.

(33) Papoulis, A., "On the accumulation of errors in the numerical solution of differential equations", J. Appl. Phys., XXIII (1952), 173-176.

(34) Rademacher, H. A., "On the accumulation of errors in processes of integration on high-speed calculating machines", Proc., Symposium on large scale digital calculating machinery. The Annals of the Computation Laboratory of Harvard University, XVI (1948), 176-187.

(35) Raymond, F. H., "Sur un type général de machines mathematiques algebriques", Ann. Telecommun., V (1950), 2-20.

(36) Rice, S. O., "Mathematical analysis of random noise", Bell Sys. Tech. J., XXIII (1944), 282-332.

(37) Todd, J., "Notes on modern numerical analysis I", Math. Tables and Other Aids to Comput., IV (1950), 39-44.

(38) Tihonov, A. N., "On systems of differential equations containing parameters", Mat. Sbornik, N. S. XXVII (1950), 147-156. (Russian).

(39) Turton, F. J., "The errors in the numerical solution of differential equations", Phil. Mag. XXVIII (1939), 359-363.

(40) Turton, F. J., "Two notes on the numerical solution of differential equations", Phil. Mag. XXVIII (1939), 381-384.

(41) Vasil′eva, A. B., "On differentiation of solutions of systems of differential equations containing a small parameter", Doklady Akad. Nauk SSSR, N. S. LXXV (1950), 483-486. (Russian).

(42) Vasil′eva, A. B., "On the differentiation of solutions of differential equations containing a small parameter", Doklady Akad. Nauk SSSR, N. S. LXI (1948), 597-599. (Russian).

1.2 Objectives

A mathematical theory of errors should provide a framework within which errors can be studied and their effects evaluated. The study of errors of individual components should be oriented to such a framework so that one can estimate their effects on the solution. Error studies for components are highly desirable, but their value depends upon a knowledge of how the solution will be affected by them.

The effect of component or other individual errors on the solution depends on the problem considered. One can readily verify that certain problems are much more sensitive to errors than others. Consequently, it is very desirable to understand and specify this sensitivity. This can only be done by methods having the generality of the present paper. Few conclusions can be drawn from specific examples, even when a reference solution of undoubted correctness is available.

Another and very important reason for such an error study is that it gives an insight into the mathematical or theoretical structure of the system of differential equations considered. One should appreciate the great technical advance represented by the differential analyzer. Without machine computations one is practically limited to linear systems with constant coefficients. However, it is also true that the structure of the latter is well known. To obtain the equivalent information for the general problems handled on differential analyzers, one needs additional theoretical structure, part of which is indicated by this error analysis. Computation by itself cannot furnish this structure, and many problems involving errors of statistical nature are better considered by means of the theoretical structure given here than by mass computation.

1.3 The α, β and λ Errors

We suppose that the system to be solved is given to us in the form of a system of n first order differential equations

$$F_i(\dot{x}_1, \ldots, \dot{x}_n, x_1, \ldots, x_n, t) = 0 \qquad i = 1, \ldots, n. \tag{1.1}$$

The precise conditions on the F_i will be stated in § 2.1, but we may indicate the situation assumed as follows. Equations (1.1) are to be considered on a region

R in n+1 dimensional space of the variables $x_1,\ldots,x_n,t$. We suppose that R can be divided into subregions on each of which we can solve for the $\dot{x}_i$ in terms of $x_1,\ldots,x_n$ and t:

$$\dot{x}_i = f_i(x_1,\ldots,x_n,t) \qquad i = 1,\ldots,n \tag{1.2}$$

where f_i is analytic in $x_1,\ldots,x_n$ and continuous in $x_1,\ldots,x_n,t$ for the subregion. A rule is given for continuing a solution from one subregion to another when a boundary point is reached.

A system such as Eqs. (1.1) will be realized on the machine in the form

$$G_i = 0 \qquad i = 1,\ldots,n \tag{1.3}$$

where the G_i's differ from the F_i's due to our inability to realize the original system accurately on the machine.

The variations from F_i to G_i may affect the order of the system, i.e., they may introduce higher derivatives of the x_j. Errors of this type we call λ errors. Variations which do not affect the order will be called α errors. For each type of error we introduce parameters, λ and $\alpha_1,\ldots,\alpha_N$, (independent of time) respectively, into the system of Eq. (1.3) so that if all the parameters are zero, Eq. (1.3) becomes Eq. (1.1), while if they assume certain other non-zero values we obtain Eq. (1.3) as realized on the machine. In particular if there are no λ errors Eq. (1.3) becomes

$$G_i(\dot{x}_1,\ldots,x_n,x_1,\ldots,x_n,t,\alpha_1,\ldots,\alpha_N) = 0 \qquad i = 1,\ldots,n. \tag{1.4}$$

Equations (1.1) and (1.3) are both special cases of Eq. (1.4) for certain values of the parameters.

We illustrate the introduction of these parameters by a simple example. Suppose we try to solve the system

$$\dot{x} = -x \tag{1.5}$$

Because of backlash, we actually set up our machine to realize, say

$$\dot{x} = -x + .001(\text{sign } \dot{x}). \tag{1.6}$$

We would write Eq. (1.6) as

$$\dot{x} = -x + \alpha \text{ sign } \dot{x}. \tag{1.7}$$

In addition to α and λ errors the machine solution may be affected by another type of error which we will term β errors. β errors arise in the course of a machine computation as the result of instantaneous disturbances of the solution. These are disturbances which appear in the solution but which do not appear in the differential equations of motion, Eq. (1.3), for the machine. For example, suppose we have a solution of the machine equations, but at a certain point in the solution process the variable x is arbitrarily disturbed and instantaneously changed by an amount β. After the change, the machine continues on a solution of the original system of equations, as it did before. But the effect of this instantaneous perturbation is to jar the machine from one solution to another. We may also consider as β errors the errors made in setting up the initial conditions of a given solution. Furthermore, in general, when a machine passes from one region of analyticity to another in the course of the solution, it will not follow out precisely the rule given for this change. This discrepancy can be described in terms of a β error. Still another example is given by integrating amplifier output noise. Such noise can be described in terms of a "shot" effect, that is, in terms of disturbances of the above sort in the solution which occur in time according to a certain probability distribution and whose magnitudes are governed by another probability distribution.

1.4 The α, β Error Theory

Our plan of attack on the general error theory involves first a discussion of the case in which no λ errors occur. The results obtained in this case are used to show that the case in which λ errors appear can be referred to a problem involving linear systems with constant coefficients. The latter problem can be solved. (Cf. Chapter 4, also Macnee [29]). In the remainder of this chapter, we outline the above procedure with emphasis on formulae and discuss its significance for sensitivity.

We begin by considering a situation in which we have only α and β errors, i.e., no λ errors. Suppose, then, that we have N α errors $\alpha_1, \ldots, \alpha_N$ and M β errors, $\beta_1, \ldots, \beta_M$. In this case it is possible to apply a generalization of the usual existence theorems for systems of ordinary differential equations which depend on parameters. (Cf. Chapter 2.) As a consequence one can show that the functions given by the machine

$$x_i = x_i(t, \alpha_1, \ldots, \alpha_N, \beta_1, \ldots, \beta_M) \tag{1.8}$$

depend analytically on the parameters $\alpha_1, \ldots, \alpha_N, \beta_1, \ldots, \beta_M$. (We suppose the initial conditions are fixed. Any error in realizing them will be a β error.)

Equation (1.8) reduces to the correct solution when we set the α's and β's equal to zero, and corresponds to the inaccurate machine solution for certain values of these parameters. We can use the analytic dependence of our solution in this case to express the solutions x_j as power series in the α's and β's.

$$x_j = x_j(t, 0, \ldots, 0, 0, \ldots, 0) + \sum_k \frac{\partial x_j}{\partial \alpha_k} \alpha_k + \sum_l \frac{\partial x_j}{\partial \beta_l} \beta_l$$

$$+ \frac{1}{2!} [\sum_{k_1, k_2} \frac{\partial^2 x_j}{\partial \alpha_{k_1} \partial \alpha_{k_2}} \alpha_{k_1} \alpha_{k_2} + [\sum_{k_1, l_1} \frac{\partial^2 x_j}{\partial \alpha_{k_1} \partial \beta_{l_1}} \alpha_{k_1} \beta_{l_1} \qquad (1.9)$$

$$+ [\sum_{l_1, l_2} \frac{\partial^2 x_j}{\partial \beta_{l_1} \partial \beta_{l_2}} \beta_{l_1} \beta_{l_2}] + \ldots$$

We must, of course, assume that the errors α and β are not so large that the series diverges. One might even advance the argument that if our error analysis requires that we go further than, say, the third degree terms in the above expansion, then the solution is so far off as to be of little value. However, theoretically, as long as the series converges we may use the above expression to obtain estimates for the effect of errors. Thus, we see that our error analysis can be reduced, in the case considered, to a study of the partial derivatives,

$$\frac{\partial x_j}{\partial \alpha}, \frac{\partial x_j}{\partial \beta}, \frac{\partial^2 x_j}{\partial \alpha^2}, \frac{\partial^2 x_j}{\partial \alpha \partial \beta}, \frac{\partial^2 x_j}{\partial \beta^2}, \text{ etc.} \qquad (1.10)$$

evaluated at $\alpha = 0, \beta = 0$.

Let γ stand for either an α or a β parameter. In view of the fact that $x_1(t, \ldots, \gamma, \ldots), \ldots, x_n(t, \ldots, \gamma, \ldots)$ satisfies Eq. (1.4), we may substitute these functions in Eq. (1.4) and take partial derivatives of each of Eqs. (1.4) with respect to γ and obtain

$$\sum_j \frac{\partial G_i}{\partial \dot{x}_j} \frac{\partial \dot{x}_j}{\partial \gamma} + \sum_j \frac{\partial G_i}{\partial x_j} \frac{\partial x_j}{\partial \gamma} + \frac{\partial G_i}{\partial \gamma} = 0, \qquad (1.11)$$

(if γ is a β, then $\partial G_i / \partial \gamma = 0$.)

Now let

$$y_j = \frac{\partial x_j}{\partial \gamma} . \tag{1.12}$$

Then

$$\frac{\partial}{\partial \gamma} \dot{x}_j = \frac{\partial^2}{\partial \gamma \partial t} x_j = \frac{\partial}{\partial t} y_j = \dot{y}_j, \tag{1.13}$$

and we see then that the partials $\frac{\partial x_j}{\partial \gamma} = y_j$ satisfy the linear system of differential equations

$$\sum_j \frac{\partial G_i}{\partial \dot{x}_j} \dot{y}_j + \sum_j \frac{\partial G_i}{\partial x_j} y_j + \frac{\partial G_i}{\partial \gamma} = 0. \tag{1.14}$$

Setting $\alpha = \beta = 0$ yields

$$\sum_j \frac{\partial F_i}{\partial \dot{x}_j} \dot{y}_j + \sum_j \frac{\partial F_i}{\partial x_j} y_j + \frac{\partial G_i}{\partial \gamma}\bigg|_0 = 0. \tag{1.15}$$

We can, therefore, find the y's for either an α parameter or a β parameter if we know their initial conditions. Now, one can show that the following initial conditions are appropriate. Suppose our computation begins at the time $t = 0$, and γ is an α. At this time the α errors have had no effect and the solution is still equal to its initial value. Consequently, $y_j = 0$ for all $j = 1, \ldots, n$ at $t = 0$. On the other hand, for $\gamma = \beta$, $y_i = 1$ at $t = t'$ if the perturbation β occurs in the variable x_i at $t = t'$ and $y_j = 0$ for $j \neq i$. There is a certain amount of compensation in these conditions since for $\gamma = \alpha$, the term $\frac{\partial G_i}{\partial \alpha}$ has to be considered in Eq. (1.15), but we have zero initial conditions while, on the other hand, $\frac{\partial G_i}{\partial \beta} = 0$ for a β error.

Let us return now to Eq. (1.14) and suppose we indicate by γ_1 the parameter previously considered. Let us now differentiate this with respect to another parameter γ_2. If we let $z_j = \frac{\partial x_j}{\partial \gamma_2}$ and w_j indicate $\frac{\partial^2 x_j}{\partial \gamma_1 \partial \gamma_2}$, we obtain

$$\sum_j \frac{\partial G_i}{\partial \dot{x}_j} \dot{w}_j + \sum_k \frac{\partial G_i}{\partial x_k} w_k + \sum_{j_1, j_2} \frac{\partial^2 G_i}{\partial \dot{x}_{j_1} \partial \dot{x}_{j_2}} \dot{y}_{j_1} \dot{z}_{j_2}$$

$$+ \sum_{j_1, k_1} \frac{\partial^2 G_i}{\partial \dot{x}_{j_1} \partial x_{k_1}} (\dot{y}_{j_1} z_{k_1} + \dot{z}_{j_1} y_{k_1}) + \sum_{k_1, k_2} \frac{\partial^2 G_i}{\partial x_{k_1} \partial x_{k_2}} y_{k_1} z_{k_2}$$

$$+ \sum_l \left[\frac{\partial^2 G_i}{\partial \gamma_1 \partial \dot{x}_l} \dot{z}_l + \frac{\partial^2 G_i}{\partial \gamma_1 \partial x_l} z_l \right] + \sum_l \left[\frac{\partial^2 G_i}{\partial \gamma_2 \partial x_l} \dot{y}_l + \frac{\partial^2 G_i}{\partial \gamma_2 \partial x_l} y_l \right]$$

$$+ \frac{\partial^2 G_i}{\partial \gamma_1 \partial \gamma_2} = 0. \qquad (1.16)$$

Now it should be appreciated that before we calculate the second partials we will have found all first partials by solving the linear system of Eq. (1.15). Consequently, the y's and z's in this system are to be considered as known functions of t, so we can abbreviate the expression in the form

$$\sum_j \frac{\partial G_i}{\partial \dot{x}_j} \dot{w}_j + \sum_k \frac{\partial G_i}{\partial x_k} w_k + T^i_{\gamma_1, \gamma_2} = 0. \qquad (1.17)$$

Here again, we set $\alpha = 0$ and obtain

$$\sum_j \frac{\partial F_i}{\partial \dot{x}_j} \dot{w}_j + \sum_k \frac{\partial F_i}{\partial x_k} w_k + T^i_{\gamma_1, \gamma_2} \Big|_o = 0. \qquad (1.18)$$

This, of course, is a linear system on the w's with precisely the same homogeneous part as Eqs. (1.15) on the y's. By well known theorems on linear differential systems, we see that our problem of evaluating the second partial derivatives reduces to one in which one has to find n linearly independent solutions of the system

$$\sum_j \frac{\partial F_i}{\partial \dot{x}_j} \dot{w}_j + \sum_k \frac{\partial F_i}{\partial x_k} w_k = 0 \qquad (1.19)$$

and then perform certain integrations.

1.5 Sensitivity

It is now apparent that if we were to try to evaluate the higher partial derivatives, we would come upon exactly this same linear homogeneous system of ordinary differential equations, Eq. (1.19). Consequently, the question of sensitivity to error of our solutions can be referred back to the study of this system of linear homogeneous differential equations. We will refer to Eqs. (1.19) as the sensitivity equations.

When one has obtained n linearly independent solutions of Eqs. (1.19), well known elementary procedures permit one to evaluate all the partial derivatives which appear in the expansion of Eq. (1.9). These procedures are referred to as the method of variation of parameters, and in general the partials can be expressed in terms of these n solutions explicitly by quadratures. Consequently, the growth and general characteristics of these n solutions indicate the sensitivity of the solution to various errors.

One can study the n linearly independent solutions of Eqs. (1.19) either by general theoretical methods or, in case one has a machine in which initial conditions can be entered accurately, by means of the machine itself. The latter process is based on the fact that a perturbation of the initial condition is a β error and the linear terms of the latter satisfy the homogeneous system of Eq. (1.19). For instance, suppose one has set up the given problem on the machine and has obtained a solution for a particular set of initial conditions. Now, take one of the dependent variables x_j and change its initial value, leaving the others unchanged. Suppose one can find the difference between this second solution and the first. In general, one is justified in regarding the set of n differences for the two solutions as yielding a solution of the variational Eqs. (1.19) and one can get n linearly independent solutions by successively applying the above process for each dependent variable $x_1, \dots, x_n$.

On the other hand it is highly desirable to supplement such purely computational procedures by theoretical investigations. For instance, it is possible even in the case the sensitivity equations do not have constant coefficients to specify the notion of stability and introduce a measure for the amount of damping present. (Cf. § 3.3.) In addition to such stability phenomena, another phenomenon is present in these sensitivity considerations which we will call resonance. "Resonance" for the constant coefficient case is discussed in § 5.4. However, even in the case of non-constant coefficients, the equivalent of resonance appears in the integrals obtained by the method of variation of parameters. We plan to discuss these questions in a future paper.

Thus, in the case in which there are no λ errors, we have succeeded in referring our problem to the solution of a system of linear differential equations without any unjustified linearization of the given problem.

1.6 The λ Errors

In Chapter 3, the discussion of the case in which no λ errors appear is carried further than indicated by the above discussion. Thus, when the β errors have a statistical character, one may express the linear and quadratic β terms in Eqs. (1.9) as chance variables and study their behavior as such. In regard to the α errors, one should normally use as few α parameters as is consistent with one's objectives. If the error is all of a determinate nature one might be wise to use only one α , as is done in § 3.5. On the other hand, independent chance α errors require a plurality of parameters.

We now return to a discussion of the λ errors. For the moment we ignore α and β errors. Furthermore, suppose here for simplicity that the order is raised at most one, and that a rise occurs in every equation. We then can write the disturbed system of equations in the form

$$G_i(\lambda \ddot{x}_1, \ldots, \lambda \ddot{x}_n, \dot{x}_1, \ldots, \dot{x}_n, x_1, \ldots, x_n, t) = 0. \qquad (1.20)$$

Equation (1.20) is now the equation of motion for the machine. A parameter λ is introduced as a multiplier of the second derivatives so that one can make them all vanish from the equations simultaneously. One thinks of λ as small, while the dependence on $\lambda \ddot{x}_j$ is to be of reasonable size. Let $x_1', \ldots, x_n'$ be the solution of Eq. (1.1) corresponding to the given initial conditions and let $x_1, \ldots, x_n$

be the machine solution which satisfies Eq. (1.20). Suppose we define u_j by the equations

$$x_j = x_j' + u_j \tag{1.21}$$

and substitute this into Eq. (1.20).

We then obtain

$$G_i(\lambda\ddot{x}_1'+\lambda\ddot{u}_1,\ldots,\lambda\ddot{x}_n'+\lambda\ddot{u}_n,\dot{x}_1'+\dot{u}_1,\ldots,\dot{x}_n'+\dot{u}_n,x_1'+u_1,\ldots,x_n'+u_n,t) = 0. \tag{1.22}$$

We can expand this in powers of $\lambda\ddot{u}_j$, $\dot{u}_j$ and u_j, $j = 1,\ldots,n$,

$$\begin{aligned} G_i = {} & G_i(\lambda\ddot{x}_1',\ldots,\lambda\ddot{x}_n',\dot{x}_1',\ldots,\dot{x}_n',x_1',\ldots,x_n', t) \\ & + \sum_j \frac{\partial G_i}{\partial(\lambda\ddot{x}_j)}\lambda\ddot{u}_j + \sum_j \frac{\partial G_i}{\partial\dot{x}_j}\dot{u}_j + \sum_j \frac{\partial G_i}{\partial x_j}u_j \\ & + R_i \end{aligned} \tag{1.23}$$

where R_i, of course, depends on higher powers of $\lambda\ddot{u}_j$, $\dot{u}_j$, and u_j. The partial derivatives $\frac{\partial G_i}{\partial(\lambda\ddot{x}_j)}$, $\frac{\partial G_i}{\partial\dot{x}_j}$, and $\frac{\partial G_i}{\partial x_j}$ now depend only on the variable t and the parameter λ since the other variables $\ddot{x}_i$, $\dot{x}_i$, and x_i have been replaced by the functions $\ddot{x}_i'$, $\dot{x}_i'$, and x_i' of t. These partials are supposed to be continuous in the variable t. Consequently, if we divide our region into small enough pieces we can suppose that on each piece these partials are constants. Thus, we write

$$\begin{aligned} G_i = {} & G_i\,(\lambda\ddot{x}_1',\ldots,\lambda\ddot{x}_n',\ \dot{x}_1',\ldots,\ \dot{x}_n',\ x_1',\ldots,\ x_n',\ t) \\ & + \sum_j A_{ij}\,\lambda\ddot{u}_j + \sum_j B_{ij}\dot{u}_j + \sum_j C_{ij}u_j \qquad (1.24) \\ & + \sum_j \left(\frac{\partial G_i}{\partial(\lambda\ddot{x}_j)} - A_{ij}\right)\lambda\ddot{u}_j + \sum_j \left(\frac{\partial G_i}{\partial\dot{x}_j} - B_{ij}\right)\dot{u}_j + \sum_j \left(\frac{\partial G_i}{\partial x_j} - C_{ij}\right)u_j + R_i. \end{aligned}$$

Now, this equation differs from the equation with constant coefficients on the subregions by small terms. So, we can introduce into this expression a parameter η

$$
\begin{aligned}
& G_i(\ldots,\ldots, t, \eta) \\
& = \eta G_i(\lambda \ddot{x}'_j, \ldots, \dot{x}'_j, \ldots, x'_j, \ldots, t) \\
& + \sum_j A_{ij} \lambda \ddot{u}_j + \sum_j B_{ij} \dot{u}_j + \sum_j C_{ij} u_j \\
& + \eta \ \Big[\sum_j \Big(\frac{\partial G_i}{\partial(\lambda \ddot{x}_j)} - A_{ij} \Big) \lambda \ddot{u}_j \\
& + \sum_j \Big(\frac{\partial G_i}{\partial \dot{x}_j} - B_{ij} \Big) \dot{u}_j + \sum_j \Big(\frac{\partial G_i}{\partial x_j} - C_{ij} \Big) u_j \Big] + \eta R_i .
\end{aligned}
\qquad (1.25)
$$

We now compare Eq. (1.20) with the system

$$\sum_j A_{ij} \lambda \ddot{u}_j + \sum_j B_{ij} \dot{u}_j + \sum_j C_{ij} u_j = 0 \qquad (1.26)$$

where, of course, the A_{ij}, B_{ij} and C_{ij} vary with the subdivision. We can pass from Eqs. (1.26) to our original problem Eqs. (1.20) by introducing the parameter η as in Eq. (1.25). Consequently, the analytic behavior of the solution of Eq. (1.22) must be the same as that of Eq. (1.26). This means, then, that we can use our α techniques to reduce the study of time delay errors to the linear case. [For $\eta = 0$, Eq. (1.25) reduces to Eq. (1.26) and for $\eta = 1$, Eq. (1.25) reduces to Eq. (1.23).]

We can regard Eq. (1.26) as the machine equation for the system

$$\sum_j B_{ij} \dot{u}_j + \sum_j C_{ij} u_j = 0 \qquad (1.27)$$

and pass from Eq. (1.26) to Eq. (1.20) by supposing u to depend on a parameter η as in Eq. (1.25).

The above is the basic principle upon which Chapter 4 rests. The formulae above have been simplified by assuming a uniform rise in the order. This assumption is not made in Chapter 4.

The discussion of Chapter 4 proceeds to consider the effect of λ errors on the system of Eq. (1.27). One can show that, in general, the roots of the indicial equation of Eq. (1.26), which would be a polynomial of degree 2n, can be divided into two sets $\mu_1, \dots, \mu_n$ and $\nu_1/\lambda, \dots, \nu_n/\lambda$. Both the $\mu_1, \dots, \mu_n$ and $\nu_1, \dots, \nu_n$ are, in general, analytic in λ at $\lambda = 0$. The corresponding solutions of Eq. (1.26) have exponential factors

$$e^{\mu_i t} \qquad \text{and} \qquad e^{(\nu_i/\lambda)t} .$$

The μ_i terms correspond to what one would normally consider to be the long range effect of errors. On the other hand, the ν_i/λ exponentials are not analytic in λ at $\lambda = 0$ and either destroy the solution completely when the real part of ν_i is positive or if the real part of ν_i is negative they become very small in a brief t interval. The total error u when λ is present can be expressed in two sets of terms one of which has the above mentioned μ properties, the other the above mentioned ν properties.

In § 4.8, the combined effect of α, β and λ errors are considered. Let us consider again the machine Eq. (1.20) with, however, α errors added and with β errors permitted in the solution. We may again introduce the parameter η to yield the equivalent of Eqs. (1.25). η is an α parameter. Thus we may obtain for the machine solution an expansion about the point where $\alpha = 0$, $\beta = 0$ in the form

$$x(t,\lambda,\alpha,\beta) = x(t,\lambda,0,0) + \sum_{\gamma} x_{\gamma}(t,\lambda,0,0)\gamma + \dots . \qquad (1.28)$$

Here, $x(t,\lambda,0,0)$ is the solution obtained under the assumption of no α or β errors, while the partials of x with respect to the various γ's are obtained by an argument similar to that for the α, β error, using, however, Eqs. (1.26) as sensitivity equations.

2. BASIC THEORY FOR α AND β ERRORS

2.1 The Given Mathematical Problem

The purpose of the present section is to describe precisely the types of problems to which the present error analysis is applicable. This description is given at the end of this section. Before the precise statement is given, however, it seems desirable to give an introductory discussion which is not stated in mathematically precise language but which will indicate certain necessary considerations.

A differential analyzer presents the solution of a system of differential equations,

$$F_j(\dot{x}_1, \ldots, \dot{x}_n, x_1, \ldots, x_n, t) = 0, \qquad j = 1, \ldots, n$$

where t is the independent variable, $x_1, \ldots, x_n$ are n unknown functions of t and $\dot{x}_i$ stands for the derivatives of x_i with respect to t. In addition to Eq. (2.1), initial values for the n variables $x_1, \ldots, x_n$ at t_o are given. It is always possible to express a differential equation problem in a form in which only first derivatives appear and we will suppose that this has been done.

In a number of cases, it may be necessary to expand the original differential equation system by the introduction of new variables and new equations in order to obtain a system which is of the first order and capable of being realized on a specific machine. In our discussion, we will suppose that this expansion has already taken place.

In many applications of interest one is not justified in considering equations $F_j = 0$ as given by a single analytic expression defined by a region in $n+1$ dimensional space of variables $x_1, \ldots, x_n, t$. Instead, it is frequently necessary to consider the region of interest R as broken up into smaller regions on each of which Eq. (2.1) can be solved for the $\dot{x}_i$,

$$\dot{x}_i = f_i(x_1, \ldots, x_n, t),$$

where f_i is analytic on each subregion in $x_1, \ldots, x_n$ for t fixed and is Riemann integrable in t when continuous functions $x_1(t), \ldots, x_n(t)$ are substituted for $x_1, \ldots, x_n$. The boundaries between these regions are analytic manifolds of not more than n dimensions. We are justified in including a boundary point in one of the subregions if the functions f_i satisfy the above mentioned criteria at such a boundary point.

We will suppose that the operation of the machine indicated the existence of a solution. In general we will be interested in solutions which extend through one or more of the boundary surfaces between regions of analyticity. If a point on such a solution is on the boundary between two regions and the f's associated with both regions are analytic at this boundary point, then the continuation of the solution through the boundary point is determined by the uniqueness theorem for ordinary differential equations. Furthermore, the analytic character of the dependence of the solution on its initial conditions remains after passing a boundary point in this fashion. On the other hand, we require this analytic dependence on the initial conditions in our discussion below and we shall take this as our hypothesis rather than specifications concerning the behavior of f on the boundary. For simplicity, the possibility of a trajectory passing through a multiple corner or being reflected at a boundary is rejected. We now give a precise statement of the hypotheses needed in the discussion below.

Let

$$F_j(\dot{x}, x, t) = 0, \qquad j = 1, \ldots, n \tag{2.1}$$

denote a system of n first order ordinary differential equations in which t is the independent variable; x stands for n dependent variables $x_1, \ldots, x_n$ and $\dot{x}$ stands for the n derivatives $\dot{x}_1, \ldots, \dot{x}_n$ of $x_1, \ldots, x_n$ with respect to t. A solution of this system consisting of n functions $x_1(t), \ldots, x_n(t)$ is desired which at $t = t_o$ assumes specified values $x_{1,o}, \ldots, x_{n,o}$, e.g. $x_i(t_o) = x_{i,o}$.

The Eqs. (2.1) are to be considered on a region R in the $n+1$ dimensional space of the variables $x_1, \ldots, x_n, t$. This region R is subdivided into subregions $R_1, R_2, \ldots$ which do not have interior points in common. On R_k it is possible to solve Eqs. (2.1) for $\dot{x}_1, \ldots, \dot{x}_n$ in terms of $x_1, \ldots, x_n, t$;

$$\dot{x}_i = f_i^k(x_1, \ldots, x_n, t) \tag{2.2}$$

where:

(A) If $x_1', \ldots, x_n', t'$ is a point of R_k then for t' fixed, $f_i^k(x_1, \ldots, x_n, t')$ is analytic in $x_1, \ldots, x_n$ at $x_1', \ldots, x_n'$ and:

(B) If $x_1(t), \ldots, x_n(t)$ are n continuous functions of t defined for an interval $a \leq t \leq b$ and such that for every t in this interval, the point $x_1(t), \ldots, x_n(t), t$ is in R_k, then $f_i^k(x_1(t), \ldots, x_n(t), t)$ is a Riemann integrable function of t for $a \leq t \leq b$.

Let B_{kl} denote the intersection of the boundary of R_k and R_l. We suppose that if B_{kl} is not null it can be specified by a finite number of analytic equations and inequalities. If the addition of B_{kl} to R_k does not destroy the properties (A) and (B) above, we shall consider B_{kl} to be in R_k. Similarly for R_l. Thus, R_k and R_l may have boundary points in common.

Let $x_{1,o}, \ldots, x_{n,o}, t_o$ be an interior point of a subregion which we suppose has been denoted R_1. Since f^1 satisfies conditions (A) and (B) on R_1, there is a unique solution, $x_1(t), \ldots, x_n(t)$, of Eq. (2.2) which passes through $x_{1,o}, \ldots, x_{n,o}, t_o$. Suppose that as $t \rightarrow t_1$, $t_o \leq t < t_1$, $x_j(t) \rightarrow x_{j1}$ where $x_{11}, \ldots, x_{n1}, t_1$ is a point of B_{12}, but not of B_{1j} for $j \neq 2$. A continuation rule is a rule which associates with every such solution defined for $t_o \leq t \leq t_1$ a unique solution of Eq. (2.2) defined for a t interval with lower endpoint t_1 and with $x_1(t), \ldots, x_n(t), t$ in R_2 for $t > t_1$. We suppose that such a continuation rule exists for every non-null boundary B_{jk}.

If B_{12} belongs to both R_1 and R_2 then there is only one continuation rule possible. This is a consequence of the usual uniqueness theorem.

We will say that the above mentioned solution penetrates the boundary analytically, if we can find a $\delta > 0$ and a $\delta' > 0$ such that if $x_j' = x_j(t_1 - \delta)$, then
$\delta' > 0$ there exists a neighborhood of $x_1', \ldots, x_n'$ such that:

(1) Each point $x_1, \ldots, x_n$ of this neighborhood is such that $x_1, \ldots, x_n, t_1 - \delta$ is in R_1,

(2) Each point, $x_1, \ldots, x_n$ of this neighborhood determines a solution $x_1(t), \ldots, x_n(t), t$ which can be continued by the continuation rule through B_{12} into R_2,

(3) The values of the continuation $x_1(t_1 + \delta'), \ldots, x_n(t_1 + \delta'), t_1 + \delta'$ are in R_2 for every such point and

(4) Each such $x_j(t + \delta')$ is an analytic function of $x_1, \ldots, x_n$ in the neighborhood.

Again, it may be stated that if B_{12} belongs to both R_1 and R_2, then every continued solution penetrates the boundary analytically, provided the point of penetration is not a boundary point of a region other than R_1 and R_2. If B_{12} is in R_2, the continuation $x_1(t), \ldots, x_n(t), t$, $t \geqslant t_1$ may remain in B_{12}.

2.2 The Machine Realization

The effort to obtain a solution of the system of differential equations, Eq. (2.1), with given initial conditions $x_{1,o}, \ldots, x_{n,o}, t_o$ is beset with two major difficulties. In the first place, the system of equations $F_i = 0$ cannot be accurately realized on the machine. Instead, we actually realize another system

$$G_i(\dot{x}, x, t) = 0$$

which in the present chapter is assumed to be of the same order as the original F_i equations. We suppose that the difference between the two systems can be described as follows. The G_i's are dependent on certain parameters $\alpha_1, \ldots, \alpha_N$ such that at $\alpha_j = 0$; the equations $G_i = 0$ reduce to the system $F_i = 0$.

In addition, however, the process of solution on the machine may also be subject to perturbations during the course of the solution. Normally, we expect that the solution given by the machine, $x_1(t), \ldots, x_n(t)$, consists of n continuous differentiable functions $x_j(t)$, defined for a specified t interval and satisfying $G_j = 0$ on this interval and at $t = t_o$ the given initial conditions. However, in actual practice, it may turn out that at some specific time t' the variable x_i may have a jump or "saltus" by the amount β_i' while there exist two t intervals, one with t' as an upper end point, the other with t' as the lower end point on which the functions $x_1(t), \ldots, x_n(t)$ are continuous and differentiable and satisfy $G_j = 0$. One would ordinarily describe this situation by saying that the machine jumps from one solution of $G_j = 0$ to another at $t = t'$. At a given t', a number of β_i' may be zero.

These perturbations normally arise in three ways. One source may be the error made in realizing the initial conditions which we can handle as a perturbation at $t = t_o$. Also, when the solution passes through a boundary point between two regions, the continued solution may differ somewhat from the desired one, and this also can be treated as a perturbation. In general, it is desirable to assume that this latter perturbation occurs at $t_b + \delta$, $\delta > 0$, interior to the new region, R_2. This can be done as follows. Suppose that at $t_b + \delta$, the correct continuation assumes values $x_1', \ldots, x_n'$ but the machine solution due to the perturbation at the boundary actually assumes values $x_1' + \beta_1, \ldots, x_n' + \beta_n$. The uniqueness of the solution of $G_j = 0$ in R_2 assures us that we may assume that for an interval with lower end point $t_b + \delta$, the machine solution coincides with the result of perturbing the correct continuation by an amount $\beta_1, \ldots, \beta_n$ in the corresponding variables at $t_b + \delta$. It will be convenient to assume that such equivalent perturbations may always be used to replace the inaccuracies involved in realizing the continuation conditions.

The third source of perturbations is "noise". If we assume that the noise present in the system arises from a noise generator analogous to the "shot effect," it may be described as a series of perturbations whose occurrence and magnitude are a matter of chance.

We now describe the equations G_i which govern the action of the machine.

We assume that the motion of the machine is described by n functions $x_1(t), \ldots, x_n(t)$ of the time t which satisfy a system of differential equations

$$G_j(\dot{x}, x, t, \alpha_1, \ldots, \alpha_N) = 0 \tag{2.3}$$

dependent on N parameters α_j. When every $\alpha = 0$, the system of Eq. (2.3) reduces to a system of equations $F_i = 0$, [Eq. (2.1)].

The Eqs. (2.3) are to be considered on a region R in $n+1+N$ dimensional space of the variables x, t, α. This region is to be subdivided into subregions $R_1, R_2, \ldots$ which do not have interior points in common. On R_k it is possible to solve Eqs. (2.3) for $\dot{x}_1, \ldots, \dot{x}_n$ in terms of the other variables.

$$\dot{x}_i = g_i^k(x_1, \ldots, x_n, t, \alpha_1, \ldots, \alpha_N) \tag{2.4}$$

where:

(A) If x', t', α' is a point of R_k then for t' fixed, g_i^k is analytic in $x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_N$ at $x_1', \ldots, x_n', \alpha_1', \ldots, \alpha_N'$ and

(B) If $x_1(t), \ldots, x_n(t)$ are n continuous functions of t defined for an interval $a \leq t \leq b$ and $\alpha_1, \ldots, \alpha_N$ are fixed values of α such that for every t in the given interval, $x_1(t), \ldots, x_n(t), t, \alpha_1, \ldots, \alpha_N$ is in R_k, then $g_i^k(x_1(t), \ldots, x_n(t), t, \alpha_1, \ldots, \alpha_N)$ is Riemann integrable in t for $a \leq t \leq b$.

The boundary B_{kl} between the regions R_k and R_l is subject to the same description word for word as B_{kl} in the previous section and for a fixed set of values $\alpha_1, \ldots, \alpha_N$, the definition of a continuation rule given in the previous section applies here.

Suppose that for a fixed set of values, $\alpha_1', \ldots, \alpha_N'$ of the parameters, a solution of Eq. (2.3) is given with initial conditions $x_{1,o}, \ldots, x_{n,o}, t_o$ such that $x_{1,o}, \ldots, x_{n,o}, t_o, \alpha_1', \ldots, \alpha_N'$ is in R_1. Suppose that as $t \rightarrow t_1$, $t_o \leq t < t_1$, $x_j(t) \rightarrow x_{j1}$ where $x_{11}, \ldots, x_{n1}, t_1, \alpha_1', \ldots, \alpha_N'$ is a point of B_{12} but not of B_{1j} for $j \neq 2$. For $t > t_1$ let $x_j(t)$ denote the continuation of this solution into the region R_2. We say that this solution penetrated the boundary analytically, if we can find a $\delta > 0$ and a $\delta' > 0$ such that if $x_j' = x_j(t_1 - \delta)$, then, there exists a neighborhood of $x_1', \ldots, x_n', \alpha_1', \ldots, \alpha_N'$ such that:

(1) Each point of this neighborhood $x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_N$ is such that $x_1, \ldots, x_n, t_1 - \delta, \alpha_1, \ldots, \alpha_N$ is in R_1,

(2) Each point $x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_N$ determines a solution which can be continued by the continuation rule through B_{12} into R_2,

(3) The values of the continuation at $t + \delta'$ determine a point $x_1(t + \delta'), \ldots, x_n(t + \delta'), t + \delta', \alpha_1, \ldots, \alpha_N$ which is in R_2 for every such point in the neighborhood and,

(4) Each such $x_j(t + \delta')$ is analytic in $x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_N$ in this neighborhood.

Let the initial conditions $x_{1,o}, \ldots, x_{n,o}$ at t_o and $\alpha_1, \ldots, \alpha_N$ be fixed. A set of functions $x_1(t), \ldots, x_n(t)$ will be said to describe a perturbed motion of the machine if a set of values $t_1, \ldots, t_M$ of t are given such that:

(i) For $t_i < t < t_{i+1}$, $x_1(t), \ldots, x_n(t)$ is a solution of Eqs. (2.3) except possibly for a finite number of values of t where this solution analytically penetrates a boundary and:

(ii) At each t_j we have parameters β_{ij} such that $x_i(t_j+) = x_i(t_j) + \beta_{ij}$. [The form of the last equation permits us to assume that $x_i(t_o) = x_{i,o}$, even when t_o is a point of perturbation, t_1, i.e. $x_i(t_o+) = x_{i,o} + \beta_{i1}$.]

2.3 The Nature of the Machine Solution

The actual running of the machine will permit us to infer that a solution of Eqs. (2.3) exists for certain values of the parameter α and we could proceed to build a theory based on the assumption of the existence of such a solution. However, normally one would prefer to come to a conclusion concerning the existence of a solution of Eqs. (2.1) or use as an assumption the existence of a solution of Eqs. (2.1). Also, one might want to perform an analysis prior to the running of the machine. For these reasons, then, it is convenient to base our discussion on the existence of a solution to Eqs. (2.1).

THEOREM 2.1. Let the systems of Eqs. (2.1) and (2.3) be described as above. Let $a \leq t \leq b$ determine an interval t on which a set of n functions $x_1(t), \ldots, x_n(t)$ is given which for t in this interval are such that $x_1(t), \ldots, x_n(t), 0, \ldots, 0, t$ is in the region R of the previous section. $x_1(t), \ldots, x_n(t)$ either satisfies Eqs. (2.1) at t, or for a finite set of values of t analytically penetrates a boundary B_{j1} in the sense of the preceding section. Suppose b is not a boundary penetration point. Except at points of boundary *penetration,* $x_1(t), \dots, x_n(t)$ *is interior to an* R_k. *At* $t = a, x_j(a) = x_{j,o}$ *where* $x_{j,o}$ *is a given set of initial conditions. Let a set of values* $t_1, \dots, t_M$ *of* t *be given at which perturbations* β_{ij} *are to be permitted. As explained above these perturbed* t's *are not to coincide with a point of boundary penetration.*

Then there exists a neighborhood in $n+N+nM$ *dimensional space of* $x_{1,o}, \dots, x_{n,o}, 0, \dots, 0, 0, \dots, 0,$ *such that if* $x_1', \dots, x_n', \alpha_1, \dots, \alpha_N, \beta_{11}, \dots, \beta_{nM}$ *is a point in this neighborhood we can find a perturbed motion of the machine. Furthermore,* x_j *for this perturbed motion at* b, i.e. $x_j(b)$, *are analytic functions of* $x_1', \dots, x_n', \alpha_1, \dots, \alpha_N, \beta_{11}, \dots, \beta_{nM}$ *on this neighborhood.*

We prove this theorem by the use of various subintervals $a' \leq t \leq b'$. In this discussion the perturbation points $t_1, \dots, t_M$ are to be fixed and such as occur in the subinterval constitute the perturbation points for the subinterval.

LEMMA 2.1. Let $a' \leq t \leq b'$ *and* $a'' \leq t \leq b''$ *be two subintervals of* $a \leq t \leq b$ *with* $b' = a''$ *and* b'' *not points of boundary penetration. Then if the theorem holds for these two subintervals it holds for the interval* $a' \leq t \leq b''$.

Proof. Our hypothesis yields a neighborhood in $n+N+nM$ space at $t = a'$ for which the values of x_j at b' are analytic. We also have a neighborhood of $x_1(b'), \dots, x_n(b'), 0, \dots, 0, 0, \dots, 0$ at $t = a'' = b'$ for which the values of the continuation at b'' are analytic. Now, the continuity of the values at b' as functions defined in the a' neighborhood insures that we can find a subneighborhood of the a' neighborhood such that on it $x_1(b'), \dots, x_n(b'), \alpha_1, \dots, \alpha_N, \beta_{11}, \dots, \beta_{nM}$ is in the b' neighborhood mentioned above. Since an analytic function of analytic functions is analytic, the values at b'' are analytic functions of $x_1, \dots, x_n, \alpha_1, \dots, \alpha_N, \beta_{11}, \dots, \beta_{nM}$ in this subneighborhood for $t = a'$. This yields the result for the interval $a' \leq t \leq b''$.

Lemma 1 obviously generalizes to any finite number of adjacent subintervals.

LEMMA 2.2 Let t' *be any point in the interval* $a \leq t \leq b$ *not a point of boundary penetration. We can find an* $\eta > 0$ *such that for every* a' *and* b' *with* $t' - \eta \leq a' < t < b' \leq t' + \eta$, b' *is not a point of boundary penetration and the theorem holds for the interval* $a' \leq t \leq b'$.

Proof. Since t' is not a point of boundary penetration $x_1(t'), \dots, x_n(t'), 0, \dots, 0, 0, \dots, 0, t'$ is an interior point of some R_k. Consequently, the usual existence theorem gives us an interval η such that for a' and b' as specified, a unique solution of Eq. (2.3) exists whose values at b' are analytic functions of the values at a' and of the α's in an appropriate neighborhood. Now if t' is not a perturbation point, this interval η can be chosen so small that no perturbation point appears in $t' - \eta \le t \le t' + \eta$ and our result holds. If t' is a perturbation point, t_i, we can choose η so small that t' is the only perturbation point in this interval. Our perturbed motion is, for $a' \le t \le t'$, that given by a solution of Eq. (2.3) on this interval and for which the values $x_i(t')$ clearly have the correct analytic character; while for $t' < t \le b'$, the perturbed motion is a solution of Eq. (2.3) with initial values at $t' = t_i$ of $x_j(t') + \beta_{ji}$. This also will yield the desired result.

We can now establish the theorem as follows. Around each point t^r of boundary penetration, we take a neighborhood $t^r - \delta^r = a'$ and $t^r + \delta^r = b'$ for which the theorem holds by the hypothesis of analytic penetration. Now if t^r and t^{r+1} are successive boundary penetration points, we suppose that $t^r + \delta^r \le t \le t^{r+1} - \delta^{r+1}$ is a non-null interval of positive length. For every point t' of this interval, Lemma 2.2 is applicable and thus is associated with an interval of length 2η. The Heine-Borel theorem then applies, and states that a finite number of these intervals will cover the interval from $t^r + \delta^r$ to $t^{r+1} - \delta^{r+1}$. Within these finite intervals we can choose non-overlapping adjacent intervals $a' \le t \le b'$ which stretch from $t^r + \delta^r$ to $t^{r+1} - \delta^{r+1}$ and for which the theorem holds. Lemma 2.1 then tells us that the theorem holds for $t^r + \delta^r \le t \le t^{r+1} - \delta^{r+1}$ and a final application of Lemma 2.1 will yield the theorem.

The solution postulated in the theorem is not permitted to remain on a boundary for a t interval. The hypothesis that if t is not a point of boundary penetration, then the corresponding point on the solution $x_1(t), \ldots, x_n(t), t, 0, \ldots, 0, 0, \ldots, 0$ must be an interior point of R_k is made necessary by the hypothesis of the usual existence theorem which requires a region of analyticity or at least of continuity around the initial point. This condition in turn results from the necessity in the usual proof of repeatedly substituting the Picard iterants into the various functions $g_j^k(x_1, \ldots, x_n, t, \ldots)$. Now, in practice when a point t not of boundary penetration is on a boundary which is part of an R_k region, we may usually substitute the iterants into the g_j's and still stay in the region R_k. Then, the usual argument goes through and under these circumstances, our initial solution can be permitted to run along a boundary B_{jk}.

Let us note the following corollary to the theorem of this section.

COROLLARY 2.1. *In the above theorem, if in the description of the* G's *we replace* " g_j^k *is analytic in* $x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_N$ " *by* " g_j^k *has 1st* order continuous partial derivatives in $x_1, \ldots, x_n$, $\alpha_1, \ldots, \alpha_N$" and if "analytic penetration of the boundary" is replaced by a corresponding property obtained by making a similar substitution in property (4) of the definition of analytic penetration, then the theorem holds provided that in the conclusion of the theorem we substitute for "are analytic in $x_{1,o}, \ldots, x_{n,o}$, $\alpha_1, \ldots, \alpha_N$, $\beta_{11}, \ldots, \beta_{nM}$" the expression "have l th order continuous partials in $x_{1,o}, \ldots, x_{n,o}$, $\alpha_1, \ldots, \alpha_N$, $\beta_{11}, \ldots, \beta_{nM}$."

Since the corresponding basic existence theorems hold, a corresponding proof can be given.

3. THE EFFECT OF α AND β ERRORS

3.1 The Analytic Expansion for the Error

In the case in which the x_i's are analytic in the α 's and β's, we can expand

$$x_i = \varphi_i(t, \alpha_1, \ldots, \alpha_N, \beta_{11}, \ldots, \beta_{nM}) \tag{3.1}$$

in a Taylor series about the point $(t, 0, \ldots, 0, 0, \ldots, 0)$,

$$\begin{aligned} x_i = \varphi_i(t, 0, \ldots, 0, 0, \ldots, 0) &+ \sum_k \frac{\partial \varphi_i}{\partial \alpha_k} \alpha_k + \sum_{j,r} \frac{\partial \varphi_i}{\partial \beta_{jr}} \beta_{jr} \\ &+ \frac{1}{2}\left[\sum_{k,l} \frac{\partial^2 \varphi'_i}{\partial \alpha_k \partial \alpha_l} \alpha_k \alpha_l + \sum_{r,k,j} \frac{\partial^2 \varphi_i}{\partial \alpha_k \partial \beta_{jr}} \alpha_k \beta_{jr} \right. \\ &\left. + \sum_{r,s,l,j} \frac{\partial^2 \varphi_i}{\partial \beta_{lr} \partial \beta_{js}} \beta_{lr} \beta_{js} \right] + \text{higher powers}, \end{aligned} \tag{3.2}$$

$i = 1, 2, \ldots, n.$

Since we suppose that the α and β are not too large, a few terms of this expansion (say one, two or three) should be adequate to determine the error. Hence, the next problem to be considered is the determination of the partial derivatives.

$$\frac{\partial \varphi_k}{\partial a}, \quad \frac{\partial \varphi_k}{\partial \beta}, \quad \frac{\partial^2 \varphi_k}{\partial a \partial \beta}, \quad \text{etc.}$$

Now, $\varphi_1, \ldots, \varphi_n$ satisfy the system of equations

$$G_i\left(\frac{\partial \varphi}{\partial t}, \varphi, t, a_1, \ldots, a_N\right) = 0, \quad i = 1, \ldots, n \tag{3.3}$$

except at a finite number of points of perturbation and at a finite number of points at which the boundary is penetrated. Let γ stand for either an a or a β .

Let

$$y_i = \frac{\partial \varphi_i}{\partial \gamma} .$$

Since

$$\frac{\partial}{\partial \gamma} \left(\frac{\partial \varphi_i}{\partial t}\right) = \frac{\partial}{\partial t}\left(\frac{\partial \varphi_i}{\partial \gamma}\right) = \dot{y}_i ,$$

we obtain on differentiating Eq. (3.3) with respect to γ ,

$$\sum_j \frac{\partial G_i}{\partial \dot{x}_j} \dot{y}_j + \sum_j \frac{\partial G_i}{\partial x_j} y_j + \frac{\partial G_i}{\partial \gamma} = 0. \tag{3.4}$$

Since we are interested in the values of y_i at $\alpha = 0$, this equation becomes

$$\sum_j \left(\frac{\partial F_i}{\partial \dot{x}_j} \dot{y}_j + \frac{\partial F_i}{\partial x_j} y_j\right) + \frac{\partial G_i}{\partial \gamma} = 0. \tag{3.5}$$

(If γ is a β , the last term is zero.)

This system of linear differential equations will specify $y_1, \ldots, y_n$ if we know the initial conditions for these y's. If γ is an α , then at $t = t_o$, $\varphi_i(t_o, \alpha_1, \ldots, \alpha_N, \beta_{11}, \ldots, \beta_{nM}) = x_{i,o}$ and consequently, at $t = t_o$, $y_i = \partial\varphi_i/\partial\gamma = 0$. (This also applies to higher partials of φ_i relative to the α's.) If $\gamma = \beta_{ij}$ is a perturbation of x_i at $t = t_j$, then

$$x_i(t_j+) = x_i(t_j) + \beta_{ij}$$

and for $k \neq i$

$$x_k(t_j+) = x_k(t_j).$$

Consequently, at $t = t_j$, $y_k = \delta_{ik}$ and one might add that the higher derivatives of φ_k relative to β_{ij} are zero at $t = t_j$, as well as the higher cross derivatives relative to other parameters.

Thus Eqs. (3.4) with these initial values determine the $y_i = \partial\varphi_i/\partial\gamma$ in every case. Suppose, then, that the first derivatives y_i are known. Let γ_1 and γ_2 be two parameters, not necessarily distinct. We denote the corresponding y's by y^1 and y^2 respectively,

$$y_i^1 = \frac{\partial \varphi_i}{\partial \gamma_1}, \qquad y_i^2 = \frac{\partial \varphi_i}{\partial \gamma_2}$$

and the second derivative by z_j,

$$z_j = \frac{\partial^2 \varphi_j}{\partial \gamma_1 \; \partial \gamma_2} .$$

If now we differentiate Eq. (3.4) relative to γ_2 (letting $\gamma = \gamma_1$) we obtain the equation:

$$\sum_j \frac{\partial G_i}{\partial \dot{x}_j} \dot{z}_j + \sum_j \frac{\partial G_i}{\partial x_j} z_j + T_i^{1,2} + \frac{\partial^2 G_i}{\partial \gamma_1 \quad \partial \gamma_2} = 0 \qquad (3.6)$$

where

$$T_i^{1,2} = \sum_{j,k} \frac{\partial^2 G_i}{\partial \dot{x}_j \; \partial \dot{x}_k} \dot{y}_j^{\,1} \; \dot{y}_k^{\,2} + \sum_{j,k} \frac{\partial^2 G_i}{\partial \dot{x}_j \; \partial x_k} (\dot{y}_j^{\,1} \; y_k^{\,2} + y_k^{\,1} \; \dot{y}_j^{\,2})$$

$$+ \sum_{j,k} \frac{\partial^2 G_i}{\partial x_j \; \partial x_k} \; y_j^{\,1} \; y_k^{\,2}$$

$$+ \sum_j \frac{\partial^2 G_i}{\partial \gamma_2 \; \partial \dot{x}_j} \dot{y}_j^{\,1} + \sum_j \frac{\partial^2 G_i}{\partial \gamma_2 \; \partial x_j} \; y_j^{\,1}$$

$$+ \sum_j \frac{\partial^2 G_i}{\partial \gamma_1 \; \partial \dot{x}_j} \dot{y}_j^{\,2} + \sum_j \frac{\partial^2 G_i}{\partial \gamma_1 \; \partial x_j} y_j^{\,2} .$$

If we set $\alpha_j = 0$, Eq. (3.6) becomes

$$\sum_j \left[\frac{\partial F_i}{\partial \dot{x}_j} \dot{z}_j + \frac{\partial F_i}{\partial x_j} z_j \right] + T^{1,2}_{i,o} + \left. \frac{\partial^2 G_i}{\partial \gamma_1 \;\; \partial \gamma_2} \right|_o = 0 \qquad (3.7)$$

where $T^{1,2}_{i,o}$ has the expected definition.

Equation (3.7) is a system of linear differential equations of the first order in the z_j whose homogeneous part is identical with the homogeneous part of Eq. (3.5). If we consider ourselves to have found the first partials, the problem of finding the second partials is essentially the same as solving a linear system with the same homogeneous part. Similarly, the problem of finding the third partials

$$w_k = \frac{\partial^3 \varphi_k}{\partial \gamma_1 \; \partial \gamma_2 \; \partial \gamma_3}$$

involves the solution of a system

$$\sum_j \left[\frac{\partial F_i}{\partial \dot{x}_j} \dot{w}_j + \frac{\partial F_i}{\partial x_j} w_j \right] + T^{1,2,3}_{i,o} + \left. \frac{\partial^3 G_i}{\partial \gamma_1 \; \partial \gamma_2 \; \partial \gamma_3} \right|_o = 0 \qquad (3.8)$$

whose homogeneous part is again identical with the homogeneous part of the partial derivatives of lower order. It is clear that a similar situation holds for partials of any order.

3.2 The Sensitivity Equations

Thus the evaluation of the Taylor expansion of the error in a differential analyzer reduces to the problem of solving a system of linear differential equations

$$\sum_j \frac{\partial F_i}{\partial \dot{x}_j} \dot{y}_j + \sum_j \frac{\partial F_i}{\partial x_j} y_j + Q_i = 0, \qquad i = 1, \ldots, n \qquad (3.9)$$

for various functions Q_i and various initial conditions. This applies even in the general nonlinear case and we are justified in referring to Eqs. (3.9) as the sensitivity equations. The error in each application of such a device depends on the equipment and the problem to be solved. Eqs. (3.9) with the appropriate boundary conditions determine the manner in which the individual problem affects the total error. The present section is concerned with the solution of Eqs. (3.9).

The determinant of the coefficients of the $\dot{y}_j$ in Eq. (3.9) is the Jacobian for the system $F_i = 0$, and the fact that we can solve the F_i system of equations for the $\dot{x}_j$ uniquely and analytically means that the Jacobian J is not zero. Let us solve then for the $\dot{y}_j$. Let K_j^k denote the determinant obtained from J by replacing the column of elements $\partial F_i / \partial \dot{x}_j$ by the column $\partial F_i / \partial x_k$. Let J_j^i denote the algebraic complement of $\partial F_i / \partial \dot{x}_j$ in J.

Then Eqs. (3.9) are equivalent to

$$J\dot{y}_j + \sum_k K_j^{\,k} y_k + \sum_i J_j^{\,i} Q_i = 0, \qquad j = 1, \dots, n. \tag{3.10}$$

We now make the further assumption that for the given solution of the system $F_i = 0$, the first partials of F_i relative to $\dot{x}_j$ are continuous in t and J is bounded away from zero when the solution is substituted.

The existence theory for linear differential equations, which is not confined to the small, then shows that for each l, $l = 1, \dots, n$, there exists a vector solution

$$Y^l = \left\{ Y_1^{\,l}, Y_2^{\,l}, \dots, Y_n^{\,l} \right\} \qquad \text{such that}$$

$$Y_j^{\,l}(t_o) = \delta_j^{\,l}$$

and

$$J\dot{Y}_j^{\,l} + \sum_k K_j^{\,k} Y_k^{\,l} = 0. \tag{3.11}$$

If the dependence of the $Y_j^{\,l}$ on t_o is to be indicated explicitly we shall write $Y_j^{\,l} = Y_j^{\,l}(t, t_o)$ and we can, of course, find solutions $Y_j^{\,l}(t, \tau)$ of Eq. (3.11) with the property that $Y_j^{\,l}(t, \tau) = \delta_j^{\,l}$ for any value of τ in the t interval. The theory also tells us that the $Y_j^{\,l}(t, \tau)$ are linear combinations of the $Y_j^{\,k}(t, t_o)$ with constant coefficients, that is,

$$Y_j^{\,l}(t, \tau) = \sum_k A_k^{\,l}\, Y_j^{\,k}(t, t_o). \tag{3.12}$$

This relationship can be written in matrix form as

$$Y(t, \tau) = A(\tau, t_o)\; Y(t, t_o) \tag{3.13}$$

where $A(\tau, t_0) = \| A_k^{\,l}(\tau, t_o) \|$.

The initial conditions for which we wish to solve Eq. (3.10) are, for some $t = \tau$ in the given interval, either in the form

(1) $y_j = \delta_j^{\,k}$

or

(2) $y_j = 0.$

The Condition (1) applies to $\partial\varphi_j/\partial\beta$ and for these, Q_i is zero and the $Y_j^{\,l}(t,\tau)$ are the desired solutions. For Condition (2) which would apply, for example, to $\partial\varphi_j/\partial\alpha$ we look for a solution of Eq. (3.10) by the method of variation of parameters. We consider, then, solutions of Eq. (3.10) in the form

$$y_j = \sum_l W_l(t,\tau) \quad Y_j^{\,l}(t,t_o)$$

for $t \geqq \tau$, and $y_j = 0$ for $t < \tau$. For such a solution of Eq. (3.10), we have

$$J[\sum_l \dot{W}_l(t,\tau) \quad Y_j^{\,l}(t,t_o)] + J[\sum_l W_l(t,\tau) \; \dot{Y}_j^{\,l}(t,t_o)]$$

$$+ \sum_{k,l} K_j^{\,k} W_l(t,\tau) \; Y_k^{\,l}(t,t_o) + \sum_i J_j^{\,i} Q_i = 0$$

and Eq. (3.10) implies, by virtue of Eq. (3.11),

$$J[\sum_l \dot{W}_l(t,\tau) \; Y_j^{\,l}(t,t_o)] + \sum_i J_j^{\,i} \; Q_i = 0. \tag{3.14}$$

Now, if we substitute $\tau = t$ in the matrix Eq. (3.13), we see that $A(t,t_o)$ is the inverse matrix for $Y(t,t_o)$. If we multiply Eq. (3.14) by $A_k^{\,j}(t,t_o)$ and sum over j,

$$J[\sum_{l,j} \dot{W}_l(t,\tau) \; Y_j^{\,l}(t,t_o) \; A_k^{\,j}(t,t_o)] + \sum_{i,j} J_j^{\,i} \; Q_i A_k^{\,j} = 0$$

or

$$J\dot{W}_k(t,\tau) + \sum_{i,j} J_j^{\,i}(t) \; Q_i(t) \; A_k^{\,j}(t,t_o) = 0.$$

Thus for $t \geq \tau$,

$$y_j(t) = [\sum_l \int_\tau^t \dot{W}_l(\zeta,\tau)d\zeta] \quad Y_j^{\,l}(t,t_o)$$

$$= \sum_{l,i,k} \int_\tau^t J^{-1}(\zeta) \; J_k^{\,i}(\zeta) \; Q_i(\zeta) \; A_l^{\,k}(\zeta,t_o) Y_j^{\,l}(t,t_o) d\zeta$$

$$= - \sum_{i,k} \int_\tau^t J^{-1}(\zeta) \; J_k^{\,i}(\zeta) \; Q_i(\zeta) \; Y_j^{\,k}(t,\zeta) d\zeta. \tag{3.15}$$

3.3 A Remark on Stability

The above formulas can be supplemented in practical cases by using the stability of the system. If we multiply Eq. (3.11) by $Y_j^{\,1}$ and sum over j,

$$J \sum_j \dot{Y}_j^{\,1} Y_j^{\,1} + \sum_{j,k} Y_j^{\,1} K_j^{\,k} Y_k^{\,1} = 0. \qquad (3.16)$$

The second term is a quadratic form with which is associated a symmetric matrix

$$\left\| \frac{1}{2}(K_j^{\,k} + K_k^{\,j}) \right\|$$

whose characteristic roots are all real. If they all have the same sign as J on the given t interval we shall say that the system is unconditionally stable. Suppose that $\lambda\,(t)$ is the least characteristic root in absolute value and let ρ be the ratio λ/J, which of course, is positive. Then

$$\rho \sum_j (Y_j^{\,1})^2 = \lambda J^{-1} \sum_j (Y_j^{\,1})^2 \leq J^{-1} \sum_{j,k} Y_j^{\,1} K_j^{\,k} Y_k^{\,1}.$$

Thus if

$$\mu = \sum_j (Y_j^{\,1})^2,$$

Eq. (3.16) implies

$$\frac{1}{2}\dot{\mu} = -J^{-1} \sum_{j,k} Y_j^{\,1} K_j^{\,k} Y_k^{\,1} \leq -\rho\mu$$

and consequently,

$$\frac{d}{dt} \log \mu \leq -2\rho.$$

Thus, if μ' is the value of μ at $t = \tau$,

$$\log\left(\frac{\mu}{\mu'}\right) = \log \mu - \log \mu' \leq -2 \int_\tau^t \rho(\zeta)\,d\zeta$$

or

$$\mu \leq \mu' \exp\left[-2 \int_\tau^t \rho(\zeta)\,d\zeta\right].$$

With the initial conditions $Y_j^1(\tau , \tau) = \delta_j^1$,

$$\mu' = \mu(\tau) = \sum_j [Y_j^1(\tau , \tau)]^2 = 1, \quad \text{and}$$

$$\sum_j [Y_j^1(t, \tau)]^2 \leq \exp\left[-2 \int_\tau^t \rho(\zeta) d\zeta \right] .$$

Thus,

$$Y_j^1(t, \tau) \leqq \exp\left[- \int_\tau^t \rho(\zeta) d\zeta \right] \leqq e^{-(t-\tau)\rho *}$$

where $\rho *$ is the greatest lower bound for $\rho(\zeta)$ in the interval (τ, t). This can be used in the obvious way in the unconditionally stable case to obtain overestimates for y_j based on Eq. (3.15).

3.4 The β Errors

Let us first consider the linear terms in the perturbation error, that is,

$$e_j^1 = \sum_{u,v} \frac{\partial \varphi^j}{\partial \beta_{uv}} \beta_{uv}, \tag{3.17}$$

where β_{uv} is a perturbation on x_u occurring at the time t_v. For this our earlier results show that

$$e_j^1 = \sum_{u,v} Y_j^u(t, t_v) \beta_{uv}. \tag{3.18}$$

(Cf. § 3.2, in particular the discussion following Eq. (3.13).)

Certain of these β's are chance variables ("noise"), others are not and are generated by non-chance phenomena. At the level of generality of the present discussion, we can do nothing further with these latter except to point out that in the unconditionally stable case, the inequality obtained on the Y_j^u in §3.3 may permit a convenient overestimate. It also is possible to introduce the adjoint system for the system of Eq. (3.11) and express the $Y_j^u(t, t_v)$ in terms of the solution of the adjoint system. If this is done, however, one finds it rather difficult to utilize the stability that may be present.

If we suppose a specific method of generating the chance variable β, our procedure permits us to go further. It is reasonable to assume that for each variable x_u the expected number n_o of noise perturbations per unit interval of time is independent of the time and that when such a perturbation occurs, the amplitude distribution is also independent of the time with mean zero. Let σ_o denote the variance of this last distribution. Both n_o and σ_o may depend on u, in which case we write n_{ou} and σ_{ou}. However, the chance variables involved are independent for different values of u. In the usual noise theory (see, for example, [36]) one shows that under these circumstances one can divide the interval from t_o to t into subintervals in such a way that the possibility of more than one perturbation occurring in an interval may be neglected. In the i th interval, say, a perturbation occurred at t_i'. This will choose a time t_i' in the i th interval. If no perturbation occurred, t_i' may be chosen arbitrarily in the interval and $\beta_u(t_i') = 0$. For a given interval, the probability of an occurrence may be taken as $n_o \Delta t$ (where of course, Δt is so small that $n_o \Delta t \ll 1$). Actually, the probability of precisely k occurrences in an interval $t_o \leq t \leq t_1$ is

$$\frac{[n_o(t_1 - t_o)]^k}{k!} \, e^{-n_o(t_1 - t_o)} \quad , \tag{3.19}$$

according to certain results in noise theory; cf. Rice, loc. cit. In particular, the occurrence of a perturbation at time t_i' is independent of previous or successive occurrences. Consequently, the variance of $\beta_u(t_i')$ is readily seen to be

$$\sigma_o \sqrt{n_{ou} \Delta t} \quad . \tag{3.20}$$

(We shall assume σ_o is independent of u.) The linear noise effect is then a chance variable ${}^n e_j^{\,1o}$ (the presuperscript n refers to "noise"),

$${}^n e_j^{\,1} = \sum_{u,i} Y_j^u(t, t_i) \beta_{ui}' \tag{3.21}$$

where the $\{\beta_{ui}'\}$ is that subset of the β_{uv} which are chance variables. Since the β_{ui}' are now considered to be independent and large in number, we may apply the Central Limit Theorem and obtain the result that ${}^n e_j^{\,1}$ is a normally distributed chance variable with mean zero and variance σ_j,

$$\sigma_j^2 = \sum_u \sigma_o^2 n_{ou} \int_{t_o}^{t} [Y_j^u(t, \zeta)]^2 d\zeta \, . \tag{3.22}$$

(Again, in the unconditionally stable case, we can obtain upper bounds for this expression.)

In the case of noise, it may also be necessary to consider, in addition to the linear term whose expected value is usually zero, the second degree terms

$$n_{e_j}^{2} = \frac{1}{2} \sum_{u,v,u',v'} \frac{\partial^2 \varphi_j}{\partial\beta_{uv} \; \partial\beta_{u'v'}} \beta_{uv} \, \beta_{u'v'} \, . \qquad (3.23)$$

(We are again considering only those β's which are chance variables.) Hence, z_j,

$$z_j = \frac{\partial^2 \varphi_j}{\partial\beta_{uv} \; \partial\beta_{u'v'}} \qquad (3.24)$$

satisfies the differential equation

$$J\dot{z}_j + \sum_k K_j^{\,k} z_k + \sum_i J_j^{\,i} \, T_{i,o}^{\beta,\beta'} = 0 \qquad (3.25)$$

(cf. Eqs. (3.6) and (3.10)) where

$$\begin{aligned} T_{i,o}^{\beta,\beta'} = & \sum_{r,s} \frac{\partial^2 F_i}{\partial \dot{x}_r \; \partial \dot{x}_s} \dot{Y}_r^{\,u} \, \dot{Y}_s^{\,u} \\ & + \sum_{r,s} \frac{\partial^2 F_i}{\partial \dot{x}_r \; \partial x_s} (\dot{Y}_r^{\,u} \, Y_s^{\,u'} + \dot{Y}_r^{\,u'} \, Y_s^{\,u}) \qquad (3.26) \\ & + \sum_{r,s} \frac{\partial^2 F_i}{\partial x_r \; \partial x_s} Y_r^{\,u} \, Y_s^{\,u'} \, . \end{aligned}$$

Now, from Eq. (3.11)

$$\left. \begin{aligned} \dot{Y}_r^{\,u} &= -J^{-1} \sum_l K_r^{\,l} \; Y_l^{\,u} \\ \dot{Y}_r^{\,u'} &= -J^{-1} \sum_l K_r^{\,l} \; Y_l^{\,u'} \end{aligned} \right\} \qquad (3.27)$$

and consequently

$$\begin{aligned} T_{i,o}^{\beta,\beta'} = \sum_{l,k} Y_l^{\,u}(t,t_v)\, Y_k^{\,u'}(t,t_{v'}) \left\{ \frac{\partial^2 F_i}{\partial x_l\, \partial x_k} \right. & \\ - J^{-1} \sum_r \left[\frac{\partial^2 F_i}{\partial \dot{x}_r\, \partial x_k} K_r^{\,l} + \frac{\partial^2 F_i}{\partial \dot{x}_r\, \partial x_l} K_r^{\,k} \right] & \qquad (3.28) \\ \left. + J^{-2} \sum_{r,s} \frac{\partial^2 F_i}{\partial \dot{x}_r\, \partial \dot{x}_s} K_r^{\,l} K_s^{\,k} \right\} . & \end{aligned}$$

We shall write this as

$$T_{i,o}^{\beta,\beta'}(t) = \sum_{l,k} Y_l^{\,u}(t,t_v)\ \ Y_k^{\,u'}(t,t_{v'})\, \Gamma_i^{\,l,k}(t). \qquad (3.29)$$

This is a quadratic form or bilinear form in the $Y_l^{\,u} Y_k^{\,u'}$. Our previous formulas, Eq. (3.15), show that

$$z_j = - \sum_{i,k} \int_{\tau}^{t} J^{-1} J_k^{\,i}(\zeta) T_{i,o}^{\beta,\beta'}(\zeta)\ Y_j^{\,k}(t,\zeta)\, d\zeta \qquad (3.30)$$

where $\tau = \max(t_v, t_{v'})$.

Now ${}^{n}e_j^{\,2}$ may be considered as a quadratic form in independent normally distributed variables β_{uv} and hence as such is subject to well known statistical methods. However, the quantity of immediate practical interest is the expected value of ${}^{n}e_j^{\,2}$. Since the β_{uv} are independent with mean zero, the cross terms have expected value zero. Thus

$$\begin{aligned} E\,[{}^{n}e_j^{\,2}] &= \frac{1}{2} \sum_{u,v} \frac{\partial^2 \varphi_i}{\partial \beta^2_{uv}}\, E\,[\,\beta^2_{uv}\,] \\ &= -\frac{1}{2} \sum_{u,v} E[\,\beta^2_{uv}\,] \int_{t_v}^{t} J^{-1}(\zeta)\, J_k^{\,i}(\zeta) \sum_{l,h} Y_l^{\,u'}(\zeta,t_v) Y_h^{\,u}(\zeta,t_v) \\ &\quad \cdot Y_j^{\,k}(t,\zeta)\, \Gamma_i^{\,l,h}(\zeta)\, d\,\zeta . \qquad (3.31) \end{aligned}$$

Now $E\,[\beta_{uv}^{2}] = \sigma_o^{2}\, n_{ou}\,\Delta t$. If we pass to the limit as $\Delta t \rightarrow 0$, the summation relative to v becomes an integral and

$$E\,[{}^{n}e_j^{2}] =$$

$$- \frac{1}{2} \sum_{u,i,k,l,h} n_{ou}\, \sigma_o^{2} \int_{t_o}^{t} \int_{\tau}^{t} J^{-1}(\zeta) J_k^{i}(\zeta) Y_l^{u}(\zeta,\tau)$$

$$\cdot Y_h^{u}(\zeta,\tau) Y_j^{k}(t,\zeta)\, \Gamma_i^{l,h}(\zeta)\, d\zeta\, d\tau . \qquad (3.32)$$

This second degree noise term does in general not have expected value zero. Another important quantity that could be computed is the variance of ${}^{n}e_j^{2}$. Similar expressions can be obtained for higher order perturbation errors ${}^{n}e_j^{r}$ similar to those for ${}^{n}e_j^{1}$ and ${}^{n}e_j^{2}$.

3.5 The α Errors

In the previous section we treated the first and second order perturbation terms explicitly and indicated how higher β errors could be handled. We shall now consider the "forcing" errors or α terms. From our treatment of the α -terms it will also be clear how mixed terms such as $\partial^{2}\varphi_j/\partial\alpha\partial\beta$ can be treated. We note that the α -terms may or may not be chance variables. If some of them are, they may be treated as the β's were, that is, we can compute their variance. While this method will not be considered in treating the partials of φ with respect to the α's alone, it should be kept in mind when dealing with the mixed partials.

The first degree forcing errors we denote by ϵ_j^{1},

$$\epsilon_j^{1} = \sum_u \frac{\partial \varphi_j}{\partial \alpha_u}\, \alpha_u \qquad (3.33)$$

and we have seen that y_j $(= \partial\varphi_j/\partial\alpha)$ satisfies the differential equation

$$\sum_j \left(\frac{\partial F_i}{\partial \dot{x}_j} \dot{y}_j + \frac{\partial F_i}{\partial x_j} y_j \right) + \frac{\partial G_i}{\partial \alpha} \Bigg|_o = 0 \qquad (3.34)$$

with the boundary conditions

$$y_i(t_o) = 0.$$

The solution to this equation is (cf. Eq. (3.15))

$$y_j(t) = -\sum_{i,k} \int_{t_o}^{t} J^{-1}(\zeta) J_k^{\,i}(\zeta) \frac{\partial G_i}{\partial \alpha} Y_j^{\,k}(t,\zeta)\, d\zeta \tag{3.35}$$

where $\frac{\partial G_i}{\partial \alpha}$ is a function of ζ and hence

$$\epsilon_j^{\,1} = -\sum_{i,k,u} \int_{t_o}^{t} J^{-1}(\zeta) J_k^{\,i}(\zeta) \frac{\partial G_i}{\partial \alpha_u} \alpha_u\, Y_j^{\,k}(t,\zeta)\, d\zeta \tag{3.36}$$

(since the α_u are independent of time).

If we let

$$\delta^1 G_i = \sum_u \frac{\partial G_i}{\partial \alpha_u} \alpha_u, \tag{3.37}$$

then Eq. (3.36) becomes

$$\epsilon_j^{\,1} = -\sum_{i,k} \int_{t_o}^{t} J^{-1}(\zeta) J_k^{\,i}(\zeta)\, \delta^1 G_i\, Y_j^{\,k}(t,\zeta)\, d\zeta\ . \tag{3.38}$$

Hence we have an explicit representation of the linear forcing errors.

Consider now the second degree effects, $\partial^2 \varphi_j / \partial\alpha_u\, \partial\alpha_v$. If we let

$$z_j = \frac{\partial^2 \varphi_j}{\partial\alpha_u\, \partial\alpha_v}, \tag{3.39}$$

then the z_j satisfy the linear differential equation

$$\sum_j \left(\frac{\partial F_i}{\partial \dot{x}_j} \dot{z}_j + \frac{\partial F_i}{\partial x_j} z_j \right) + \left. \frac{\partial^2 G_i}{\partial \alpha_u\, \partial \alpha_v} \right|_o + T^{u,v}_{i,o} = 0 \tag{3.40}$$

(cf. Eq. (3.7)) whose solution is

$$z_j(t) = - \sum_{i,k} \int_{t_o}^{t} J^{-1}(\zeta) J_k^{\,i}(\zeta) \left[\frac{\partial^2 G_i}{\partial \alpha_u \; \partial \alpha_v} + T^{u,v}_{i,o} \right] Y_j^{\,k}(t, \zeta) d\zeta . \quad (3.41)$$

Now the second order forcing error is

$$\epsilon_j^{\,2} = \frac{1}{2!} \sum_{u,v} \frac{\partial^2 \varphi_j}{\partial \alpha_u \; \partial \alpha_v} \alpha_u \, \alpha_v \qquad (3.42)$$

and hence

$$\epsilon_j^{\,2} = - \frac{1}{2} \sum_{u,v,i,k} \int_{t_o}^{t} J^{-1}(\zeta) J_k^{\,i}(\zeta) \frac{\partial^2 G_i}{\partial \alpha_u \; \partial \alpha_v} \alpha_u \, \alpha_v \, Y_j^{\,k}(t, \zeta) d\zeta + S_j^{\,2} \qquad (3.43)$$

where

$$S_j^{\,2} = - \frac{1}{2} \sum_{u,v,i,k} \int_{t_o}^{t} J^{-1}(\zeta) J_k^{\,i}(\zeta) T^{u,v}_{i,o} \; \alpha_u \, \alpha_v \, Y_j^{\,k}(t, \zeta) d\zeta \; . \qquad (3.44)$$

If we let

$$\delta^2 G_i = \frac{1}{2} \sum_{u,v} \frac{\partial^2 G_i}{\partial \alpha_u \; \partial \alpha_v} \alpha_u \, \alpha_v \, , \qquad (3.45)$$

then

$$\epsilon_j^{\,2} = - \sum_{i,k} \int_{t_o}^{t} J^{-1}(\zeta) J_k^{\,i}(\zeta) \, \delta^2 G_i \; Y_j^{\,k}(t, \zeta) d \zeta \; + S^2 . \qquad (3.46)$$

The total error due to the forcing errors alone is clearly

$$\epsilon_j = \epsilon_j^{\,1} + \epsilon_j^{\,2} + \epsilon_j^{\,3} + \ldots \qquad (3.47)$$

and the total disturbance of the original equations is ΔG_i,

$$\Delta G_i = G_i(\frac{\partial \varphi}{\partial t}, \varphi, \; t, \; \alpha_1, \ldots, \alpha_N) - F_i \, (\frac{\partial \varphi}{\partial t}, \varphi, \; t) = \quad \delta^1 G_i + \delta^2 G_i + \ldots \quad (3.48)$$

Hence,

$$\epsilon_j(t) = \sum_{i,k} \int_{t_o}^{t} J^{-1}(\zeta) J_k^{\,i}(\zeta) Y_j^{\,k}(t, \zeta) \; \Delta G_i \; d\zeta \; + \sum_{n=2}^{\infty} S_j^{\,n} . \qquad (3.49)$$

Here we have the total α error. The first term in this expression corresponds to the direct effect of errors on the solution. The remaining terms correspond to iterated results of errors. While the first term depends only on the total error ΔG_i in the various equations, the other terms cannot be expressed integrally in this way. What one can do is illustrated by the following. Note that S_j^2 depends on

$$T^i = \sum_{u,v} T^{u,v}_{i,o}\, \alpha_u\, \alpha_v = \sum_{l,k} \frac{\partial^2 G_i}{\partial \dot{x}_l\, \partial \dot{x}_k} \Big(\sum_u \dot{y}_l^{\,u} \alpha_u \Big) \Big(\sum_v \dot{y}_k^{\,v} \alpha_v \Big)$$

$$+ \sum_{l,k} \frac{\partial^2 G_i}{\partial \dot{x}_l\, \partial x_k} \Big[\Big(\sum_u \dot{y}_l^{\,u} \alpha_u \Big) \Big(\sum_v y_k^{\,v} \alpha_v \Big) + \Big(\sum_u \dot{y}_l^{\,u} \alpha_u \Big) \Big(\sum_v y_k^{\,v} \alpha_v \Big) \Big]$$

$$+ \sum_{l,k} \frac{\partial^2 G_i}{\partial x_l\, \partial x_k} \Big(\sum_u y_l^{\,u} \alpha_u \Big) \Big(\sum_v y_k^{\,v} \alpha_v \Big) \tag{3.50}$$

$$+ \sum_l \Big(\sum_u \frac{\partial^2 G_i}{\partial \alpha_u\, \partial \dot{x}_l} \alpha_u \Big) \Big(\sum_v \dot{y}_l^{\,v} \alpha_v \Big) + \sum_l \Big(\sum_u \frac{\partial^2 G_i}{\partial \alpha_u\, \partial x_l} \alpha_u \Big) \Big(\sum_v y_l^{\,v} \alpha_v \Big)$$

$$+ \sum_l \Big(\sum_v \frac{\partial^2 G_i}{\partial \alpha_v\, \partial \dot{x}_l} \alpha_v \Big) \Big(\sum_u \dot{y}_l^{\,u} \alpha_u \Big) + \sum_l \Big(\sum_v \frac{\partial^2 G_i}{\partial \alpha_v\, \partial x_l} \alpha_v \Big) \Big(\sum_u y_l^{\,u} \alpha_u \Big) .$$

The partials of G are to be evaluated at $\alpha = 0$.

Now let d_α denote the operation of taking the α differential. Then

$$d_\alpha \varphi_l = \sum_u y_l^{\,u} \alpha_u, \qquad d_{\dot{\alpha}} \varphi_l = \sum_u \dot{y}_l^{\,u} \alpha_u \tag{3.51}$$

and

$$T^i = \sum_{l,k} \frac{\partial^2 G_i}{\partial \dot{x}_l\, \partial \dot{x}_k} d_{\dot{\alpha}} \varphi_l\, d_{\dot{\alpha}} \varphi_k + 2 \sum_{l,k} \frac{\partial^2 G_i}{\partial \dot{x}_l\, \partial x_k} d_{\dot{\alpha}} \varphi_l\, d_\alpha \varphi_k$$

$$+ \sum_{l,k} \frac{\partial^2 G_i}{\partial x_l\, \partial x_k} d_\alpha \varphi_l\, d_\alpha \varphi_k + 2 \sum_l \Bigg[\frac{\partial (\delta^1 G_i)}{\partial \dot{x}_l} d_\alpha \varphi_l \tag{3.52}$$

$$+ \frac{\partial (\delta^1 G_i)}{\partial x_l} d_{\dot{\alpha}} \varphi_l \Bigg] .$$

This is as far as we can go in expressing T^i in "integral" form.

3.6 Numerical Integration

The β discussion (§ 3.4) yields a very natural method of treating the growth of error in numerical integration. It is customary to utilize a difference equation approach both for stability considerations and for a study of the growth of error. Now one does not have in general the basic existence results for difference equations which are known for differential equations. Consequently, the use of difference equations in error growth leads either to some form of "linearization" or to gross overestimates.

The numerical integration of a system of differential equations

$$\dot{y}_i = f_i(y_1, \ldots, y_n, t) \qquad i = 1, \ldots, n \tag{3.53}$$

will yield tabulated values for the y's, $y_i^o, \ldots, y_n^o$ at intervals of h in the independent variable. These also yield tabulated values for the f_i, say f_i^o. At each value of t for which the computation is performed, we can suppose that an error $\beta_i(t)$ is made in y_i. This error β_i can be regarded as the sum of two errors β_i' and β_i''. β_i' is the truncation error, which results from the fact that there is no precise way of numerically integrating Eq. (3.53); various "numerical integration" procedures must be used. β_i'' is the "round off" error due to the fact that even these numerical procedures can not be precisely carried out since the number of places available in the registers of the machine is finite. β'' can be regarded either as a chance variable or a periodic function of t.

It should be obvious from the above discussion that the total effects of the β_i error can be evaluated by the above theory provided we know both β_i' and β_i''. For instance Eq. (3.22) will permit us to evaluate the linear effect of these errors, and Eqs. (3.23), (3.31) and (3.32) will express the quadratic effects. Of course we must express each β as the sum of two errors, β' and β''.

Thus one must evaluate β' and β''. β'' must be obtained by a specific investigation of the numerical procedure used. It is customary to consider β'' as a chance variable with known expected value and variance. It is clear from the above that in general this is adequate to determine the distribution of the corresponding effect on the solution.

It remains therefore to compute β'. There are a number of ways in which the f_i's can be regarded as functions of t defined for the intermediate values and then the integration error $\epsilon_i(h)$ made in y_i at the k th step can be defined as

$$\epsilon_i(t_{k+1}) = y_i^o(t_{k+1}) - y_i^o(t_k) - \int_{t_k}^{t_{k+1}} f_i^o dt \tag{3.54}$$

where $t_k = t_o + kh$. Let $y_i(t_{k+1})$ denote the values of the solution of Eq. (3.53), which at t_k has the values $y_i^o(t_k)$. Then

$$y_i(t_{k+1}) = y_i^o(t_k) + \int_{t_k}^{t_{k+1}} f_i dt \tag{3.55}$$

and the truncation error β' is given by

$$\beta_i'(t_{k+1}) = y_i^o(t_{k+1}) - y_i(t_{k+1}) . \tag{3.56}$$

In general one can find ϵ_i (Cf. Brock and Murray [2]) and we must obtain the relation between ϵ_i and β_i. The precise relation is Eq. (3.57) below; but in many practical instances Eq. (3.58), which is very easy to use, is adequate.

Assuming y_i^o to be defined for intermediate values of t we can use Eqs. (3.54) and (3.56) to define ϵ_i and β_i' for t between t_k and t_{k+1} by replacing t_{k+1} by t in these equations, yielding

$$\beta_i' - \epsilon_i = \int_{t_k}^{t} [f_i(y_1^o(t), \ldots, t) - f_i(y_1^o + \beta_1', \ldots, t)] \, dt . \tag{3.57}$$

The relation of Eq. (3.57) can be differentiated to yield a differential system for β'. The solution of this system by Picard's method is obtained by substituting a given approximation β' in the right hand integral and using the resulting expression for β' as an improved approximation. Normally this can be carried out to any degree of accuracy. Usually ϵ_i itself will do for a good first approximation to β_i.

However, for the accuracy needed for most purposes, the following is effective. In most methods, $\epsilon_i(t)$ is effectively linear in t (even when the interval h appears to a relatively high power). Now if we assume that this is also true of β_i and that the $A_{ij} = \frac{\partial f_i}{\partial y_j}$ are constant over the range from y_i^o to $y_i^o + \beta_i'$, then Eq. (3.57) becomes

$$\beta_i' + \frac{1}{2} \sum_j h \, A_{ij} \beta_j' = \epsilon_i . \tag{3.58}$$

(Note that if $\beta'(t) = bt$, then $\int_0^h \beta_j dt = \frac{1}{2} \beta_j h$.) Equation (3.58) can be applied directly; give ϵ_i and one obtains β_i' to the first order in ϵ_i by this formula. While one can obtain more precise formulas than Eq. (3.58), the accuracy obtained is seldom worth the effort. On the other hand, the improvement obtained by using the β_i' of Eq. (3.58) instead of ϵ_i in error analysis is frequently significant.

4. THE λ ERRORS

4.1 Introduction

The theory of the previous chapters is not applicable to the case in which the G's are of higher order than the F's. Time delays and imperfections in the response of integrators can easily result in a system of equations $G = 0$ for the device which are of higher order than the given system.

The purpose of the present chapter is to introduce a method of treating the general situation. One can by straightforward computational methods solve the problem in the linear case with constant coefficients. It is also possible by means of the theory established above for the α and β errors to reduce the study of the general nonlinear problem to a linear problem with coefficients constant on intervals.

The detailed exposition of this development begins in the next section. However, in the remainder of this section, we give certain examples which indicate the possibilities that arise when the G's are of higher order than the F's.

An error parameter λ corresponding to an error which affects the order of the system realized will not appear analytically in the machine solution. Thus the methods of the previous chapters cannot be used directly to obtain the answer.

We give an example to show that a λ or order-raising error will not appear analytically in the machine solution.

Suppose we wish to solve the simple equation

$$\dot{x} = -x.$$

Now, due to time lags, the best that we can realize on the machine is a system

$$\lambda\ddot{x} + \dot{x} = -x.$$

The general solution of this is, of course,

$$x(t) = a \exp\left[- \frac{2t}{1+\sqrt{1-4\lambda}} \right] + b \exp\left[- \frac{1+\sqrt{1-4\lambda}}{2\lambda} \; t \right].$$

We cannot depend on the initial conditions to be put in so perfectly that $b = o$ and thus, in general, we must expect the presence of the last term which is nonanalytic in λ at $\lambda = o$.

The above example illustrates how nonanalyticities arise due to time delays. In the next example we shall show that even if arbitrarily small time delays are present it is possible to obtain an unstable solution for an otherwise stable system. The example we shall use is

$$\dot{x} + 2\dot{y} = \left(\frac{3+\sqrt{17}}{2}\right) x$$

$$\dot{x} + \dot{y} = \left(\frac{3-\sqrt{17}}{2}\right) y \, .$$

We readily verify that for this case the Jacobian is unequal to zero and that -1 and -2 are the characteristic roots. Clearly one has a wide margin of stability. Suppose, due to time delays, the above (F system) is modified to

$$\dot{x} + \lambda\ddot{x} + 2\dot{y} = \left(\frac{3+\sqrt{17}}{2}\right) x$$

$$\dot{x} + \dot{y} + \lambda\ddot{y} = \left(\frac{3-\sqrt{17}}{2}\right) y$$

(the G system). One finds in this case that the four characteristic roots μ_1, μ_2, $\frac{\nu_1}{\lambda}$, $\frac{\nu_2}{\lambda}$, are:

$$\mu_1(\lambda) = -1 + r_1(\lambda)\cdot\lambda \quad ; \quad \mu_2(\lambda) = -2 + r_2(\lambda)\cdot\lambda$$

$$\frac{\nu_1(\lambda)}{\lambda} = \frac{\sqrt{2}-1}{\lambda} + r_3(\lambda) \quad ; \quad \frac{\nu_2(\lambda)}{\lambda} = -\left(\frac{\sqrt{2}+1}{\lambda}\right) + r_4(\lambda)$$

where r_1, r_2, r_3, r_4 are power series in λ . Hence we see that no matter how small $\lambda > 0$ is we have a solution of the form

$$e^{(\sqrt{2}-1)t/\lambda} \; e^{r_3(\lambda)t} \, ,$$

that is, a positive exponential. As in the first example, we cannot hope that the boundary conditions can be put in so perfectly as to eliminate this solution.

4.2 The Order of the Machine Equations

We shall now proceed to treat the general case and describe a process whereby the present problem may be analyzed using the results of Chapter 2. Our original system of equations is again of the form

$$F_i(\dot{x}, x, t) = 0, \qquad i = 1, 2, \ldots, n. \qquad (4.1)$$

The system as realized on the machine will be in the form

$$G_i(\lambda \ddot{x}_1, \ldots, \lambda \ddot{x}_n, \dot{x}_1, \ldots, \dot{x}_n, x_1, \ldots, x_n, t) = 0 \qquad (4.2)$$

$$i = 1, \ldots, n.$$

or as we prefer to write it

$$G(\lambda \ddot{x}, \dot{x}, x, t) = 0.$$

The parameter λ has been introduced for the purpose of convenience. Thus, when we set $\lambda = 0$, we obtain again the system of Eq. (4.1). It is convenient to think of λ as small and that $\frac{\partial G}{\partial(\lambda \ddot{x})}$ is of normal size. The assumption that the order only increases by one is made in order to obtain reasonably compact formulas. The arguments will generalize to the cases in which higher derivatives appear. We ignore α errors since they can be treated just as readily here as in our previous chapters and by the same methods.

We shall say that the G system of Eq. (2.3) (§ 2.2, Chapter 2) is <u>satisfactory</u> if it satisfies the conditions imposed upon the G_i in that section. The system of Eq. (4.2) above can be expanded into a first order system by introducing new equations and new variables. We shall say that Eq. (4.2) is <u>λ -satisfactory</u> if for $\lambda \neq 0$ this expanded system is satisfactory.

We must first, however, specify the extent to which Eqs. (4.2) are actually of the second degree. Suppose the Eqs. (4.2) imply exactly $n-r$ independent functional relations

$$G_j'(\dot{x}_1, \ldots, \dot{x}_n, x_1, \ldots, x_n, t) = 0, \qquad j = r+1, \ldots, n \qquad (4.3)$$

which do not involve the second derivatives. Now one can readily see that if Eq. (4.2) is such that we can explicitly solve for s of the expressions $\lambda \ddot{x}_1, \ldots, \lambda \ddot{x}_s$ in terms of $\lambda \ddot{x}_{s+1}, \ldots, \lambda \ddot{x}_n, \dot{x}_1, \ldots, \dot{x}_n, x_1, \ldots, x_n, t,$ then when we substitute these in the remaining $n-s$ equations of Eq. (4.2) either the $\lambda \ddot{x}_{s+1}, \ldots, \lambda \ddot{x}_n$ appear, and we can solve for another $\lambda \ddot{x}$ or they disappear from the remaining equations and we have $n-s$ relations of the type Eq. (4.2). Since there are exactly $n-r$ relations in Eq. (4.3), we can carry out the above elimination process for $s = 1, \ldots, r$. Suppose then that the first r equations

$$G_i(\lambda \ddot{x}_1, \ldots, \lambda \ddot{x}_n, \dot{x}_1, \ldots, \dot{x}_n, x_1, \ldots, x_n, t) = 0 \qquad (4.4)$$

$$i = 1, \ldots, r$$

can be used for solving for $\lambda \ddot{x}_1, \dots, \lambda \ddot{x}_r$ and that Eqs. (4.4) and (4.3) form a system equivalent to Eq. (4.2).

Now suppose the system of Eq. (4.3) is adequate to specify $\dot{x}_{r+1}, \dots, \dot{x}_n$ in terms of $\dot{x}_1, \dots, \dot{x}_r, x_1, \dots, x_n, t$. This means that a certain Jacobian is not zero. Thus, if we differentiate Eq. (4.3) with respect to t,

$$\sum_i \frac{\partial G'_j}{\partial \dot{x}_i} \ddot{x}_i + \sum_i \frac{\partial G'_j}{\partial x_i} \dot{x}_i + \frac{\partial G'_j}{\partial t} = 0, \quad j = r+1, \dots, n, \qquad (4.5)$$

and these equations can be used to eliminate $\lambda \ddot{x}_{r+1}, \dots, \lambda \ddot{x}_n$ from Eqs. (4.4). In order to present the following discussion with a reasonable economy we will suppose that this latter elimination has been carried out. Otherwise the requisite discussions of λ-satisfaction would be quite complex.

HYPOTHESIS 4.1. The system of Eq. (4.2) can be written

$$G'_i(\lambda \ddot{x}_1, \dots, \lambda \ddot{x}_r, \dot{x}_1, \dots, \dot{x}_n, x_1, \dots, x_n, t) = 0$$

$$i = 1, \dots, r$$

$$G'_i(\dot{x}_1, \dots, \dot{x}_n, x_1, \dots, x_n, t) = 0, \qquad i = r+1, \dots, n \qquad (4.6)$$

where the first r equations can be used to solve for $\lambda \ddot{x}_1, \dots, \lambda \ddot{x}_r$ in terms of $\dot{x}_1, \dots, \dot{x}_n, x_1, \dots, x_n, t$ and the remaining equations can be used to solve for $\dot{x}_{r+1}, \dots, \dot{x}_n$ in terms of $\dot{x}_1, \dots, \dot{x}_r, x_1, \dots, x_n$. Hypothesis 4.1 is to be understood as stating that certain Jacobians are not zero plus the existence of certain values satisfying the equations. We suppose then that the equations of motion of the machine can be written in the form of Eq. (4.6) and that at $\lambda = 0$, these become equivalent to Eq. (4.1).

In the following discussion r will be considered to be fixed for the given problem. One can generalize the discussion to a case in which r changes a number of times during the run.

Such a system is readily seen to be λ-satisfactory since we can solve for $\lambda \ddot{x}_1, \dots, \lambda \ddot{x}_r, \dot{x}_{r+1}, \dots, \dot{x}_n$ in terms of $\dot{x}_1, \dots, \dot{x}_r, x_1, \dots, x_n$. Now we introduce r new unknowns and r equations,

$$\dot{x}_j = z_j \qquad j = 1, \dots, r. \qquad (4.7)$$

The result will be an $n+r$ first order system which we can solve explicitly for the derivatives if $\lambda \neq 0$. (The $\lambda \ddot{x}_j$ become, of course, $\lambda \dot{z}_j$ after the above substitution.)

4.3 The Linearization

Now let $x_1', \ldots, x_n'$ be a given solution of Eq. (4.1) corresponding to a prescribed set of initial conditions $x_{1,o}', \ldots, x_{n,o}'$. We shall write the solution of Eq. (4.6) obtained when one tries to solve Eq. (4.1) on the machine in the form

$$x_i = x_i' + u_i. \tag{4.8}$$

Clearly the initial values of $x_1, \ldots, x_n$ and $\dot{x}_1, \ldots, \dot{x}_r$ are determined by corresponding initial values of $u_1, \ldots, u_n, \dot{u}_1, \ldots, \dot{u}_r$. We may consider Eq. (4.6) then as determining a system of equations for the u_i

$$G_i'(\lambda' \ddot{x}_1' + \lambda \ddot{u}_1, \ldots, \lambda' \ddot{x}_r' + \lambda \ddot{u}_r, \dot{x}_1' + \dot{u}_1, \ldots, \ldots, x_n' + u_n, t) = 0$$

$$i = 1, \ldots, r.$$

$$G_i'(\dot{x}_1' + \dot{u}_1, \ldots, \dot{x}_n' + \dot{u}_n, x_1' + u_1, \ldots, x_n' + u_n, t) = 0 \tag{4.9}$$

$$i = r+1, \ldots, n.$$

We distinguish between λ as a coefficient of $\ddot{u}$ and λ' as a coefficient of $\ddot{x}$. The latter λ' can be conveniently regarded as an α parameter from now on.

Normally one is justified in regarding the initial values of the u_i's as small. This means that the x_i's have been entered into the machine with approximately the correct values. However, we do not have any assurance that the values of the $\dot{u}_j$'s are small. We shall give a discussion which will clearly have a t interval of validity if $\dot{u}$ is small and this will certainly happen if the initial value of $\dot{u}$ is small. On the other hand, in general this discussion will have an adequate range of validity, but this cannot be presupposed if $\dot{u}_i$ is arbitrary. We shall make this matter precise later.

We may write Eq. (4.9) in the form

$$G_i'(\lambda'\ddot{x}', \dot{x}', x', t) + \sum_{j=1}^{r} \frac{\partial G_i'}{\partial(\lambda \ddot{x}_j)}(\lambda \ddot{u}_j) + \sum_{j=1}^{n} \frac{\partial G_i'}{\partial \dot{x}_j}\dot{u}_j + \sum_{j=1}^{n} \frac{\partial G_i'}{\partial x_j} u_j + R_i = 0$$

$$i = 1, \ldots, r \qquad (4.10)$$

$$G_i'(\dot{x}', x', t) + \sum_{j=1}^{n} \frac{\partial G_i'}{\partial \dot{x}_j}\dot{u}_j + \sum_{j=1}^{n} \frac{\partial G_i'}{\partial x_j} u_j + R_i = 0 \qquad (4.10')$$

$$i = r+1, \ldots, n \; .$$

λ' as a coefficient of $\ddot{x}'$ is to be regarded as a constant. Consequently, $\frac{\partial G_i'}{\partial(\lambda \ddot{x}_j)}$, $\frac{\partial G_i'}{\partial \dot{x}_j}$ and $\frac{\partial G_i'}{\partial x_j}$ are now functions of t independent of λ and u. R_i contains the higher powers of $\lambda \ddot{u}$, $\dot{u}$ and u.

The time interval $0 \leq t \leq t^*$ is now to be subdivided into intervals $0 = t_o < t_1 < \ldots < t_m = t^*$ and in each of these subintervals a point t_k', $t_{k-1} \leq t_k' \leq t_k$ is chosen at which we evaluate the coefficients

$$\frac{\partial G_i'}{\partial(\lambda \ddot{x}_j)} = A_{ij}, \quad i,j = 1, \ldots, r \qquad (4.11)$$

$$\frac{\partial G_i'}{\partial \dot{x}_j} = B_{ij}, \quad i,j = 1, \ldots, n \qquad (4.11')$$

$$\frac{\partial G_i'}{\partial x_j} = C_{ij}, \quad i,j = 1, \ldots, n. \qquad (4.11'')$$

These permit us to rewrite Eqs. (4.10) and 4.10') in the form

$$0 = G_i'(\lambda \ddot{x}', \dot{x}', x', t) + \sum_j A_{ij} \lambda \ddot{u}_j + \sum_j B_{ij}\dot{u}_j + \sum_j C_{ij} u_j$$

$$+ \quad \eta \Bigg[\sum_j \Big(\frac{\partial G_i'}{\partial(\lambda \ddot{x}_j)} - A_{ij} \Big) \lambda \ddot{u}_j + \sum_j \Big(\frac{\partial G_i'}{\partial x_j} - B_{ij} \Big) \dot{u}_j$$

$$+ \quad \sum_j \Big(\frac{\partial G_i'}{\partial x_j} - C_{ij} \Big) u_j + R_i \Bigg] \qquad i = 1, \ldots, r$$

$$0 = G_i'(\dot{x}', x', t) + \sum_j B_{ij} \dot{u}_j + \sum_j C_{ij} u_j$$

$$+ \quad \eta \Bigg[\sum_j \Big(\frac{\partial G_i'}{\partial \dot{x}_j} - B_{ij} \Big) \dot{u}_j + \sum_j \Big(\frac{\partial G_i'}{\partial x_j} - C_{ij} \Big) u_j + R_i \Bigg]$$

$$i = r+1, \ldots, n \qquad (4.12)$$

where η is an α-like parameter which we introduce. For $\eta = 1$ Eq. (4.12) is equivalent to the original system, Eqs. (4.6), and for $\eta = 0$, the system of Eq. (4.12) is a system with piecewise constant coefficients.

<u>LEMMA 4.1.</u> <u>If Hypothesis 4.1 above is satisfied for the given problem originally, it is also satisfied by the system of Eq. (4.12) at</u> $\eta = 0$.

<u>Proof</u>. The conditions of Hypothesis 4.1 are to be interpreted as the statement that certain Jacobians are not zero and certain additional statements. If we set $\eta = 0$, the corresponding Jacobians are readily seen to coincide with certain values of the former set of Jacobians and hence are not zero. Since at $\eta = 0$ we have a linear system in the u's, the existence of solutions at specific points is readily established. This establishes the Lemma.

The fact that Hypothesis 4.1 is satisfied for $\eta = 0$ and the fact that Eq. (4.12) for $\eta = 0$ is a linear system insures that given any initial condition $\dot{u}_{1,o}, \ldots, \dot{u}_{r,o}, u_{1,o}, \ldots, u_{n,o}$ we can find a solution corresponding to $\eta = 0$.

Furthermore, since under these circumstances the hypothesis of § 2.1 of Chapter 2 applies to this system and this solution, the $\dot{u}_1, \ldots, \dot{u}_r$ and $u_1, \ldots, u_r$ are analytic functions of η which, for a given value of $t = t*$, are analytic for some η neighborhood around zero. If this η neighborhood includes 1 we shall say the set of initial conditions $\dot{u}_{1,o}, \dots, \dot{u}_{r,o}, u_{1,o}, \dots, u_{n,o}$ is in case 1.

We suppose now we are dealing with a set of initial conditions in case 1. For such a case the u_i's can be written as

$$u_i(t,\eta,\lambda) = u_i(t,0,\lambda) + u_{i,\eta}(t,0,\lambda)\eta + \frac{1}{2!} u_{i,\eta,\eta}(t,0,\lambda)\eta^2 \qquad (4.13)$$

$$+ \dots .$$

The $u_i(t,0,\lambda)$ satisfy the system of equations

$$0 = G_i^!(\lambda \ddot{x}', \dot{x}', x', t) + \sum_j A_{ij} \lambda \ddot{u}_j + \sum_j B_{ij}\dot{u}_j + \sum_j C_{ij} u_j, \quad i = 1, \dots, r$$

$$0 = G_i^!(\dot{x}', x', t) + \sum_j B_{ij}\dot{u}_j + \sum_j C_{ij} u_j \qquad i = r+1, \dots, n \qquad (4.14)$$

which is obtained from Eq. (4.12) by setting $\eta = 0$. Corresponding equations for the higher derivatives are obtained by differentiating Eq. (4.12) with respect to η and setting $\eta = 0$ in the result. We abbreviate the coefficients of η in Eq. (4.12):

$$S_i = \sum_j \left(\frac{\partial G_i^!}{\partial(\lambda \ddot{x}_j)} - A_{ij}\right) \lambda \ddot{u}_j + \sum_j \left(\frac{\partial G_i^!}{\partial \dot{x}_j} - B_{ij}\right)\dot{u}_j + \sum_j \left(\frac{\partial G_i^!}{\partial x_j} - C_{ij}\right) u_j + R_i$$

$$i = 1, \dots, r \qquad (4.15)$$

$$S_i = \sum_j \left(\frac{\partial G_i^!}{\partial \dot{x}_j} - B_{ij}\right)\dot{u}_j + \sum_j \left(\frac{\partial G_i^!}{\partial x_j} - C_{ij}\right) u_j + R_i, \quad i = r+1, \dots, n. \quad .$$

In this expression u, $\dot{u}$, $\ddot{u}$ are to be regarded as functions of η. If the superscript is to denote differentiating with respect to η then differentiating Eq. (4.12) k times with respect to η and setting $\eta = 0$ in the result yields

$$0 = \sum_j A_{ij} \lambda \ddot{u}_j^{(k)} + \sum_j B_{ij}\dot{u}_j^{(k)} + \sum_j C_{ij} u_j^{(k)} + kS_i^{(k-1)} \quad i = 1, \dots, r \qquad (4.16)$$

$$0 = \sum_j B_{ij}\dot{u}_j^{(k)} + \sum_j C_{ij} u_j^{(k)} + kS_i^{(k-1)} \qquad i = r+1, \dots, n.$$

This, of course, is again a system which can be solved successively for the various values of k. To do so we must first solve the homogeneous system

$$0 = \sum_j A_{ij} \lambda \ddot{u}_j^{(k)} + \sum_j B_{ij}\dot{u}_j^{(k)} + \sum_j C_{ij}u_j^{(k)} \qquad i = 1,\ldots,r \tag{4.17}$$
$$0 = \sum_j B_{ij}\dot{u}_j^{(k)} + \sum_j C_{ij}u_j^{(k)} \qquad i = r+1,\ldots,n.$$

This system is not one with constant coefficients but one in which the coefficients are constant over successive time intervals. Nevertheless we can construct $n+r$ linearly independent solutions and use these in turn to solve the non-homogeneous systems of Eqs. (4.14) and(4.16).

4.4 The Sensitivity Characteristics

We first consider the system of Eq. (4.17) for a fixed interval. Since this is a system with constant coefficients, we consider the indicial equation which can be written in determinantal form as

$$|A_{ij} \lambda m^2 + B_{ij}m + C_{ij}| = 0. \tag{4.18}$$

(We let $A_{ij} = 0$ for $i = r+1,\ldots,n.$)

If we set $\lambda = 0$ in the above, we obtain an n th order polynomial

$$|B_{ij}m + C_{ij}| = 0. \tag{4.19}$$

The matrix $B = || B_{ij} ||$ corresponds to the Jacobian of the original system of Eq. (4.1) although it is not precisely equal to it as long as λ' is not zero. In view of this though, it is reasonable to assume that $|| B_{ij} ||$ is not singular. If we multiply the determinant of Eq. (4.19) by the determinant of the inverse matrix B^{-1} we have a characteristic equation

$$| mI + B^{-1}C | = 0$$

which clearly is of the n th degree in m. ($C = ||C_{ij}||$.)

ASSUMPTION 4.1. $B = || B_{ij} ||$ is not singular.

ASSUMPTION 4.2. Equation (4.19) does not have multiple roots.

LEMMA 4.2. Assumption 4.1 implies that Eq. (4.19) is an equation of the n th degree.

LEMMA 4.3. There is a λ neighborhood of $\lambda = 0$ such that corresponding to each root m_o of Eq. (4.19) which is not a multiple root of Eq. (4.19) there exists a root $\mu(\lambda)$ of Eq. (4.18) such that $\mu(0) = m_o$ and $\mu(\lambda)$ is analytic in λ at $\lambda = 0$.

Proof. Consider

$$L(\lambda, m) = | A_{ij} \lambda m^2 + B_{ij} m + C_{ij} | \qquad (4.20)$$

and consider Eq. (4.18), i.e., $L = 0$ as an equation intended to determine m as a function of λ. If we set $\lambda = 0$, m_o is a value of m satisfying $L(0, m_o) = 0$ and we may apply the usual implicit function theorem to this situation and obtain the desired result provided $\frac{\partial L}{\partial m} \neq 0$ at $\lambda = 0$, $m = m_o$. But L reduces to Eq. (4.19) for $\lambda = 0$ and the statement $\frac{\partial L}{\partial m} \neq 0$ is immediately equivalent to the statement that m_o is not a double root of Eq. (4.19).

(Multiple roots m_o lead to expansions in fractional powers of λ rather than a simple power series. Otherwise the results are analogous, but it should be appreciated that small variations in λ yield far larger variations in, say, the square root of λ. Consequently, in the multiple root case, the root $\mu(\lambda)$ depends much more sensitively on λ than in the analytic case for small values of λ.)

Under Assumptions 4.1 and 4.2, then, we have n roots $\mu_1(\lambda), \ldots, \mu_n(\lambda)$ of Eq. (4.18) associated with corresponding roots of Eq. (4.19). (One might well point out here that in the case of no λ errors we could still apply the process of the previous section in order to obtain a system of equations with piecewise constant coefficients. The result would be Eq. (4.19). For these we would have the n roots $m_1, \ldots, m_n$ and correspondingly n solutions. The introduction of the λ errors has affected these only slightly under Assumptions 4.1 and 4.2. We have n analytic functions of λ, $\mu_1(\lambda), \ldots, \mu_n(\lambda)$ each of which reduces to an m at $\lambda = 0$. The same holds for the corresponding solutions.)

Thus these n solutions correspond to relatively minor variations of the basic situation introduced by the λ error.

However, there are other solutions which must be investigated.

LEMMA 4.4. Equation (4.18) is of degree n+r in m.

Proof. Consider first the determinant of the first r rows and r columns of Eq. (4.18)

$$|A_{ij} \lambda m^2 + B_{ij} m + C_{ij}| \qquad i,j = 1,\ldots,r \tag{4.21}$$

This determinant is a polynomial in m of degree at most 2r. Since the coefficient of m^{2r} is precisely the determinant $|A_{ij}| \lambda^{2r}$ and the last is not zero by Hypothesis 4.1, we see that this polynomial is of the 2r degree. Now one can also readily show that any other rth row determinant from the first r rows is of degree $\leq$ 2r-1 since only the first r columns are of the second degree in m.

In the last n-r rows we take the last n-r columns and consider the corresponding minor

$$|B_{ij} m + C_{ij}| \qquad i,j = n-r+1,\ldots,n. \tag{4.22}$$

This is of degree $\leq$ n-r with $|B_{ij}|$ as the coefficient of m^{n-r}. Hypothesis 4.1 states that this determinant is not zero and consequently Eq. (4.22) is of degree n-r. Any other minor from the last n-r rows is of degree $\leq$ n-r.

Apply then the Laplace expansion of L for the first r rows. This expansion contains the product of Eqs. (4.21) and (4.22) which is of the n+r degree. Since the degree of the other minors of the first r rows is less than 2r and the degree of the minors of the last n-r rows does not exceed n-r, the degree of any other product in this expansion is less than n+r. Hence L is of degree n+r.

Thus there are r solutions of Eq. (4.18) which we must still obtain. Multiply Eq. (4.18) by λ^n. This will permit us to multiply each element in the determinant by λ,

$$|A_{ij}(\lambda m)^2 + B_{ij}(\lambda m) + C_{ij} \lambda| \quad = \quad 0. \tag{4.23}$$

Let $\nu = \lambda m$. Then Eq. (4.23) becomes

$$|A_{ij} \nu^2 + B_{ij} \nu + C_{ij} \lambda| \quad = \quad 0. \tag{4.24}$$

Equation (4.24) is, of course, of the same degree in ν as Eq. (4.18) was in m, i.e., of degree n+r. Let us now consider Eq. (4.24) for $\lambda = 0$, obtaining

$$|A_{ij}\nu^2 + B_{ij}\nu| = 0. \tag{4.25}$$

We can factor a ν from each column of this determinant which can then be written as

$$\nu^n |A_{ij}\nu + B_{ij}| = 0. \tag{4.26}$$

Now

$$|A_{ij}\nu + B_{ij}| = 0 \tag{4.27}$$

is an equation of degree r in ν. Suppose $\nu_{1,o}, \ldots, \nu_{r,o}$ are the r roots of Eq. (4.27).

ASSUMPTION 4.3. Equation (4.27) does not have multiple roots.

Assumption 4.1 implies that no $\nu_{k,o}$ is zero.

LEMMA 4.5. Under Assumptions 4.1 and 4.3, there exists a λ neighborhood of $\lambda = 0$ such that for each root $\nu_{k,o}$ of Eq. (4.27), there exists an analytic solution $\nu(\lambda)$ of Eq. (4.24).

Proof. The proof of this is similar to that of Lemma 4.3. If we retrace the steps from Eq. (4.24) to Eq. (4.27) in reverse order we find that each of the solutions of Eq. (4.27) yields a solution of Eq. (4.24) for $\lambda = 0$. Assumption 4.1 shows that $\nu = 0$ is not one of the $\nu_{k,o}$ and by Assumption 4.3, the $\nu_{k,o}$ are distinct. Consequently, no $\nu_{k,o}$ is a multiple root of Eq. (4.24) at $\lambda = 0$, and $\frac{\partial L}{\partial \nu} \neq 0$ at each of the values $\lambda = 0$, $\nu = \nu_{k,o}$. The implicit function theorem can now be applied to obtain the result specified.

Corresponding to the r solutions $\nu(\lambda)$ of Eq. (4.24) there are r solutions $m = \nu(\lambda)/\lambda$ of Eq. (4.23) and consequently, r solutions of Eq. (4.18).

The situation for any multiple root of Eq. (4.27) is analogous to that for any multiple root of Eq. (4.22). We summarize our results in the following theorem.

THEOREM 4.1 Under Hypothesis 4.1 of § 4.2 and Assumptions 4.1, 4.2 and 4.3 above, for each interval of constancy the indicial equation Eq. (4.18) of the system Eq. (4.17) possesses n roots $\mu_1(\lambda), \dots, \mu_n(\lambda)$ analytic in λ and r roots $\nu_1(\lambda)/\lambda, \dots, \nu_r(\lambda)/\lambda$ where $\nu_1(\lambda), \dots, \nu_r(\lambda)$ are analytic in λ. If Assumptions 4.1 or 4.2 do not hold, the μ 's or the ν 's can be expressed in fractional powers of λ.

4.5 The Solution for a Single Interval

To each solution $\mu_j(\lambda)$ of the theorem of the previous section, we can find a solution of Eq. (4.17) in the form

$$u_i^j = D_{ij} e^{\mu_j(\lambda)t} \tag{4.28}$$

on the interval of constancy. The D_{ij} are the nontrivial solutions of the homogeneous system

$$\sum_k (A_{ik}\lambda \mu_j^2 + B_{ik}\mu_j + C_{ik})D_{kj} = 0, \quad i = 1, \dots, n \tag{4.29}$$

whose determinant is zero. ($A_{ij} = 0$, $i = r+1, \dots, n.$). Similarly for each of the $\nu_j(\lambda)$ we have solutions in the form

$$v_i^j = E_{ij} e^{(\nu_j(\lambda)/\lambda)t} \tag{4.30}$$

where for each j

$$\sum_k (A_{ik}\nu_j^2 + B_{ik}\nu_j + \lambda C_{ik})E_{kj} = 0. \tag{4.31}$$

Notice that the Eq. (4.29) and (4.31), which determine D_{kj} and E_{kj} when a suitable normalization is supposed, are well defined at $\lambda = 0$ and are analytic in λ at $\lambda = 0$.

The general solution of the system

$$\sum_k A_{ik}\lambda \ddot{u}_k + \sum_k B_{ik}\dot{u}_k + \sum_k C_{ik}u_k = 0 \tag{4.32}$$

is a linear combination with constant coefficients of Eqs. (4.28) and (4.30). In the vector notation

$$u = \sum_{j=1}^{n} U_j u^j + \sum_{k=1}^{r} V_k v^k . \tag{4.33}$$

Here u^j and v^k stand for vectors with elements given by Eqs. (4.28) and (4.30) above. Now let U stand for the vector with elements U_j, V for that with elements V_j, M for the diagonal matrix with elements $\mu_j(\lambda)\,\delta_{ij}$, N the r *th* order diagonal matrix with elements $\nu_j(\lambda)\,\delta_{ij}$. Then Eq. (4.33) can be written

$$u = D(\exp Mt)U + E[\exp((N/\lambda)t)]\,V. \tag{4.34}$$

(D is the matrix $||D_{ij}||$, E is an n by r matrix $||E_{ij}||$.)

Let D^r and E^r denote the matrix of the first r rows of D and E respectively. Then the initial conditions for the solution u can be written in vector form as

$$\begin{aligned} u^o &= DU + EV \\ \lambda\dot{u}^o &= \lambda D^r M\dot{U} + E^r NV. \end{aligned} \tag{4.35}$$

U and V as vectors can also be considered as one column matrices. Similarly, u^o and $\dot{u}^o$ are one column matrices. Thus,

$$U = D^{-1}u^o - D^{-1}EV$$

and

$$\lambda(\dot{u}^o - D^r MD^{-1}u^o) = (E^r N - \lambda D^r MD^{-1}E)V.$$

Now, let

$$T = E^r N - \lambda D^r MD^{-1}E. \tag{4.36}$$

Then

$$V = \lambda T^{-1}(\dot{u}^o - D^r MD^{-1}u^o) \tag{4.37}$$

$$U = D^{-1}u^o - \lambda D^{-1}ET^{-1}(\dot{u}^o - D^r MD^{-1}u^o). \tag{4.38}$$

Notice V is zero and $U = D^{-1}u^o$ provided

$$\dot{u}^o = D^r MD^{-1}u^o. \tag{4.39}$$

LEMMA 4.6. Equation (4.39) is the necessary and sufficient condition that Eq. (4.34) depend analytically on λ *at* $\lambda = 0$.

These, of course, all apply to an interval of t on which the coefficients A_{ij}, B_{ij}, C_{ij} are constant. Notice, however, that if $R(\nu_j) < 0$ (real part of ν_j) and of finite size while λ is small, the nonanalytic terms of Eq. (4.34) are negligible for non-zero positive values of t of finite size.

Equations (4.34), (4.37) and (4.38) can be combined to yield

$$u = [\,D \exp(Mt)D^{-1}\,]\, u^{o} + \lambda[E \exp(\,(N/\lambda)t) - D \exp(Mt)D^{-1}E]\; T^{-1} \\ .(\dot{u}^{o} - D^{r}MD^{-1}u^{o}) \qquad (4.40)$$

$$= \exp(\overline{M}t)u^{o} + \lambda[E \exp(\,(N/\lambda)t) - \exp(\overline{M}t)E]\; T^{-1}(\dot{u}^{o} - D^{r}D^{-1}\overline{M}u^{o})$$

where $\overline{M} = DMD^{-1}$.

4.6 The Continuation of the Solution

Equation (4.40) applies to each interval of constancy and relates a solution of the homogeneous system Eq. (4.17), u, to its initial values $\dot{u}^{o}$ and u^{o} for the interval. u^{o} is an n-dimensional vector, $\dot{u}^{o}$ is r-dimensional. A subscript following a colon will indicate the interval, i.e. $u_{j:k}$ is the j th component of the solution vector on the k th interval. Equation (4.40) can then be written

$$u_{:k} = \exp[\,\overline{M}_{:k}(t - t_{k-1})\,]\, u^{o}_{:k} + \lambda\Big\{E_{:k}\exp[\,(N_{:k}/\lambda)(t - t_{k-1})] \\ - \exp[\,\overline{M}_{:k}(t - t_{k-1})]\; E_{:k}\Big\}\quad T^{-1}_{:k}(\dot{u}^{o}_{:k} - D^{r}_{:k}D^{-1}_{:k}\overline{M}_{:k}u^{o}_{:k}) \qquad (4.41)$$

which implies

$$\dot{u}_{:k} = \overline{M}_{:k}\exp[\,\overline{M}_{:k}(t - t_{k-1})]\; u^{o}_{:k} + \Big\{E_{:k}N_{:k}\exp[\,(N_{:k}/\lambda)(t - t_{k-1})] \\ - \lambda\overline{M}_{:k}\exp[\,\overline{M}_{:k}(t - t_{k-1})\,]\, E_{:k}\Big\}\quad T^{-1}_{:k}(\dot{u}^{o}_{:k} - D^{r}_{:k}D^{-1}_{:k}\overline{M}_{:k}u^{o}_{:k}). \qquad (4.42)$$

Let H denote the r by n dimensional matrix whose first r by r minor matrix is the identity and the rest zero. Then

$$D^{r} = HD \qquad \text{and} \qquad D^{r}D^{-1} = H.$$

It is now desirable to obtain the relation between $u^o_{:k-1}$ and $\dot{u}^o_{:k-1} - H\bar{M}_{:k-1}u^o_{:k-1}$ and $u^o_{:k}$ and $\dot{u}^o_{:k} - H\bar{M}_{:k}u^o_{:k}$, (cf. Eq. (4.39)). Let $\Delta t_{k-1} = t_{k-1} - t_{k-2}$.

The given solution is continuous at $t = t_{k-1}$ if

$$u^o_{:k} = \exp(\bar{M}_{:k-1}\Delta t_{k-1})u^o_{:k-1} + \lambda\Big\{E_{:k-1}\exp[(N_{:k-1}/\lambda)\Delta t_{k-1}] - \exp(\bar{M}_{:k-1}\Delta t_{k-1})E_{:k-1}\Big\} T^{-1}_{:k-1}(\dot{u}^o_{:k-1} - H\bar{M}_{:k-1}u^o_{:k-1}). \tag{4.43}$$

The derivatives of the first r of the u_j's are continuous if

$$\dot{u}^o_{:k} = H\bar{M}_{:k-1}\exp(\bar{M}_{:k-1}\Delta t_{k-1})u^o_{:k-1} \tag{4.44}$$

$$+ H\Big\{E_{:k-1}N_{:k-1}\exp[(N_{:k-1}/\lambda)\Delta t_{k-1}] - \lambda\bar{M}_{:k-1}\exp(\bar{M}_{:k-1}\Delta t_{k-1})E_{:k-1}\Big\}$$

$$\cdot T^{-1}_{:k-1}(\dot{u}^o_{:k-1} - H\bar{M}_{:k-1}u^o_{:k-1}) .$$

Now Let $v^o_{:k} = u^o_{:k} - H\bar{M}_{:k}u^o_{:k}$. Equation (4.43) can then be written as

$$u^o_{:k} = \exp(\bar{M}_{:k-1}\Delta t_{k-1})u^o_{:k-1} + \lambda\Big\{E_{:k-1}\exp[(N_{:k-1}/\lambda)\Delta t_{k-1}] - \exp(\bar{M}_{:k-1}\Delta t_{k-1})E_{:k-1}\Big\} T^{-1}_{:k-1}v^o_{:k-1} \tag{4.45}$$

and

$$\begin{aligned} v^o_{:k} = H\Big\{ &(\bar{M}_{:k-1} - \bar{M}_{:k})\exp(\bar{M}_{:k-1}\Delta t_{k-1})u^o_{:k-1} \\ &+ [E_{:k-1}(N_{:k-1} - \lambda\bar{M}_{:k})\exp((N_{:k-1}/\lambda)\Delta t_{k-1}) \\ &- \lambda(\bar{M}_{:k-1} - \bar{M}_{:k})\exp(\bar{M}_{:k-1}\Delta t_{k-1})E_{:k-1}]T^{-1}_{:k-1}v^o_{:k-1}\Big\} . \end{aligned} \tag{4.46}$$

Thus, Eqs. (4.45) and (4.46) permit us to continue a solution of the homogeneous solution, through intervals. For suitably stable matrices, the exponential matrices should cause such a solution to decay.

In general, the solution Eq. (4.41) consists of two parts. One part is analytic in λ, i.e., the term involving $\exp \overline{M}(t - t_{k-1})$. Normally one would expect this to decay slowly. On the other hand, the matrix $\exp(N(\lambda)/\lambda)(t - t_{k-1})$, if it is properly stable, should have very large negative real parts for the characteristic roots and hence should decay quickly. This is, however, non-analytic in λ. The possibility that $\exp \overline{M}(t-t_{k-1})$ is unstable must be considered (cf. § 4.1).

4.7 Construction of the Solution

The results of the previous section can be used to solve the system of Eq. (4.16) as follows: the solution desired is that for the non-homogeneous system, Eq. (4.16), which at $t = t_o$ has zero values for $u_1, \ldots, u_n, \dot{u}_1, \ldots, \dot{u}_r$. For the first interval $t_o \leq t \leq t_1$ a solution $y_{:1}$ will be found of Eq. (4.16). This will have certain values at t_1 which can be used as the initial values for a solution of the homogeneous equation for the second interval and this solution can be continued as a solution of the homogeneous solution for the remaining intervals. In general, let $y_{:k}$ denote the vector solution $y_{1:k}, \ldots, y_{n:k}$, which is zero for the intervals preceding the k th, is a solution of Eq. (4.16) on the k th interval with initial values 0, and is continued continuously by the method of the previous section as a solution of the homogeneous equation over the remaining intervals.

Since it is obvious that the desired solution $u_1, \ldots, u_n$ is the sum of the $y_{1:k}, \ldots, y_{n:k}$ for all intervals I_k, it remains only to solve the problem of finding, for a given interval I_k, the solution of the non-homogeneous system Eq. (4.16) which at the lower end point has zero initial values. Thus again one needs to consider only a single interval and it is not necessary to make explicit reference to the interval.

It is desirable to solve the first r equations of Eq. (4.16) for $\lambda \ddot{z}_j$, the remaining for $\dot{z}_j$. Thus,

$$\lambda \ddot{z}_i = \sum_{j=1}^{n} \beta'_{ij} \dot{z}_j + \sum_{j=1}^{n} \gamma'_{ij} z_j + s'_i \qquad i = 1, \ldots, r \qquad (4.47)$$

$$\dot{z}_i = \sum_{j=1}^{n} \gamma'_{ij} z_j + s'_i \qquad i = r+1, \ldots, n$$

where s_i' is a linear combination of the original non-homogeneous terms. It is desirable to use the last equations to eliminate $\dot{z}_{r+1}, \ldots, \dot{z}_n$ from the first r which will then become

$$\lambda \ddot{z}_i = \sum_{j=1}^{r} \beta_{ij}\dot{z}_j + \sum_{j=1}^{n} \gamma_{ij} z_j + s_i \qquad i = 1, \ldots, r$$

$$\dot{z} = \sum_{j=1}^{n} \gamma_{ij} z_j + s_i \qquad i = r+1, \ldots, n. \tag{4.48}$$

Referring to § 4.5, one may recall that the customary method of finding a solution of the non-homogeneous equation, Eq. (4.48), is to consider the expression Eq. (4.33) with the constants U_j and V_k replaced by functions $\varphi_j(t)$ and $\psi_k(t)$ of t

$$z_i = \sum_{j=1}^{n} \varphi_j(t) u_i^j + \sum_{k=1}^{r} \psi_k(t) v_i^k$$

$$= \sum_{j=1}^{n} D_{ij}\ \varphi_j(t) e^{\mu_j t} + \sum_{k=1}^{r} E_{ij}\ \psi_j(t) e^{(\nu_j/\lambda)t}, \tag{4.49}$$

where u_i^j and v_i^j are given by Eqs. (4.28) and (4.30). Since Eq. (4.48) is of the second order in z_i, $i = 1, \ldots, r$, one must introduce auxiliary conditions

$$\sum_{j=1}^{n} \dot{\varphi}_j u_i^j + \sum_{k=1}^{r} \dot{\psi}_k v_i^k = 0 \qquad i = 1, \ldots, r. \tag{4.50}$$

Since the u_i^j and v_i^k are solutions of the homogeneous system, one can readily show that Eq. (4.48) is equivalent to

$$\lambda \left[\sum_{k=1}^{r} \dot{\psi}_k \dot{v}_i^k + \sum_{j=1}^{n} \dot{\varphi}_j \dot{u}_i^j \right] = s_i, \qquad i = 1, \ldots, r \tag{4.51}$$

$$\sum_{k=1}^{r} \dot{\psi}_k v_i^k + \sum_{j=1}^{n} \dot{\varphi}_j u_i^j = s_i \qquad i = r+1, \ldots, n .$$

Combining these with Eqs. (4.28) and (4.30), one obtains

$$\sum_{k=1}^{r} E_{ik}\,\nu_k\,\dot{\psi}_k e^{(\nu_k/\lambda)t} + \sum_{j=1}^{n} D_{ij}\,\lambda\,\mu_j\,\dot{\varphi}_j e^{\mu_j t} = s_i, \quad i = 1,\dots,r$$

$$\sum_{k=1}^{r} E_{ik}\,\dot{\psi}_k e^{(\nu_k/\lambda)t} + \sum_{j=1}^{n} D_{ij}\,\dot{\varphi}_j e^{\mu_j t} = 0, \quad i = 1,\dots,r$$

$$\sum_{k=1}^{r} E_{ik}\,\dot{\psi}_k e^{(\nu_k/\lambda)t} + \sum_{j=1}^{n} D_{ij}\,\dot{\varphi}_j e^{\mu_j t} = s_i, \quad i = r+1,\dots,n. \qquad (4.52)$$

Now, if one introduces $\Psi_k = \dot{\psi}_k e^{(\nu_k/\lambda)t}$, $\Phi_j = \dot{\varphi}_k e^{\mu_j t}$

Eq. (4.52) becomes

$$\sum_{k=1}^{r} E_{ik}\,\nu_k\,\Psi_k + \sum_{j=1}^{n} D_{ij}\,\lambda\,\mu_j\,\Phi_j = s_i, \quad i = 1,\dots,r$$

$$\sum_{k=1}^{r} E_{ik}\,\Psi_k + \sum_{j=1}^{n} D_{ij}\,\Phi_j = 0, \quad i = 1,\dots,r \qquad (4.53)$$

$$\sum_{k=1}^{r} E_{ik}\,\Psi_k + \sum_{j=1}^{n} D_{ij}\,\Phi_j = s_i, \quad i = r+1,\dots,n.$$

This is a linear system of equations with constant coefficients on the Φ_k and Ψ_k as far as t is concerned. Furthermore, for $\lambda = 0$, the determinant is in the form of the product of the determinant of E^r and D evaluated at $\lambda = 0$ (Cf. § 4.5) and hence is not zero. Since this determinant is a continuous function of λ, one may suppose the existence of a range of λ around $\lambda = 0$, where this determinant is not zero.

Solving for Ψ_k and Φ_k expresses these also as linear combinations with constant coefficients of the original inhomogeneous terms.

$$\Psi_k = s_k^*(t), \qquad \Phi_j = s_j''(t). \qquad (4.54)$$

The coefficients depend analytically on λ and thus,

$$\dot{\Psi}_k = e^{-(\nu_k/\lambda)t} s_k^*(t), \qquad \dot{\varphi}_j = e^{-\mu_j t} s''_j(t) . \tag{4.55}$$

Since at the initial point t_{k-1}, ψ_j and φ_j are zero one has for the desired solution of Eq. (4.47)

$$z_i(t) = \sum_{j=1}^{n} D_{ij} e^{\mu_j t} \int_{t_{k-1}}^{t} e^{-\mu_j \tau} s''_j(\tau) d\tau$$

$$+ \sum_{j=1}^{r} E_{ij} e^{(\nu_j/\lambda)t} \int_{t_{k-1}}^{t} e^{-(\nu_j/\lambda)\tau} s_j^*(\tau) d\tau \tag{4.56}$$

valid in the interval I_k.

We conclude therefore:

LEMMA 4.7. The solution $y_{i:k}$ *defined at the beginning of this section has the form of Eq. (4.56) on the* k *th interval. The quantities and functions* D_{ij}, μ_j, $s''_j(\tau)$, E_{ij}, ν_j, $s^*(\tau)$ *depend analytically on* λ *in some neighborhood of* $\lambda = 0$.

For each interval then the solution $u_1, \ldots, u_n$ of Eq. (4.16) is in the form $\sum_{j=1}^{k} y_{:j}$. For $j < k$, the y_{ij} are linear combinations of terms in the form $e^{\mu_j t}$ and $e^{(\nu_j/\lambda)t}$. Normally one would expect that the terms $e^{\mu_j t}$ would decay slowly and those in the form $e^{(\nu_j/\lambda)t}$ would decay quickly. The above lemma shows that $y_{i:k}$ may also be divided into two terms similar in nature to these two types. Thus it is clear that while the λ errors do introduce errors in the solution, nonanalytic in λ, these errors decay quickly if the ν_j have negative real parts. (However, cf. § 4.1.)

4.8 Concluding Discussion

It is now possible to set up a general theory of the error, including all types of errors, α, β and λ. We return now to Eqs. (4.2) of § 4.2 above and permit α errors to appear. We will suppose that the system of Eq. (4.2) is equivalent to a system of Eq. (4.6) subject to Hypothesis 4.1. We now introduce u and pass to the system with two additional α-like parameters, λ'(cf. Eq. (4.9)) and η (cf. Eq. (4.12)). For λ fixed, the resulting system can readily be expanded to a first order system of the type treated in Chapter 2 and thus our previous theory applies to this system for α errors including η and λ', and β errors.

This means that for λ fixed we can expand the error function $u(t, \lambda, \alpha, \beta)$ in terms of α and β. Thus

$$u(t,\lambda,\alpha,\beta) = u(t,\lambda,0,0) + \sum_{\lambda = \alpha,\beta} u_\gamma(,\lambda,\ 0,\ 0)\,\gamma + \dots \quad . \qquad (4.57)$$

The argument of this chapter yields $u(t, \lambda, 0, 0)$. For the various partial derivatives of u relative to α and β, the discussion of Chapter 3 above is applicable to the extended first order system. However, the effect of the extension can be obtained simply by replacing Eq. (3.9) of Chapter 3 by

$$\sum_{j=1}^{r} A_{ij}\lambda \ddot{u}_j + \sum_{j=1}^{n} B_{ij}\dot{u}_j + \sum_{j=1}^{n} C_{ij}u_j + Q_i = 0, \qquad (4.58)$$

Q_i is the same as in Chapter 3, § 3.2, and the rest of the equation corresponding to the similar terms which appeared in Eq. (4.16) of the present chapter.

The discussion of the previous sections show that in each case there is a favorable situation where the expression for the partial

$$\frac{\partial^{r} u}{\partial\gamma_1^{\,r_1} \cdots \partial\gamma_u^{\,r_u}}$$

can be expressed as a sum of two terms, one of which is analytic in λ, while the other term, based on $\nu_j(\lambda)$ with negative real parts, disappears rapidly. Consequently, we have a similar resolution for $u(t, \lambda, \alpha, \beta)$ in the favorable case in which the $\nu_j(\lambda)$ have negative real parts. If the real part of $\nu_j(\lambda)$ is positive, the $\nu_j(\lambda)$ terms for λ small will be dominant and very large, and normally one would expect the machine solution to be useless.

The above discussion is based on a number of assumptions. One of these is that relative to the parameter η, the solution is in case 1. Another is that the order of the system increases by no more than 1, when introduced into the machine. However, this last assumption was introduced merely for convenience. One can readily indicate a procedure which is applicable in the higher order case. The essential part of this discussion is the generalization of the situation represented by Hypothesis 4.1 of § 4.2 above. The given system $F_i = 0$ is supposed to be in the first order case. Let s denote the order of the highest derivative which appears in the machine equations $G_i = 0$. If our given system $G_i = 0$ implies $n - r_s$ relations between the derivatives of lower order, we suppose that we can separate out r_s equations involving the s-order derivatives and that in these in turn we can eliminate all but r_s s-order derivatives. In the remaining $n - r_s$ relations, we suppose that we have $n - r_s - r_{s-1}$ relations of order less than $s-1$. A similar procedure then permits us to separate out r_{s-1} relations on r_{s-1} derivatives of order $s-1$.

Thus we can obtain, after a suitable re-enumeration of the dependent variables, a system equivalent to the original broken up into s sets of equations. The first set consists of r_s equations on the s-order derivatives of $x_1, \ldots, x_{r_s}$ and do not involve any further derivatives of order s. The next set consists of r_{s-1} equations on the $s-1$ derivatives of $x_{r_s+1}, \ldots, x_{r_s+r_{s-1}}$ and contains no higher derivatives and no other derivatives of order $s-1$. A similar situation holds for the remaining sets of equations. The parameter λ is most effectively introduced by writing the functions as depending on

$\lambda^{s-1}x_j^{(s)}$, $\lambda^{s-2}x_j^{(s-1)}$, etc.

The linearization process of § 4.2 of this chapter may now be applied. Setting $\lambda = 0$ will again yield n values $\mu_{1,o}, \ldots, \mu_{n,o}$ which determine functions $\mu_j(\lambda)$ analytic in λ at $\lambda = 0$ (cf. § 4.4 above). On the other hand, the remaining portions of § 4.4 show that one has a total of $r_2 + 2r_3 + \ldots + (s-1)r_s$ extra ν roots of the equation equivalent to Eq. (4.24) of § 4.4 above. From this point on, the results in the higher order case are precisely similar to those in the case in which the order is raised one.

Returning now to the case in which the order is raised one, we wish to discuss the alternative to the assumption above that our problem is in case 1. Normally one would want the R_i of Eq. (4.10) of § 4.3 above to be small and this requires that the $\dot{u}_1, \ldots, \dot{u}_r$ be initially small, since it is reasonable to expect that the $u_1, \dots, u_n$ are small. However, unless special precautions are taken, the $\dot{u}_1, \dots, \dot{u}_r$ will not in general be small and their size can lead to convergence difficulties or even to slow convergence in case 1.

There is a process which permits one to study the situation which arises in the case of large $\dot{u}_{1,o}, \dots, \dot{u}_{r,o}$, i.e., the case where initially the machine solution and the true solution have different rates of change.

Consider the equations

$$G_i(\lambda \ddot{u}_1, \dots, \lambda \ddot{u}_r, \dot{u}_1, \dots, \dot{u}_n, u_1, \dots, u_n, t) = 0$$

$$i = 1, \dots, r$$

$$G_i(\dot{u}_1, \dots, \dot{u}_n, u_1, \dots, u_n, t) = 0 \qquad i = r+1, \dots, n$$

and let us change the independent variable in u from t to a variable $\tau = \lambda^{-1} t$ and let the dot refer to differentiation relative to τ. The result is

$$G_i(\lambda^{-1} \ddot{u}_1, \dots, \lambda^{-1} \ddot{u}_r, \lambda^{-1} \dot{u}_1, \dots, \lambda^{-1} \dot{u}_n, \lambda \tau) = 0$$

$$G_i(\lambda^{-1} \dot{u}_1, \dots, \lambda^{-1} \dot{u}_n, u_1, \dots, u_n, \lambda \tau) = 0.$$

The effect of the λ^{-1} is to emphasize those variables which it multiplies and in general one can consider this system equivalent to

$$H_i(\ddot{u}_1, \dots, \ddot{u}_r, \dot{u}_1, \dots, \dot{u}_n, \lambda, u_1, \dots, u_n, \lambda \tau) = 0 \quad i = 1, \dots, r$$

$$H_i(\dot{u}_1, \dots, \dot{u}_n, \lambda, u_1, \dots, u_n, \lambda \tau) = 0 \qquad i = r+1, \dots, n,$$

in which λ plays the role of a parameter. Thus, if we set $\lambda = 0$ we obtain relations

$$H_i^o(\ddot{u}_1, \dots, \ddot{u}_r, \dot{u}_1, \dots, \dot{u}_n) = 0 \qquad i = 1, \dots, r$$

$$H_i^o(\dot{u}_1, \dots, \dot{u}_n) = 0 \qquad i = r+1, \dots, n.$$

These equations can be considered as a system of equations for $\dot{u}_1, \dots, \dot{u}_n$. Now the favorable phenomena which is desired is that the solutions $\dot{u}_1, \dots, \dot{u}_n$

of the above system should approach zero asymptotically as $\tau \to \infty$ and that the solutions be near zero after not too large a τ interval. Now suppose that λ is quite small. For the original variable t, this means that $\dot{u}_j$ becomes small after a small t interval and from then on one has a situation corresponding to small $u_1, \ldots, u_n$ and small $\dot{u}_1, \ldots, \dot{u}_r$. This process of examining the result of passing to the variable τ is basic in any discussion of λ errors and may be effectively used in case 1 as well.

In the above discussion, we have utilized only one parameter λ. If we have more than one type of integrator in the system, λ will be associated with the type which has the greatest time delay and other time delays will be represented by a λ where a is small.

4.9 The Necessity for Interval by Interval Analysis

Given the differential equation

$$\lambda \ddot{y} + a(x)\dot{y} + b(x)y = 0 \tag{4.59}$$

with λ a parameter and $a(x)$, $b(x)$ analytic in the neighborhood of $x = 0$, it was hoped that the general solution of Eq. (4.59) could be written in the form

$$g(x, \lambda) = A \cdot X(x, \lambda) + B \cdot Y(x, \lambda)$$

where X and Y are analytic in x for a suitable interval and $Y(x,\lambda)$ is analytic in λ. Essentially this amounts to the requirement that Eq. (4.59) have one non-trival solution which is analytic in λ. This, unfortunately, is not the case in general, as is shown by the following example.

EXAMPLE: Consider the differential equation

$$\lambda \ddot{y} + \dot{y} + \frac{1}{1+x} y = 0. \tag{4.60}$$

If there exists a solution $Y(x, \lambda)$ analytic in λ,

$$Y(x,\lambda) = u_o(x) + \lambda u_1(x) + \lambda^2 u_2(x) + \ldots \tag{4.61}$$

which can be differentiated with respect to x, twice, term by term, i. e. such that

$$\dot{Y}(x,\lambda) = \dot{u}_o(x) + \lambda\dot{u}_1(x) + \lambda^2\dot{u}_2(x) + \ldots \qquad (4.62)$$

$$\ddot{Y}(x,\lambda) = \ddot{u}_o(x) + \lambda\ddot{u}_1(x) + \lambda^2\ddot{u}_2(x) + \ldots$$

and $Y(0,0) = 1$, then the λ radius of convergence of Eq. (4.61) for $x > 0$ is zero. (This is also true for $x < 0$.)

Proof: Substituting Eq. (4.62) in Eq. (4.60) we obtain the following set of differential systems:

$$\dot{u}_o + \frac{1}{1+x} u_o = 0 \qquad u_o(0) = 1 \qquad (4.63)$$

$$\dot{u}_n + \frac{1}{1+x} u_n = -\ddot{u}_{n-1}, \quad u_n(0) = 0, \; n = 1, 2, \ldots \; . \qquad (4.64)$$

By a direct calculation,

$$u_o(x) = \frac{1!}{x+1}$$

$$u_1(x) = \frac{2!}{(1+x)^2} - \frac{2}{1+x}$$

$$u_2(x) = \frac{3!}{(1+x)^3} - \frac{4}{(1+x)^2} - \frac{2}{1+x} \; .$$

We shall now prove

(a) $$u_n(x) = \frac{(n+1)!}{(1+x)^{n+1}} - \sum_{k=1}^{n} \frac{a_k^{\,n}}{(1+x)^k}$$

(b) $$\sum_{k=1}^{n} a_k^{\,n} = (n+1)!$$

(c) $$a_k^{\,n} > 0 \qquad k = 1, 2, \ldots, n.$$

The proof is by complete induction on n. Assume (a), (b), (c). Then substituting in Eq. (4.64):

$$\dot{u}_{n+1} + \frac{1}{1+x} u_{n+1} = - \frac{(n+1)\,(n+2)!}{(1+x)^{n+3}} + \sum_{k=1}^{n} \frac{k(k+1)}{(1+x)^{k+2}} a_k^{n} \quad .$$

Assume

$$u_{n+1}(x) = \frac{a}{(1+x)^{n+2}} - \sum_{k=1}^{n+1} \frac{a_k^{n+1}}{(1+x)^{k}} .$$

Then

$$\dot{u}_{n+1} + \frac{1}{1+x} u_{n+1} = - \frac{(n+2)a}{(1+x)^{n+3}} + \sum_{k=1}^{n+1} \frac{k a_k^{n+1}}{(1+x)^{k+1}} + \frac{a}{(1+x)^{n+3}} - \sum_{k=1}^{n+1} \frac{a_k^{n+1}}{(1+x)^{k+1}}$$

$$= \frac{-(n+1)a}{(1+x)^{n+3}} + \sum_{k=1}^{n+1} \frac{(k-1)a_k^{n+1}}{(1+x)^{k+1}}$$

$$= - \frac{(n+1)\,(n+2)!}{(1+x)^{n+3}} + \sum_{k=1}^{n} \frac{a_k^{n} k(k+1)}{(1+x)^{k+2}} \quad .$$

Equate coefficients of $\frac{1}{(1+x)^k}$:

$$-(n+1)a = -(n+1)\,(n+2)!$$

$$(k-1)a_k^{n+1} = (k-1)ka_{k-1}^{n} \qquad k = 2, \ldots, n+1$$

$$a_k^{n+1} = ka_{k-1}^{n} \qquad k = 2, 3, \ldots, n+1$$

$$a = (n+2)!$$

$$a_k^{n+1} > 0 \qquad k = 2, 3, \ldots, n+1$$

Let $\quad a_1^{n+1} = a - \sum_{k=2}^{n+1} a_k^{n+1} \quad .$

Then if we show

$$a_1^{n+1} > 0,$$

conditions (a), (b), (c) will be satisfied. Now,

$$a_k^{n+1} \leqq (n+1)a_{k-1}^n \qquad k = 2, 3, \ldots, n+1$$

$$\sum_{k=2}^{n+1} a_k^{n+1} < (n+1)\sum_{k=2}^{n+1} a_{k-1}^n \qquad (n+1)\left(\sum_{l=1}^{n} a_l^n\right)$$

$$= (n+1)(n+1)!$$

$$a_1^{n+1} > a - (n+1)(n+1)! = (n+2)! - (n+1)(n+1)!$$

$$= (n+1)! > 0.$$

Now, suppose $x > 0$. Then since a_k^n is positive

$$u_n(x) = (1+x)^{-n-1}\left[(n+1)! - \sum_{k=1}^{n} (1+x)^{n+1-k} a_k^n\right]$$

$$< (1+x)^{-n-1}\left[(n+1)! - (1+x)\sum_{k=1}^{n} a_k^n\right]$$

$$= -x(1+x)^{-n-1}(n+1)! .$$

Thus for $x > 0$, $u_n(x)$ is more negative than

$$-x \frac{(n+1)!}{(1+x)^{n+1}}$$

and the series

$$\sum_{n=0}^{\infty} u_n(x) \lambda^n$$

has zero radius of convergence in λ . (A similar argument holds for $x < 0$.)

This example indicates that one cannot hope to break down the solutions of the linear variational equation into, say, an n-dimensional set of solutions analytic in λ at $\lambda = 0$ and an r-dimensional set which is nonanalytic. The set of analytic solutions may have dimension less than n, even zero.

Thus, in order to obtain the long range part of the error, which in the more desirable cases contains the greater part of the error, one must use some procedure in addition to solving the variational equations, for instance, the process given above of breaking the interval into smaller intervals on which the coefficients may be considered to be constant. In addition, the latter process does yield a clear picture of the phenomenon and even in the case when λ errors do not occur it may correspond to a desirable numerical procedure.

5. EXAMPLES

In this chapter four illustrative examples are given. These examples indicate many possibilities for further developments.

5.1 Example of an α Error

As an example of an α error we briefly treat the equation

$$\dot{y} = -y^2 \tag{5.1}$$

assuming that y^2 is obtained by means of a multiplier which unfortunately has certain rather common inaccuracies. One factor can be taken as the correct y. The other factor can be described mathematically as follows.

Let z be a certain time delayed value of y, i.e., $z = y - \tau\dot{y}$. The second factor is obtained from z by means of a step-like function, i.e., let $[x]$ denote as usual the largest integer not exceeding x. Then there is a large integer N such that the remaining factor is

$$\frac{1}{N}\,[zN]. \tag{5.2}$$

Let $f(x)$ be defined by

$$\frac{1}{N}\,[zN] = z + f(z). \tag{5.3}$$

Thus, the equation as actually realized is

$$\dot{y} = -[z+f(z)]y = -[y - \tau\dot{y} + f(y - \tau\dot{y})]y. \tag{5.4}$$

When one introduces α this becomes

$$\dot{y} = -y^2 + \alpha[\tau\,\dot{y}y - y\,f(y - \tau\,\dot{y})]. \tag{5.5}$$

The solution of Eq. (5.5) depends on two variables, t and α, and will be denoted by $y(t, \alpha)$.

We take as the initial value for Eq. (5.1), $y(0) = y_o$ and as the initial value for Eq. (5.5), $y(0, \alpha) = y_o$.

$f(z)$ as defined by Eq. (5.3) is a sawtooth function with slope -1 which ranges from 0 to $-1/N$ as z changes from k/N to just below $(k+1)/N$. Therefore, as Eq. (5.5) stands it violates the analyticity requirements for G as given in § 2.2 of Chapter 2. However, as a practical matter one would be willing to substitute $f(y(t,0) - \gamma\dot{y}(t,0))$ for $f(z)$, i.e., ignore the dependence of z on α in regard to $f(z)$. Let

$$g(t) = f(y(t,0) - \gamma\dot{y}(t,0)) = f(y_o(y_ot+1)^{-1} + \gamma y_o^{\,2}(y_ot+1)^{-2}) . \tag{5.6}$$

Here one has used the fact that if $y(0,0) = y_o$, then

$$y(t,0) = y_o[y_ot+1]^{-1} . \tag{5.7}$$

Thus, the form of G to which the error analysis will be applied is

$$\dot{y} = -y^2 + \alpha[\gamma\dot{y}y - y\,g(t)]. \tag{5.8}$$

The solution we are concerned with is

$$y(t,1) = y(t,0) + y_\alpha(t,0) + \frac{1}{2}y_{\alpha\alpha}(t,0) + \ldots \tag{5.9}$$

$$= y(t,0) + z(t) + \frac{1}{2}w\,(t) + \ldots$$

where

$$z(t) = y_\alpha(t,0), \qquad w(t) = y_{\alpha\alpha}(t,0). \tag{5.10}$$

If we differentiate Eq. (5.8) relative to α, we obtain

$$\frac{\partial\dot{y}}{\partial\alpha} = -2y\frac{\partial y}{\partial\alpha} + \gamma\dot{y}y - y\,g(t) + \alpha[\gamma\frac{\partial\dot{y}}{\partial\alpha}y + \gamma\dot{y}\frac{\partial y}{\partial\alpha} - \frac{\partial y}{\partial\alpha}g(t)]. \tag{5.11}$$

Setting $\alpha = 0$ yields

$$\dot{z} = -2yz + \gamma\dot{y}y - yg. \tag{5.12}$$

Differentiating Eq. (5.11) again and setting $\alpha = 0$ yields

$$\dot{w} = -2yw - 2z^2 + 2[\gamma(\dot{z}y + \dot{y}z) - zg(t)] . \tag{5.13}$$

It is desirable that in the region of interest z be small and that w be small relative to z. The purpose of the present analysis is to determine the range of t for which this is true. It is convenient for this purpose to eliminate $\dot{y}$ and $\dot{z}$ from Eq. (5.13) by using Eqs. (5.8) (at $\alpha=0$) and (5.12). Thus

$$\dot{w} + 2yw = -2[z^2 + (3y^2\gamma + g)z + y^2\gamma(\gamma y^2 + g)] . \tag{5.14}$$

To solve Eq. (5.14) and

$$\dot{z} + 2yz = - r y^3 - yg = -y(r y^2 + g), \tag{5.15}$$

we use the integrating factor $(y_o t + 1)^2 = y_o^2 y^{-2}$. This replaces the elaborate Green's function which is necessary in the general case.

$$(y_o t + 1)^2 z = - \int_0^t y_o(y_o x+1) (r y^2 + g)dx \tag{5.16}$$

or

$$z = -y^2 y_o^{-1} \int_0^t (y_o x + 1) (r y^2 + g)dx \tag{5.17}$$

$$= -y^2 y_o^{-1} \int_0^t r y_o^2 (y_o x + 1)^{-1} dx - y^2 y_o^{-1} \int_0^t (y_o x + 1)g(x)dx.$$

It is reasonable to replace $g(x)$ in the last integral by its average value $-(2N)^{-1}$. This yields

$$z = - r y^2 \log(y_o t + 1) + \frac{1}{4N} \left[1 - \left(\frac{y}{y_o}\right)^2 \right] \tag{5.18}$$

or

$$z = r y^2 \log \frac{y}{y_o} + \frac{1}{4N} \left[1 - \left(\frac{y}{y_o}\right)^2 \right] . \tag{5.19}$$

Now $y = y_o(y_o t + 1)^{-1}$. For y_o positive, the solution can be considered for all positive values of t. Since $y = y_o/(y_o t + 1)$, y approaches zero as $t \to \infty$ and z approaches the value $\frac{1}{4N}$. The first term is negative and has maximum absolute value $\frac{1}{2} r y_o^2 e^{-1}$. Thus z lies between $-\frac{1}{2} r y_o^2 e^{-1}$ and $\frac{1}{4N}$,

$$-\frac{1}{2} r y_o^2 e^{-1} \leq z \leq \frac{1}{4N} \qquad y_o > 0. \tag{5.20}$$

Presumably both bounds are small.

For y_o negative, the situation would be complicated in most cases by scaling difficulties. The problem can extend only until $-y$ has attained its maximum scale value which must be greater than $-y_o$. Consequently, here again the terms differ in sign. z can be written

$$z = y^2 \left[r \log \left(\frac{y}{y_o}\right) \right] - \frac{1}{4Ny_o^2} + \frac{1}{4N} . \qquad (5.21)$$

If the maximum value of $-y$ is 1 and $r = -1/y_o$ we see that the final value of z is

$$z = (r \log r \cdot - \frac{r^2}{4N}) + \frac{1}{4N} . \qquad (5.22)$$

Thus, y_o must be limited if z is to be kept small. The procedure is straightforward when r and N are known.

We next consider w, the second order error effect, and we show that it the time interval is not very large then w is small compared with z. However, if t is very large, w will become arbitrarily large. From Eq. (5.14) we may write, with the aid of the integrating factor $\left(\frac{y_o}{y}\right)^2$,

$$w = \frac{-2}{(y_o t+1)^2} \int_0^t \left(\frac{y_o}{y}\right)^2 [z^2 + gz + y^2(3 r z + r^2 y^2 + r g)]\, dx. \qquad (5.23)$$

Now w can be written

$$w = A r^2 + 2B r \left(\frac{1}{4N}\right) + C \left(\frac{1}{4N}\right)^2 \qquad (5.24)$$

when we replace g by its average value $-\frac{1}{2N}$. By Eq. (5.18),

$$A = \frac{-2}{(y_o t+1)^2} \int_0^t \left\{ y_o^2 y^2 [\log(y_o x+1)-1]^2 - y_o^2 y^2 \log(y_o x+1) \right\} dx. \qquad (5.25)$$

We let $\zeta = y_o x+1$ and integrate by parts obtaining

$$A = \frac{-2y_o^3}{(y_o t+1)^2} \int_1^{y_o t+1} \frac{1}{\zeta 2} \left\{ [\log \zeta -1]^2 - \log \zeta \right\} d\zeta$$

$$= \frac{2y_o^3}{(y_o t+1)^3} [\log^2(y_o t+1) - \log(y_o t+1)] . \qquad (5.26)$$

Similarly

$$B = \frac{-2}{(y_o t+1)^2} \int_0^t y^2 \left[\log(y_o x+1) - 1 + y_o x + \frac{y_o^2 x^2}{2}\right] dx \tag{5.27}$$

$$= \frac{-2y_o}{(y_o t+1)^2} \int_1^{y_o t+1} \frac{1}{\zeta^2}\left(\log \zeta + \frac{1}{2}\zeta^2 - \frac{3}{2}\right) d\zeta$$

$$= \frac{2y_o}{(y_o t+1)^3} \left[\log(y_o t+1) - \frac{1}{2} y_o^2 t^2\right] .$$

And

$$C = \frac{2}{(y_o t+1)^2} \int_0^t \left[\left(\frac{y_o}{y}\right)^2 - \left(\frac{y}{y_o}\right)^2 \right] dx \tag{5.28}$$

$$= \frac{2}{y_o(y_o t+1)^2} \left[\frac{1}{3}(y_o t+1)^3 + (y_o t+1)^{-1} - \frac{4}{3}\right] .$$

For y_o positive, $| A |$ and B are zero initially, remain less than a relatively small maximum, and then approach zero. On the other hand, C is like $\frac{2}{3}t$ and we see that w will become of finite size when t is of the order of $16N^2$. This will presumably be adequate range.

On the other hand, for y_o negative, if we let

$$\frac{y}{y_o} = \frac{1}{y_o t+1} = \rho$$

we must suppose that we have a range in which $\rho \leq r$ where r is some predetermined value. Thus, if we assume $| y | \leq 1$ we may interpret Eq. (5.21) as conditions that

$$r \log r \ll 1, \quad r^2 \ll 4N. \tag{5.29}$$

The first term and Eq. (5.26) will imply that in general the term $A\tau^2$ will be small. (The t range will be limited if y_o is negative.) Similarly, if we look at the dominant terms in B and C, we see that Eq. (5.29) is adequate to yield that both $B\tau/4N$ and $C/(4N)^2$ are small.

5.2 Example of β Errors

We shall give here an example of a differential system involving β errors. For simplicity we shall assume a first order equation in which no α or λ errors are present. The equation we shall consider is

$$\dot{x} + \frac{t}{t+1} x^2 - 2x + \frac{t+2}{t+1} = 0 \tag{5.30}$$

with the initial condition

$$x(0) = 1.$$

Suppose that the integrator also functions as a noise generator. A noise generator whose output is "white" may be described by the "shot effect" perturbations described in § 3.4. For an individual integrator we may measure the output of this noise generator by supposing that we are integrating the equation

$$\dot{x} = 0, \tag{5.31}$$

say with $x_o = 0$.

Under these circumstances the output is the chance variable $n_e{}^1$ of Eq. (3.21) of § 3.4. This is a normally distributed variable whose variance is given by Eq. (3.22) of § 3.4. For one variable Eq. (3.22) becomes

$$\sigma^2 = \sigma_o{}^2 n_o \int_{t_o}^{t} [Y(t,\zeta)]^2 d\zeta . \tag{5.32}$$

For Eq. (5.31), $Y(t,\zeta)$ is readily seen to be 1 for $\zeta \leq t$ and consequently the mean square value of the output is

$$s^2(t) = \sigma_o{}^2 n_o (t - t_o). \tag{5.33}$$

If $t = t_o + T$, $\frac{s^2(t)}{T} = \sigma_o^2 n_o$; and, if the output is a voltage developed across a pure resistive load, R, then the noise power of the output of the integrator is

$$\frac{\sigma_o^2 n_o}{R} .$$

Now let us return to Eq. (5.30). The unique solution is $x = 1$ for the initial value $x(0) = 1$. On the other hand, owing to the noise present the actual output will be given by a series

$$x = 1 + n_e^1 + n_e^2 + \ldots \tag{5.34}$$

where n_e^1 is given by Eq. (3.17) or Eq. (3.21) of § 3.4 and n_e^2 by Eq. (3.23) of § 3.4.

Here again n_e^1 is a normally distributed variable whose variance is given by Eq. (5.32). However, to compute $Y(t, \zeta)$ we must find the solution of

$$\dot{y} - \frac{2}{t+1} y = 0 \tag{5.35}$$

which for $t = \zeta$ has the value 1. [Equation (5.35) is the equivalent of Eq. (3.5) in the present case and may be obtained as follows. Suppose the solution x of Eq. (5.30) depends on the variable β and take the partial derivative of both sides of Eq. (5.30) relative to β. The result is

$$\dot{y} + \left(\frac{2tx}{1+t} \quad -2\right) y = 0 \tag{5.36}$$

where $y = \frac{\partial x}{\partial \beta}$. If one sets $x = 1$ in Eq. (5.36), one obtains Eq. (5.35).]

From Eq. (5.35) cne readily infers that

$$Y(t, \zeta) = \frac{(t+1)^2}{(\zeta+1)^2} . \tag{5.37}$$

Thus, Eq. (5.32) and Eq. (5.37) yield for the variance of the chance variable n_e^1

$$\sigma^2 = \sigma_o^2 n_o \frac{1}{3} [(1+t)^4 - (1+t)]. \tag{5.38}$$

We can also use the discussion of § 3.4 to find the expected value of ${}^{n}e^{2}$, [Eq. (3.32)]. In the case of just one dependent variable, we may evaluate Eq. (3.32) simply by letting $u = i = k = l = h = 1$

$$E[{}^{n}e^{2}] = -\frac{1}{2}\sigma_o{}^2 n_o \int_{t_o}^{t}\int_{\tau}^{t} J^{-1}J_1{}^1[Y(\zeta,\tau)]^2 Y(t,\zeta) \quad \Gamma_1{}^{1,1}(\zeta)d\zeta\, d\tau . \qquad (5.39)$$

From Eqs. (3.28) and (3.29) one has

$$\Gamma_1{}^{1,1} = \frac{\partial^2 F_1}{\partial x^2} - 2J^{-1}\frac{\partial^2 F_1}{\partial \dot{x}\,\partial x}(K_1{}^1) - J^{-2}\frac{\partial^2 F}{\partial \dot{x}^2}(K_1{}^1)^2. \qquad (5.40)$$

One can verify by means of the discussion given before Eq. (3.10) that for Eq. (5.29), $J = 1$, $J_1{}^1 = 1$, $K_1{}^1 = 1$. Since Eq. (5.30) is linear in $\dot{x}$, the second partials of F relative to $\dot{x}$ are zero and

$$\Gamma(\zeta) = \frac{2\zeta}{(\zeta+1)} \quad . \qquad (5.41)$$

Thus, Eq. (5.39) becomes

$$E[{}^{n}e^{2}] = -\frac{1}{2}\sigma_o{}^2 n_o \int_0^t \int_{\tau}^{t} \frac{(t+1)^2}{(\tau+1)^4} \; 2\zeta\,(\zeta+1)d\zeta\, d\tau$$

$$= \frac{1}{3}\sigma_o{}^2 n_o(t+1)\,[\,-\frac{1}{3}(t+1)^4 + \frac{1}{2}(t+1)^3 - \frac{7}{6}(t+1) + 1$$

$$+ (t+1)\log(t+1)\,].$$

5.3 Mechanical Differential Analyzer

Consider a differential analyzer constructed by means of disk integrators, gears and differentials and certain suitable servo systems which eliminate load from the disk integrators and their inputs. The independent variable t will be obtained from a uniformly rotating shaft. Thus, if τ denotes real time

$$\frac{dt}{d\tau} = r \qquad (5.42)$$

where r is a constant.

A servo system has an input and an output which we denote by Z and z respectively. Let T_Z denote the torque exerted by the input and T_z the torque exerted by the output of the servo system. It is possible, in general, to replace T_Z by zero and we shall do so in our future discussions.

The output torque is proportional to the input signal and the latter in turn is proportional to $Z - z$,

$$T_z = k^{-1} (Z - z) \tag{5.43}$$

or

$$z = Z - kT_z. \tag{5.44}$$

For the present discussion, we shall suppose that an "integrator" consists of a disk integrator and three servo systems--one on the output and one on each of the two inputs, (cf. Fig. 1). If X denotes the output of the disk integrator and

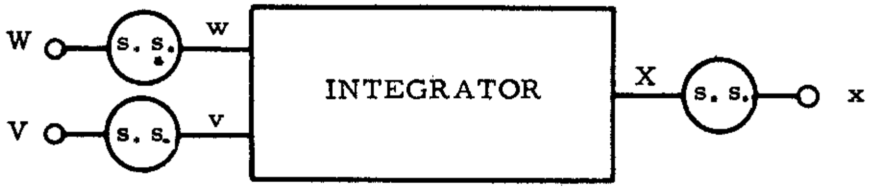

Figure 1

w and v the two inputs, the desired relation between these is

$$\dot{X} = w\dot{v}. \tag{5.45}$$

The output X of the disk integrator is the input of a servo system with output x. Thus, Eq. (5.44) yields

$$x = X - kT_x. \tag{5.46}$$

On the basis of the usual assumptions that the friction is proportional to $r\dot{x}$ and that the inertia load is proportional to $r^2\ddot{x}$,

$$T_x + k_1 r\dot{x} + k_2 r^2 \ddot{x}. \tag{5.47}$$

As long as the integrators do not load their inputs, it is reasonable to assume that k_1 and k_2 will be constant for a given problem. They will depend on the number of gears and differentials driven by the output of the servo system. We have postulated three servo systems to permit this assumption. If servo systems are not used on the inputs of the disk integrators, T_x may be a much more complex function, although it is believed that in various individual problems it should be possible to obtain it.

We can combine Eqs. (5.45), (5.46) and (5.47) to yield

$$\dot{x} + d_1\ddot{x} + d_2\dddot{x} = \dot{wv} \tag{5.48}$$

where

$$\left.\begin{aligned} d_1 &= kk_1 r \\ d_2 &= kk_2 r^2. \end{aligned}\right\} \tag{5.49}$$

Now w and v are also outputs of servo systems with inputs W and V respectively. The torque T_w is proportional to $\dot{w}$,

$$T_w = k_3 r\dot{w}. \tag{5.50}$$

From Eq. (5.43) we conclude

$$w = W - d_3\dot{w} \tag{5.51}$$

where

$$d_3 = kk_3 r.$$

It is consistent with the usual practices in error analysis to replace $\dot{w}$ by $\dot{W}$ in Eq. (5.51). We do this now but we shall return to this point at the end of this section. Thus

$$w = W - d_3\dot{W}. \tag{5.52}$$

Similarly for v,

$$v = V - d_4 \dot{V}. \tag{5.53}$$

Equations (5.48), (5.52) and (5.53) may be combined to yield

$$\dot{x} + d_1 \ddot{x} + d_2 \dddot{x} = (W - d_3 \dot{W}) (\dot{V} - d_4 \ddot{V}). \tag{5.54}$$

We now briefly discuss the applications of the above to a specific set-up of the differential analyzer. We suppose this involves n integrators in the above sense with outputs $x_1, \ldots, x_n$. For each output we obtain the equivalent of Eq. (5.54),

$$\dot{x}_j + d_1 \ddot{x}_j + d_2 \dddot{x}_j = (Wj - d_3 \dot{W}_j) (\dot{V}_j - d_4 \ddot{V}_j). \tag{5.55}$$

The W_j and V_j depend linearly on the x's. Thus

$$W_j = a_{oj} + \sum_k a_{jk} x_k \tag{5.56}$$

$$V_j = b_{oj} t + \sum_l b_{jl} x_l. \tag{5.57}$$

Equation (5.55) represents the expanded set of G equations used in Chapter 4, which replaces the correct relation

$$\dot{x}_j = W_j \dot{V}_j \ . \tag{5.58}$$

Now suppose $x'_1, \ldots, x'_n$ is a correct solution of Eqs. (5.56), (5.57) and (5.58). Suppose the solution of Eqs. (5.55), (5.56) and (5.57) is in the form

$$x'_1 + u'_1, \ldots, x'_n + u_n.$$

Substituting in Eqs. (5.56) we obtain

$$W_j = W'_j + \delta W_j \tag{5.59}$$

where

$$W'_j = a_{oj} + \sum_k a_{jk} x'_k , \qquad \delta W_j = \sum_k a_{jk} u_k .$$

Similarly,

$$\dot{V}_j = \dot{V}'_j + \delta \dot{V}_j \tag{5.60}$$

where

$$\dot{V}'_j + b_{oj} + \sum_l b_{jl} \dot{x}'_l , \qquad \delta \dot{V}_j = \sum_l b_{jl} \dot{u}_l .$$

Equation (5.55) thus becomes

$$\begin{aligned} &\dddot{u}_j[d_2] \\ &+ \ddot{u}_j[d_1] + \delta \ddot{V}_j[d_4 W'_j - d_3 d_4 \dot{W}'_j] \\ &+ \dot{u}_j - \delta \dot{V}_j[W'_j - d_3 \dot{W}'_j] + \delta \dot{W}_j[d_3 \dot{V}'_j - d_3 d_4 \ddot{V}'_j] \\ &- \delta W_j[\dot{V}'_j - d_4 \ddot{V}'_j] + R_j + S_j = 0. \end{aligned} \tag{5.61}$$

Here we have collected the terms which are linear in u_j and its derivatives; R_j involves terms of higher degree than one in u_j and its derivatives; and S_j does not invlove u_j or its derivatives at all,

$$R_j = - \delta W_j(\delta \dot{V}_j - d_4 \delta \ddot{V}_j) + \delta \dot{W}_j(d_3 \delta \dot{V}_j - d_3 d_4 \delta \ddot{V}_j) \tag{5.62}$$

$$S_j = d_1 \ddot{x}'_j + d_2 \dddot{x}'_j + d_4 W'_j \ddot{V}'_j + d_3 \dot{W}'_j \dot{V}'_j - d_3 d_4 \dot{W}'_j \ddot{V}'_j . \tag{5.63}$$

Now let $*$ denote the operation of replacing W_j', $\dot{W}_j'$, etc. by suitably chosen constants. The equivalent of Eq. (4.17) of Chapter 4 (that is, the homogeneous part of Eq. (5.61)) is

$$\begin{aligned}
&\lambda^2 \dddot{u}_j[d_2] \\
&+ \lambda[\ddot{u}_j d_1 + \delta \ddot{V}_j(d_4 W_j'* - d_3 d_4 \dot{W}_j'*)] \\
&+ [\dot{u}_j - \delta \dot{V}_j(W_j'* - d_3 \dot{W}_j'*) + \delta \dot{W}_j(d_3 \dot{V}_j'* - d_3 d_4 \ddot{V}_j'*)] \\
&+ [-\delta W_j(\dot{V}_j'* - d_4 \ddot{V}_j'*)] = 0,
\end{aligned} \tag{5.64}$$

although here the effective value of λ is one.

The μ_o's correspond to the modes of

$$\begin{aligned}
&[\dot{u}_j - \delta \dot{V}_j(W_j'* - d_3 \dot{W}_j'*) + \delta \dot{W}_j(d_3 \dot{V}_j'* - d_3 d_4 \ddot{V}_j'*)] \\
&\quad + [-\delta W_j(\dot{V}_j'* - d_4 \ddot{V}_j'*)] = 0
\end{aligned} \tag{5.65}$$

while the ν_o's correspond to the modes of

$$\begin{aligned}
&\dddot{u}_j[d_2] \\
&+ [\ddot{u}_j d_1 + \delta \ddot{V}_j(d_4 W_j'* - d_3 d_4 \dot{W}_j'*)] \\
&+ [\dot{u}_j - \delta \dot{V}_j(W_j'* - d_3 \dot{W}_j'*) + \delta \dot{W}_j(d_3 \dot{V}_j'* - d_3 d_4 \ddot{V}_j'*)] = 0 .
\end{aligned} \tag{5.66}$$

The μ_o's and ν_o's are to be extended to $\lambda = 1$.

This is as far as one can go with a general theory. Further developments must involve assumptions on the relative sizes of the quantities introduced above. However, one might also attempt to carry through the above arguments in the case where one does not assume servo amplifiers on the inputs of the integrators. The output torque T_x would then contain terms corresponding to the various integrators which receive x_j as part of W_i or V_i.

The above discussion has been considerably simplified by the use of Eqs. (5.52) and (5.53) instead of

$$w + d_3\dot{w} = W \qquad (5.67)$$

$$v + d_4\dot{v} = V \qquad (5.68)$$

respectively. We believe the substitution of $\dot{W}$ for $\dot{w}$ and $\dot{V}$ for $\dot{v}$ in these equations would be the customary practice in error analysis and is justified. (Second approximations could also be used.) On the other hand it is possible to proceed using Eqs. (5.67) and (5.68) even though the result is far more complicated. It may be of interest to compare the relative complexity of the two procedures, and for this reason we indicate the process by which w and v can be eliminated between Eqs. (5.67), (5.68) and (5.48). We write the last $L = w\dot{v}$ where $L = \dot{x} + d_1\ddot{x} + d_2\dddot{x}$.

For symmetry in our treatment, we differentiate Eq. (5.68) and let $\dot{v} = s$, $\dot{V} = S$. Thus our system of equations becomes

$$w + d_3\dot{w} = W \qquad (5.67)$$

$$s + d_4\dot{s} = S \qquad (5.69)$$

$$L = ws \qquad (5.70)$$

and we want to eliminate the w and s.

Suppose that $0 < d_4 \leq d_3$ and $d_4/d_3 = \rho$. Then, if we differentiate $L = ws$ and multiply by d_4

$$\begin{aligned} d_4\dot{L} &= d_4\dot{w}s + d_4\dot{s}w \\ &= \rho(W - w)s + w(S - s) \qquad (5.71) \\ &= -(1 + \rho)ws + \rho Ws + wS \\ &= -(1 + \rho)L + \rho Ws + wS. \end{aligned}$$

Let

$$M = (1 + \rho)L + d_4\dot{L}. \qquad (5.72)$$

We then have the equations

$$\begin{aligned} M &= wS + \rho Ws \\ L &= ws \end{aligned} \tag{5.73}$$

from which we can eliminate s and obtain

$$Sw^2 - wM + \rho WL = 0. \tag{5.74}$$

We can now eliminate w by using $w + d_3\dot{w} = W$. For differentiating Eq. (5.74) and eliminating $\dot{w}$ yields

$$(2S - d_3\dot{S})w^2 - (2SW + M - d_3\dot{M})w + MW - d_3\,\rho(\dot{W}L + \dot{L}W) = 0. \tag{5.75}$$

Thus we have two quadratic equations in w and the latter can be eliminated. The resulting equation is nonlinear and of the second order in L, which means that it is of the fifth order in x. While one could theoretically apply the methods of this paper to this situation, it is difficult to conceive of a situation in which this more complicated procedure would be of practical interest. If a more accurate analysis is desired the use of higher order approximations in Eqs. (5.67) and (5.68) would be preferred.

The authors have also considered the case in which torque amplifiers are used instead of servo systems. The results are similar but more complicated.

5.4 Sensitivity in the Constant Coefficient Case

Because of the great practical importance of linear differential equations with constant coefficients we shall consider certain phenomena which arise in such systems. The constant coefficient case has been treated by other authors (cf. : Brock and Murray, Macnee, Raymond, loc. cit.). In the examples which follow we shall give a discussion which can be generalized to n th order systems. The complete analysis will be given in a future paper.

We shall consider the effects of α errors in the coefficients. Normally the original differential equations are analytic in these parameters. If this be the case, the usual existence theorems show that the solutions are also analytic in the α, and hence we may differentiate the solutions relative to α to obtain the coefficients of ϵ_j^i error terms (see Eq. (3.35)). However, while the solutions are analytic in α, the characteristic roots need not be. Consider the simple example

$$\begin{aligned} \dot{x} &= -3x + (1 - \alpha)y \\ \dot{y} &= -x - y \end{aligned} \tag{5.76}$$

where the "α" that appears is an α error in our usual definition. The characteristic roots of Eq. (5.76) are

$$\lambda_1 = -2 + \sqrt{\alpha} \qquad , \qquad \lambda_2 = -2 - \sqrt{\alpha}$$

and are clearly not analytic in α at $\alpha = 0$. The analyticity of the solutions

$$x(t) = x_o e^{-2t}(\cosh \sqrt{\alpha}\, t - \frac{1}{\sqrt{\alpha}} \sinh \sqrt{\alpha}\, t) + y_o e^{-2t} \frac{1-\alpha}{\sqrt{\alpha}} \sinh \sqrt{\alpha}\, t$$

$$y(t) = -x_o e^{-2t}(\frac{1}{\sqrt{\alpha}} \sinh \sqrt{\alpha}\, t) + y_o e^{-2t}(\cosh \sqrt{\alpha}\, t + \frac{1}{\sqrt{\alpha}} \sinh \sqrt{\alpha}\, t)$$

is guaranteed by our existence theorems.

Suppose we have a linear differential equation with constant coefficients which has a characteristic root λ of multiplicity m. Then if we consider the same linear differential equation with a non-homogeneous term of the form $e^{\lambda t}$, the term in the solution corresponding to this non-homogeneous term is of the form $t^m e^{\lambda t}$. In analogy with physical systems we shall sometimes call "characteristic roots" by the name "characteristic frequencies" and call the non-homogeneous term a "forcing function". We shall refer to the phenomena that occur when a forcing function $f(t)$ contains a characteristic frequency as resonance, or say that the function $f(t)$ resonates (with respect to the given differential equation).

Now we will show that when a system with constant coefficients

$$\dot{x}_i = \sum_j a_{ij}(\alpha) x_j \tag{5.77}$$

contains an error parameter α in the coefficients, the final error involves a resonance. For, let

$$y_i = \frac{\partial x_i}{\partial \alpha} . \tag{5.78}$$

Then y_i satisfies

$$\dot{y}_i = \sum_j a_{ij}(\alpha)\, y_j + \sum_j \frac{\partial a_{ij}}{\partial \alpha} x_j . \tag{5.79}$$

Since x_j is a solution of Eq. (5.77), the forcing term

$$\sum_j \frac{\partial a_{ij}}{\partial \alpha} x_j$$

of Eq. (5.79) is a linear combination of exponentials $t^r e^{\lambda t}$ where λ is a characteristic root of the system of Eq. (5.77) with multiplicity exceeding r. Since the homogeneous portions of Eq. (5.79) are the same as those of Eq. (5.77), the y_i will contain resonance terms. In particular if λ is of multiplicity m, y_i may contain terms of the form $At^{2m-1}e^{\lambda t}$ or similar terms involving lower powers of t.

How badly an error resonates depends not only on the multiplicity of a characteristic root but also on the Jordan normal form of the matrix of the system. Suppose we have the system of Eq. (5.77) which we write in vector form as

$$\dot{x} = Ax \tag{5.80}$$

where $\dot{x}$ and x are column vectors and A is the coefficient matrix. Then we know there exists a non-singular square matrix M such that $B = MAM^{-1}$ is in Jordan normal form.

Now let $y = MxM^{-1}$. Then applying M on the left and M^{-1} on the right to Eq. (5.80) we obtain

$$\dot{y} = M\dot{x}M^{-1} = MAxM^{-1} = MAM^{-1}MxM^{-1} = By. \tag{5.81}$$

Equation (5.81) consists of blocks of equations of the form

$$\begin{aligned}
\dot{y}_j &= \lambda y_j \\
\dot{y}_{j+1} &= y_j + \lambda y_{j+1} \\
&\vdots \\
\dot{y}_{j+k} &= y_{j+k-1} + \lambda y_{j+k} .
\end{aligned}$$

Such a block is solved in the form

$$y_j = A_j e^{\lambda t}$$

$$y_{j+1} = A_{j+1}e^{\lambda t} + A_j te^{\lambda t}$$

$$y_{j+2} = A_{j+2}e^{\lambda t} + A_{j+1}te^{\lambda t} + A_j \frac{t^2}{2!} \quad e^{\lambda t}$$

$$. \quad . \quad . \quad . \quad . \quad . \quad .$$

$$y_{j+k} = A_{j+k}e^{\lambda t} + A_{j+k-1}te^{\lambda t} + \dots + A_j \frac{t^k}{k!} \quad e^{\lambda t}$$

where the A_j, A_{j+1}, ..., A_{j+k} for the different blocks are independent constants of integration. Consequently, the highest power, m, of t associated with a given λ in the solution is one less than the maximum length of a block for this λ rather than the multiplicity of λ. Resonance is associated with the terms which are not in purely exponential form.

We illustrate the phenomenon of resonance with a simple example. Consider the system of differential equations

$$\begin{aligned} \dot{x} &= -k^2x + \alpha y \\ \dot{y} &= x - k^2y + \alpha z \\ \dot{z} &= \alpha y - k^2z \end{aligned} \tag{5.82}$$

where k is a real constant, α is an α parameter. Clearly, this system is analytic in α and has the normal form of Eq. (5.80) (with $\lambda = -k^2$ at $\alpha = 0$). The characteristic roots of Eq. (5.82) are $\lambda_1 = -k^2$; $\lambda_2 = -k^2 + \sqrt{2\alpha}$; $\lambda_3 = -k^2 - \sqrt{2\alpha}$, and three linearly independent solutions are

$$\varphi_1(t) = e^{-k^2t}$$

$$\varphi_2(t) = e^{-k^2t} \cosh \sqrt{2\alpha}\ t$$

$$\varphi_3(t) = e^{-k^2t} \sinh \sqrt{2\alpha}\ t.$$

The general solution of Eq. (5.82) for the boundary conditions $x(0) = x_o$, $y(0) = y_o$, $z(0) = z_o$ is:

$$x = x_o[\frac{1}{2} \varphi_1 + \frac{1}{2} \varphi_2] + y_o[\frac{\sqrt{\alpha}}{\sqrt{2}} \varphi_3] + z_o[- \frac{\alpha}{2} \varphi_1 + \frac{\alpha}{2} \varphi_2]$$

$$y = x_o[\frac{1}{\sqrt{2\alpha}} \varphi_3] + y_o[\varphi_2] + z_o[\frac{\sqrt{\alpha}}{\sqrt{2}} \varphi_3]$$

$$z = x_o[- \frac{1}{2\alpha} \varphi_1 + \frac{1}{2\alpha} \varphi_2] + y_o[\frac{1}{\sqrt{2\alpha}} \varphi_3] + z_o[\frac{1}{2} \varphi_1 + \frac{1}{2} \varphi_2].$$

Differentiating x, y, z with respect to α and then setting $\alpha = 0$ gives us the coefficients $\frac{\partial x}{\partial \alpha}$, $\frac{\partial y}{\partial \alpha}$, $\frac{\partial z}{\partial \alpha}$ in the ϵ error terms. We find in this example that

$$\left(\frac{\partial x}{\partial \alpha}\right)_{\alpha=0} = x_o[\frac{1}{2} t^2 e^{-k^2 t}] + y_o[t e^{-k^2 t}] + z_o[0]$$

$$\left(\frac{\partial y}{\partial \alpha}\right)_{\alpha=0} = x_o[\frac{t^3}{3} e^{-k^2 t}] + y_o[t^2 e^{-k^2 t}] + z_o[t e^{-k^2 t}]$$

$$\left(\frac{\partial z}{\partial \alpha}\right)_{\alpha=0} = x_o[\frac{-t^4}{12} e^{-k^2 t}] + y_o[\frac{t^3}{3} e^{-k^2 t}] + z_o[\frac{1}{2} t^2 e^{-k^2 t}]$$

and hence that except for certain special boundary conditions the error varies as the fourth power of the time.