ARO Report 81-1

TRANSACTIONS OF THE TWENTY-SIXTH

CONFERENCE OF ARMY MATHEMATICIANS

AD A093562

JAN 8 1981

A

DDC FILE COPY

81 1 08 063

U. S. ARMY RESEARCH OFFICE

Report No. 81-1

January 1981

TRANSACTIONS OF THE TWENTY-SIXTH CONFERENCE

OF ARMY MATHEMATICIANS

Sponsored by the Army Mathematics Steering Committee

Host

U. S. Army Cold Regions Research and
Engineering Laboratory

Hanover, New Hampshire

10-12 June 1980

FOREWORD

__ _ _ovember 1979, Colonel B. Devereaux, Jr., Commander and Director of the __ _ _,uy Cold Regions Research and Engineering Laboratory (CRREL) sent the following letter to Dr. Jagdish Chandra, Chairman of the Army Mathematics __t _ _g Committee (AMSC).

1. "We would like to extend an invitation that the 26th Conference of Army Mathematicians be hosted by the U. S. Army Cold Regions Research and Engineering Laboratory on 10-12 June 1980.

2. Mr. Benjamin S. Yamashita, Public Affairs Officer, will serve as CRREL's point of contact for administrative arrangements for the Conference.

3. We look forward to having the group meet here."

Dr. Chandra was pleased to accept this invitation on behalf of the AMSC. This _ _ the second time that CRREL has served as host for one of those meetings. The Twelfth Conference of Army Mathematicians was held at CRREL, Hanover, New Hampshire on 22-23 June 1966. For that meeting, Dr. Shunsuke Takagi played much the same role as he did for the 1980 conference, namely he was responsible for scientific details regarding various phases of the program. The AMSC, the sponsors of these meetings, would like to take this occasion to thank Messrs. Takagi and Yamashita as well as all other members of CRREL for making this a very successful conference.

The theme of the 26th Conference was "Wave Propagation in Solids and Nondestructive Evaluation Techniques." In addition to the five invited speakers, there were more than thirty contributed papers presented at this meeting. While most of these papers were given by Army scientists, still there were a surprising number of papers, namely eight, that were delivered by university professors. The invited speakers and their topics are noted below.

Speaker and Affiliation	Title of Address
Professor Jan Achenbach Northwestern University	DIRECT AND INVERSE METHODS FOR SCATTERING BY CRACKS IN THE HIGH-FREQUENCY RANGE
Professor Constantine Dafermos Brown University	CAN DISSIPATION PREVENT THE BREAKING OF WAVES?
Professor Y. H. Pao Cornell University	THEORY OF ACOUSTIC EMISSION

iii

Professor James Tasi SHOCK WAVES AND LATTICE DYNAMICS
State University of New York
 at Stony Brook

Professor T. C. T. Ting WAVE PROPAGATION IN PERIODICALLY LAYERED
University of Illinois- MEDIA
 Chicago Circle

The members of the AMSC would like to express their thanks to the speakers
and research workers who participated in this meeting, and to all the attendees
for supporting it with their many stimulating questions. The AMSC is pleased
to be able to publish in these transactions many of the conference papers
and thus to make available to the scientific community some of the research
results presented at this meeting.

TABLE OF CONTENTS*

---

*This table of contents contains only the papers that are published in this
technical manual. For a list of all papers presented at the Twenty-Fifth
Conference of Army Mathematicians, see the Program of the meeting.

PROGRAM

26th CONFERENCE OF ARMY MATHEMATICIANS

Monday
9 June 1980

1800-2000   Registration   Gallagher's Tavern

Tuesday
10 June 1980

1800   Registration   Ballroom Foyer

0845   Opening Remarks   Ballroom East

0900   GENERAL SESSION I   Ballroom East

             CHAIRPERSON - Dr. Chunsuke Takasi, U. S. Army Cold Regions
                           Research and Engineering Laboratory, Hanover,
                           New Hampshire

             SPEAKER   - Professor Jan Achenbach, Northwestern
                            University, Evanston, Illinois

             TITLE   - DIRECT AND INVERSE METHODS FOR SCATTERING BY
                            CRACKS IN THE HIGH-FREQUENCY RANGE

1000   Break

1030   TECHNICAL SESSION I   Ballroom East

             CHAIRPERSON - Dr. Dennis Tracey, U. S. Army Materials and
                           Mechanics Research Center, Watertown,
                           Massachusetts

             SHORT WAVE REFLECTION IN INHOMOGENEOUS MATERIALS

             Professor Richard E. Meyer, Mathematics Research Center and
             University of Wisconsin, Madison, Wisconsin.

             SOLITARY WAVES AND SHOCK PROFILES IN THE THREE-DIMENSIONAL
             LATTICE

             Dr. John D. Powell, Ballistic Research Laboratory, Aberdeen
             Proving Ground, Maryland, and Dr. Jad H. Batteh, Science
             Applications, Inc., Atlanta, Georgia

             ELASTIC WAVE SCATTERING BY FLAWS FOR NDE APPLICATIONS

             Professors V. V. Varadan and V. K. Varadan, Ohio State
             University, Columbus, Ohio

NDE OF FLAWS IN COMPOSITE MATERIALS

Professors V. V. Varadan and V. K. Varadan, Ohio State
University, Columbus, Ohio

AN ENERGY CRITERION FOR FINITE ELASTICITY

Professor Reza Malek-Madani, Mathematics Research Center,
University of Wisconsin, Madison, Wisconsin

1030   TECHNICAL SESSION II   Ballroom West

CHAIRPERSON - Dr. John Polk, Ballistic Research Laboratory,
     Aberdeen Proving Ground, Maryland

A TRANSITION PHENOMENON IN THE THEORY OF COMBUSTION WAVES

Professor S. S. Lin, Mathematics Research Center, University
of Wisconsin, Madison, Wisconsin

MONOTONE METHOD AND STABILITY OF SOLUTIONS OF A SYSTEM OF
REACTION DIFFUSION EQUATIONS

Drs. Jagdish Chandra and Francis Dressel, U. S. Army Research
Office, Research Triangle Park, North Carolina, and Dr. Paul
Norman, Virginia Military Institute, Lexington, Virginia

DEVELOPMENT OF DEFLAGRATION IN INITIALLY COLD COMBUSTIBLES

Professor A. K. Kapila, Mathematics Research Center and
Rensselaer Polytechnic Institute, Troy, New York

A MATHEMATICAL QUESTION FROM COMBUSTION THEORY

Professors G. S. S. Ludford and D. S. Stewart, Cornell
University, Ithaca, New York

STABILITY OF REACTION DIFFUSION SYSTEMS UNDER RANDOM
PERTURBATION

Professor P. L. Chow, Wayne State University, Detroit,
Michigan.

1200   Lunch

1330   TECHNICAL SESSION III   Ballroom East

CHAIRPERSON - Dr. Julian Wu, Benet Weapons Laboratory,
     Watervliet Arsenal, Watervliet, New York

DIFFERENTIATION OF TABULAR DATA

Dr. Cesiocas Masaitis, Ballistic Research Laboratory,
Aberdeen Proving Ground, Maryland

COMPUTER AIDED ANALYSIS OF MECHANICAL SYSTEMS WITH
INTERMITTENT MOTIONS

Professor Edward J. Haug, Jr., University of Iowa, Iowa
City, Iowa

APPLICATIONS OF DELAY FEEDBACK IN CONTROL SYSTEM DESIGN

Drs. Norman Coleman and Edward Carroll, Fire Control and
Small Caliber Weapon Systems Laboratory, U. S. Army
Armament Research and Development Command, Dover, New Jersey

ADAPTIVE LEAD PREDICTION ALGORITHM FOR MANEUVERING TARGET
ENLARGEMENT

Drs. Pak Yip and Norman Coleman, Fire Control and Small
Caliber Weapon Systems Laboratory, U. S. Army Armament
Research and Development Command, Dover, New Jersey

1330   TECHNICAL SESSION IV   Ballroom West

CHAIRPERSON - Dr. Julian Davis, U. S. Army Armament Research
                and Development Command, Dover, New Jersey

ON VOLTERRA INTEGRAL EQUATIONS OF PULSE CONVOLUTION TYPE

Dr. Edward W. Ross, U. S. Army Natick Research and Develop-
ment Command, Natick, Massachusetts

CASE STUDIES OF SEVERAL VOLTERRA INTEGRAL EQUATIONS OCCURRING
IN APPLICATIONS

Professor Ben Noble, Mathematics Research Center, University
of Wisconsin, Madison, Wisconsin

LAME POTENTIALS FOR STATIONARY AND TRAVELING PRESSURE PULSES
IN A HOLLOW CYLINDER

Mr. Alexander S. Elder, Ballistic Research Laboratory,
Aberdeen Proving Ground, Maryland

ON THE SPLINE SOLUTION OF SINGULAR INTEGRAL EQUATIONS

Professors R. P. Srivastav and Erica Jen, State University
of New York at Stony Brook, Stony Brook, New York

1500   Break

1530	GENERAL SESSION II	Ballroom East

CHAIRPERSON - Dr. James Thompson, U. S. Army Tank-Automotive Research and Development Command, Warren, Michigan

SPEAKER - Professor Constantine Dafermos, Brown University, Providence, Rhode Island

TITLE - CAN DISSIPATION PREVENT THE BREAKING OF WAVES?

1630	Adjourn	
1830	No Host Cocktails	Ballroom Foyer
1930	BANQUET	Ballroom

SPEAKER - Dr. Wilford Weeks, U. S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire

TITLE - A NORTH POLE PERSPECTIVE

Wednesday
11 June 1980

0830	TECHNICAL SESSION V	Ballroom East

CHAIRPERSON - Dr. Norman Coleman, Fire Control and Small Caliber Weapon Systems Laboratory, U. S. Army Armament Research and Development Command, Dover, New Jersey

DEFLECTION OF FLOATING ELASTIC PLATES SOLVED BY THE VIRTUAL REACTION METHOD

Dr. Shunsuke Takagi, U. S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire

CYCLIC STRESS-STRAIN BEHAVIOR NEAR A SHARP ELLIPTICAL FLAW

Drs. Dennis M. Tracey and Colin E. Freese, U. S. Army Materials and Mechanics Research Center, Watertown, Massachusetts.

ELEASTIC-PLASTIC ANALYSIS OF SCREW THREADS

Dr. G. P. O'Hara, Benet Weapons Laboratory, Watervliet Arsenal, Watervliet, New York

GENERALIZED PLANE STRAIN PROBLEMS IN AN ELASTIC-PLASTIC
THICK-WALLED CYLINDER

Dr. Peter C. T. Chen, Benet Weapons Laboratory, Watervliet
Arsenal, Watervliet, New York

QUADRATIC AND CUBIC TRANSITION ELELEMENTS

Drs. M. A. Hussain, J. D. Vasilakis and S. L. Pu, Benet
Weapons Laboratory, Watervliet Arsenal, Watervliet, New York

DATA REDUCTION FROM THE EXPANDING CYLINDER TEST

Dr. John Bulk, Ballistic Research Laboratory, Aberdeen
Proving Ground, Maryland

0830   TECHNICAL SESSION II   Ballroom West

CHAIRPERSON - Dr. Yoshisuke Nakano, U. S. Army Cold Regions
              Research and Engineering Laboratory, Hanover,
              New Hampshire

PERTURBATION AND BIFURCATION IN A FREE BOUNDARY PROBLEM

Professors Roger Alexander and Bernard A. Fleishman,
Rensselaer Polytechnic Institute, Troy, New York

A TECHNIQUE FOR SYNTHESIZING THE MOTIONS OF A FLAT PLATE
FOUNDATION DURING AN EARTHQUAKE

Dr. Francis E. Council, U. S. Army Mobility Equipment
Research and Development Command, Ft. Belvoir, Virginia

HOPF BIFURCATION COMPUTATIONS FOR DISTRIBUTED PARAMETER
SYSTEMS

Professors R. F. Heinemann and A. B. Poore, Mathematics
Research Center, University of Wisconsin, Madison,
Wisconsin

THERMOELASTIC WAVE PROPAGATION

Mr. Julian Davis, U. S. Army Armament Research and Develop-
ment Command, Dover, New Jersey, and Professor Yu Chen,
Rutgers University, New Brunswick, New Jersey

FORMATION OF SINGULARITIES FOR THE EQUATION OF GAS FLOW
THROUGH A TUBE OF VARIABLE CROSS-SECTIONS

Professors Reza Malek-Madani and S. J. Lin, Mathematics
Research Center, University of Wisconsin, Madison, Wisconsin

xi

BEAM MOTIONS UNDER MOVING LOADS SOLVED BY FINITE ELEMENT
METHOD CONSISTENT IN SPATIAL AND TIME COORDINATES

Dr. Julian J. Wu, Benet Weapons Laboratory, Watervliet
Arsenal, Watervliet, New York.

1030	Break

1100	GENERAL SESSION III	Ballroom East

CHAIRPERSON - Dr. Edward Lenoe, U. S. Army Materials and
Mechanics Research Laboratory, Watertown,
Massachusetts

SPEAKER - Professor T. C. T. Ting, University of
Illinois-Chicago Circle, Chicago, Illinois

TITLE - WAVE PROPAGATION IN PERIODICALLY LAYERED MEDIA

1200	Lunch

1330	TOUR OF THE U. S. ARMY COLD REGIONS RESEARCH AND ENGINEERING LABORATORY

Thursday
12 June 1980

0900	GENERAL SESSION IV	Ballroom East

CHAIRPERSON - Dr. John D. Powell, Ballistic Research
Laboratory, Aberdeen Proving Ground, Maryland

SPEAKER - Professor James Tasi, State University of New
York at Stony Brook, Stony Brook, New York

TITLE - SHOCK WAVES AND LATTICE DYNAMICS

1000	Break

1030	GENERAL SESSION V	Ballroom East

CHAIRPERSON - Dr. Edward W. Ross, U. S. Army Natick Research
and Development Command, Natick, Massachusetts

SPEAKER - Professor Y. H. Pao, Cornell University, Ithaca,
New York

TITLE - THEORY OF ACOUSTIC EMISSION

1145	ADJOURN

DIRECT AND INVERSE METHODS FOR SCATTERING
BY CRACKS IN THE HIGH FREQUENCY RANGE

Jan D. Achenbach
Department of Civil Engineering
Northwestern University
Evanston, IL. 60201

ABSTRACT. An important method in quantitative non-destructive evaluation of materials (QNDE) is based on scattering of ultrasonic waves by cracks. The presence of a flaw is relatively easy to detect. The determination of its size, shape and orientation from the scattered field poses a challenging inverse scattering problem. In recent years several analytical methods have been developed to investigate scattering of elastic waves by interior cracks and surface-breaking cracks, in both the high- and the low-frequency domains. The appeal of the high-frequency approach is that the probing wavelength is of the same order of magnitude as the length-dimensions of the crack. This gives rise to interference phenomena which can easily be detected. In this paper we discuss approximate methods for the solution of the direct scattering problem in the high-frequency domain, which show good agreement with experimental results. The simple analytical solutions to the direct problem suggest the application of Fourier-type integrals to solve the inverse problem. The application of two kinds of inversion integrals to far-field high-frequency scattering data from flat cracks has been discussed briefly.

I. <u>INTRODUCTION</u>. Reliable methods of quantitative non-destructive evaluation (QNDE), that can be used not only to detect the presence and the approximate location of a flaw, but also to determine its size, shape and orientation are important cornerstones of a damage-tolerant design philosophy.

Among the most useful QNDE methods are those based on the scattering of elastic (ultrasonic) waves by flaws in solids. In the scattered field approach it is attempted to infer the geometrical configuration of a flaw from either the angular dependence of its far-field scattering amplitude at fixed frequency, or from the frequency dependence of its far-field scattering amplitude at fixed angles. In this paper analytical investigations for the scattered field approach to detection of crack-like flaws are discussed. Scattering by interior cracks, surface-breaking cracks and cracks near a boundary will be considered.

In experimental work on quantitative flaw definition by the ultrasonic pulse method either the pulse-echo method with one transducer or the pitch-catch method with two transducers is used. The transducer(s) may be either in direct contact with the specimen, or transducer(s) and specimen may be immersed in a water bath. Most experimental setups include instrumentation to gate out the relevant pulses in the scattered field on the basis of arrival times. The application of a Fast Fourier Transform to these pulses subsequently yields frequency spectra. In the frequency domain the raw scattering data can conveniently be corrected for transducer transfer functions and other characteristics of the system, which have been obtained on the basis

of appropriate calibrations. The corrected experimental data can then be compared with theoretical results that have been obtained by harmonic wave analysis.

For short pulses the frequency spectra of the diffracted signals are centered in the high-frequency (short wavelength) range. High-frequency incident waves give rise to interference processes which can easily be interpreted, and which can provide the basis for an inversion procedure. Particularly the first arriving signals, which are related to the longitudinal waves in the solid, produce a very simple structure in the frequency domain.

Elastodynamic ray theory provides a powerful tool for the computation of fields generated by scattering of time-harmonic waves by cracks, when the wavelength of the incident wave is of the same order of magnitude as characteristic length parameters of the crack. Ray theory has the advantage of simplicity and intuitive appeal. The rules that govern reflection, refraction and edge diffraction of rays are well established, and it is generally not difficult to trace rays from the source via the scatterer to an observer.

Considerable progress has been achieved in recent years in the application of elastodynamic ray theory to scattering by cracks. For cracks in unbounded solids theoretical results have been given by Achenbach et al [1]-[3]. For two-dimensional problems ray theory results have been compared with exact results in Ref.[4]-[5], and with experimental results in Ref.[6].

The basic concepts of elastodynamic ray theory have been presented by Karal and Keller [7]. For time-harmonic wave motion, ray theory provides a method to trace the amplitude of a disturbance as it propagates along a ray. In a homogeneous, isotropic, linearly elastic solid the rays are straight lines, which are normal to the wavefronts. An unbounded solid can support rays of longitudinal and transverse wave motion. These rays are denoted as L-rays and T-rays, respectively. The free surface of a solid can, in addition, support rays of surface-wave motion, which are denoted as R-rays.

When a disturbance is applied to the surface of a body, generally a ray of longitudinal motion as well as a ray of transverse motion are generated. Upon striking an interface, rays produce reflected and refracted rays. Such reflection and refraction problems are well understood. In principle, elastodynamic ray theory can be used to construct scattered fields generated by cylinders, spheres and other curved surfaces of simple geometrical shapes. These fields can be constructed by computing the fields on reflected rays according to well-established rules. The result is called the geometrical elastodynamics (GE) field. The GE field does, of course, not describe the diffracted field which penetrates into the shadow region. Another shortcoming of the GE field is that it shows discontinuities at shadow boundaries and at boundaries of zones of reflected waves. Additional considerations are

required to include the diffracted field. For the high-frequency case these considerations have resulted in the formulation of the geometrical theory of diffraction (GTD) which was formulated by Keller [8].

It should be noted that for sufficiently large frequency the geometrical elastodynamics field may require no correction, i.e., the scattering phenomenon may be entirely dominated by geometrical elastodynamics. This is the case for backscatter from smoothly curved surfaces with radii of curvature very large as compared to the wavelength. On the other hand, when the scattering obstacle has a sharp edge, the effect of edge diffraction may be quite pronounced. Edge diffraction is particularly relevant for cracks when the geometrical elastodynamics approximation only gives a shadow zone and two bundles of reflected rays.

Diffraction by smooth obstacles in elastic solids has been investigated by Resende [9], who also considered diffracted by an edge in a solid, at least for the two-dimensional case.
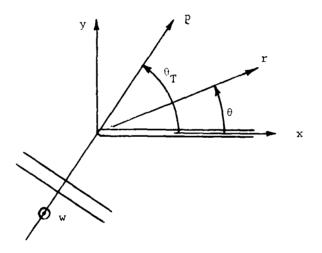
Fig. 1:   Wave of anti-plane strain incident on a
          semi-infinite crack.

II.   DIFFRACTION OF ELASTIC WAVES BY CRACKS.   It is instructive to start a discussion of the interaction of elastic waves with cracks with a brief review of the simplest problem of that type, which is concerned with incidence of waves of anti-plane strain on a semi-infinite crack. With reference to the coordinate system shown in Fig. 1, waves of anti-plane strain are defined by displacements in the z-direction which depend only on x and y. These displacement components, $w(x,y)$, satisfy the wave equation

$$\nabla^2 w + k_T^2 w = 0 \tag{2.1}$$

where ∇^2 is the two-dimensional Laplacian and $k_T^2 = \omega^2/c_T^2$, where ω is the circular frequency and c_T is the velocity of transverse waves ($c_T^2 = \mu/\rho$). Here it is implied that the waves are time harmonic, but the term $\exp(-i\omega t)$ has been omitted, as it will be in the sequel. An incident wave of anti-plane strain is defined by

$$w^i(x,y) = A \exp[ik_T r \cos(\theta-\theta_T)] \tag{2.2}$$

where $(\cos\theta_T, \sin\theta_T)$ defines the propagation direction, and (r,θ) are polar coordinates as shown in Fig. 1. The conditions on the faces of the crack are

$$\frac{\partial w}{\partial y} = 0 \quad x > 0, \ y = 0 \tag{2.3}$$

It should be noted that the problems defined by (2.1)-(2.3) is completely analogous to incidence of an acoustic wave on a rigid screen, or of a magnetically polarized wave on a perfectly conducting screen. The solution is due to Sommerfeld, and it can be found in several places.

A detailed derivation of the solution will not be given. It can be checked that the solution stated below satisfies the conditions (2.1)-(2.3). The solution is expressed in terms of the function F(z) defined by

$$F(z) = \exp(-iz^2) \int_z^\infty \exp(it^2) \, dt \tag{2.4}$$

This function has the following property

$$F(-z) = \pi^{1/2} \exp(-iz^2 + \tfrac{1}{4}\pi i) - F(z) \tag{2.5}$$

For the problem at hand z is real-valued. Integration by parts yields

$$F(z) = \frac{i}{2z} + O(z^{-3}) \qquad \text{for } z \geq 0 \tag{2.6}$$

Equations (2.5) and (2.6) imply that

$$F(-z) = \pi^{1/2} \exp(-iz^2 + \tfrac{1}{4}\pi i) - \frac{i}{2z} + O(z^{-3}) \tag{2.7}$$

To investigate F(z) for small z we rewrite Eq.(2.4) as

$$F(z) = \exp(-iz^2) \left\{ \int_0^\infty \exp(it^2)dt - \int_0^z \exp(it^2)dt \right\} \tag{2.8}$$

4

It then follows easily that for $|z| << 1$

$$F(z) = \frac{1}{2} \pi^{\frac{1}{2}} \exp(\frac{1}{4}\pi i) - z + O(z^2) \tag{2.9}$$

The solution to the problem defined by (2.1)-(2.3) is

$$w^t(x,y) = A\pi^{-\frac{1}{2}}\exp(ik_T r - \frac{1}{4}\pi i)\{F[(2k_T r)^{\frac{1}{2}} \sin \frac{1}{2}(\theta_T - \theta)]$$

$$+ F[(2k_T r)^{\frac{1}{2}}\sin \frac{1}{2}(\theta_T + \theta)]\} \tag{2.10}$$

where the superscript t indicates that this is the total solution, i.e., the incident field is included. Let us investigate some aspects of this solution. In the limit as $r \to \infty$, Eq.(2.10) yields

$$w^t = A\{\exp[ik_T r \cos(\theta - \theta_T) + \exp[ik_T r \cos(\theta + \theta_T)]\}$$

$$\text{for} \quad 2\pi - \theta_T < \theta < 2\pi \tag{2.11}$$

$$w^t = A \exp[ik_T r \cos(\theta - \theta_T)] \quad \text{for} \quad \theta_T < \theta < 2\pi - \theta_T \tag{2.12}$$

$$w^t = 0 \quad\quad\quad \text{for} \quad 0 < \theta < \theta_T \tag{2.13}$$

where (2.6) and (2.7) have been used. Equation (2.11) shows the existence of a zone of reflected waves where the incident wave has been reflected as if the crack where infinite in extent. Equation (2.11) shows a zone of incident waves only, and (2.12) shows that there is no wave motion in the shadow zone as $r \to \infty$. In analogy with geometrical optics, the expressions given by (2.11)-(2.13) are called the geometrical elastodynamics solution (henceforth denoted by w^g). The geometrical elastodynamics solution is discontinuous at the boundaries of the shadow zone and the zone of reflected waves. The diffracted field, w^d, which is defined by

$$w^t = w^g + w^d \tag{2.14}$$

secures a smooth transition across these boundaries.

When $k_T r >> 1$ the diffracted field follows from (2.6) and (2.7) as

$$w^d \sim A (k_T r)^{-\frac{1}{2}}D_{TH}(\theta_T;\theta) \exp(ik_T r) + O[(k_T r)^{-3/2}] \tag{2.15}$$

where

$$D_{TH}(\theta_T;\theta) = (\frac{2}{\pi})^{1/2} \frac{\cos\frac{1}{2}\theta \sin\frac{1}{2}\theta_T}{\cos\theta - \cos\theta_T} e^{i\pi/4} \tag{2.16}$$

is called the diffraction coefficient. Clearly Eq.(2.15) is not valid when $\theta = \theta_T$ or $\theta = 2\pi-\theta_T$, i.e., near the shadow boundary and the boundary of the zone of reflected waves. Exactly on these boundaries the fields follow from (2.10) and (2.9) as

$$w^t = \frac{1}{2} A \exp(ik_T r) + O[(k_T r)^{-1/2}] \tag{2.17}$$

In the immediate vicinities of $\theta = \theta_T$ and $\theta = 2\pi - \theta_T$ the full solutions (2.10) must be used.

The results given by (2.11)-(2.16) can conveniently be interpreted within the context of elastodynamic ray theory. The incident wave consists of an infinite number of rays. The rays that strike the crack are reflected according to the usual rules of plane-wave reflection, and they give the geometrical elastodynamics solution as given by (2.11)-(2.13). The one ray that strikes the crack tip generates a source at the crack tip with an amplitude factor which depends on the angle of observation, and whose radiated field is given by (2.15). It was recognized by Keller [8] that these elementary observations can be generalized to three-dimensions to screens (cracks) with curved edges and to other than plane incident waves.

III. GEOMETRIC THEORY OF DIFFRACTION FOR SCALAR WAVES. A more general "canonical" problem than discussed in the previous section is the one of general oblique incidence, when the propagation vector of the incident plane wave makes an angle ψ with the edge of the semi-infinite screen. For scalar waves this problem has been solved. Far away from the edge, the solution shows the interesting property that the diffracted field behaves locally as a plane wave whose propagation vector emanates from a point on the edge, and makes an angle ϕ with the edge. In terms of ray theory, the interpretation is that an incident ray which strikes the edge at point 0 under an angle ϕ, generates a cone of diffracted rays with semi-angle ϕ, whose apex is on the edge at the point of diffraction 0, and whose axis is along the edge. The fields on the diffracted rays (the generators of the cone) only vary along a ray in the distance to the point of diffraction. If $\phi = \pi/2$ the cone degenerates into a fan of rays, and the solution to the canonical problem is the one given in the previous section.

The geometric theory of diffraction (GTD) generalizes the results of the canonical problem to curves edges and incident waves with curves wavefronts. The Ansatz is that the rays behave in the same way even if the crack edge is curved, i.e., a cone of diffracted rays is generated whose axis is the tangent

6

to the edge of the screen. The fields on the diffracted rays are in terms
of diffraction coefficients (which follow from the canonical problem of plane
wave incidence on a semi-infinite screen with a straight edge), the distance
travelled along a diffracted ray, and the incident field at the point of
diffraction, and certain geometrical correction factors which involve the
curvature of the edge and the curvature of the incident wavefront.

IV. GEOMETRIC THEORY OF DIFFRACTION FOR WAVES IN ELASTIC SOLIDS. The
general ideas outlined in the previous section can be extended to elasto-
dynamic theory. A general groundwork for a three-dimensional geometric theory
of diffraction by cracks in elastic solids was given by Achenbach and Gautesen
[1] and Gautesen, Achenbach and McMaken [2]. The main difference between the
scalar and elastic wave problems lies in the appearance of both longitudinal
and transverse waves in elastic solids, which are coupled by conditions on the
boundaries.

For plane longitudinal and transverse waves, which are under arbitrary
angles of incidence with a traction-free semi-infinite crack, the fields on
the diffracted rays can be obtained by asymptotic considerations for
$\omega r/c_L > > 1$. This was shown in detail in Ref.[1]. The results of Ref.[1] pro-
vide the canonical solutions for a geometric theory of diffraction of elastic
waves. Basic to such a theory is the result that two cones of diffracted
rays are generated when a ray carrying a high-frequency body-wave strikes the
edge of a crack. The surfaces of the inner and outer cones consist of L-rays
(longitudinal) and T-rays (transverse), respectively. The half-angles of the
cones are related by Snell's law. In addition there are 2 rays of surface
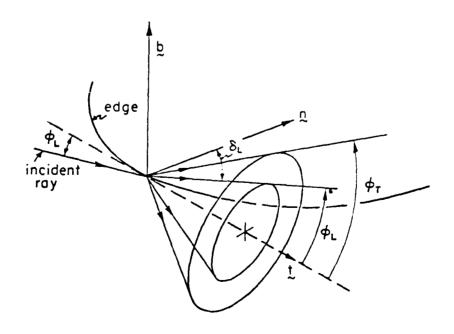waves. (R-rays) on the faces of the crack; one on each crack face.

Fig. 2: Incident ray and cones of diffracted rays

7

Figure 2 shows the cones generated by an incident longitudinal ray. For this case the diffracted longitudinal rays make the same angle ψ_L with the tangent to the edge as the incident ray, and the diffracted rays of transverse motion are under an angle ϕ_T with the edge, where $c_L \cos\psi_T = c_T \cos\psi_L$. The rays of surface wave motion are not indicated in Fig. 2.

Let us define primary diffracted rays as rays that have interacted once with the edge of the crack. Secondary diffracted rays have interacted twice with the crack edge, and n^{th} multiple diffracted rays have interacted n times with the edge of the crack. The total fields at a point of observation are now not just comprised of the fields on the primary diffracted body wave rays. At the edge of the crack rays of crack-face motion are generated, which intersect the crack edges again and generate additional diffracted body wave rays. Some of these secondary diffracted rays will pass through the point of observation. On the faces of the crack, important contributions to the diffracted fields are coming from rays of surface waves because in the first approximation the diffraction coefficients for the body wave motions vanish on the crack faces, except for diffracted horizontally polarized transverse wave motions. In addition, surface wave motions suffer less geometrical decay than body wave motions.

When a R-ray intersects the edge of a crack, a ray of reflected surface wave motion is generated, as well as cones of diffracted rays of longitudinal and transverse motions. The reflection coefficients have been discussed in Ref.[2]. The cones of diffracted L- and T-rays have also been analyzed, and the associated diffraction coefficients have been obtained. With the aid of these results the contributions to the diffracted fields of waves which travel via the crack faces can be computed. Thus the total diffracted field consists of primary diffractions and a system of higher order diffractions. Pertinent results have been summarized in Ref.[4].

With GE and GTD the total displacement field is of the form

$$\underline{u}^t = \underline{u}^g + \underline{u}^d \tag{4.1}$$

This result is still not valid at the boundaries of the shadow zone and the zone(s) of reflected waves. In a further refinement which is called uniform asymptotic theory (UAT), the fields at these boundaries are corrected.

For incident waves with curved wavefronts and for curved diffracting edges, the cones of diffracted rays have envelopes, at which the rays coalesce and the fields become singular. The envelopes are called caustics, and GTD breaks down at caustics.

Within the context of GTD theory of Refs.[1] and [2], the diffracted field at a point of observation Q is comprised of contributions corresponding to "primary" diffracted body-wave rays, which are directly generated by incident body-wave rays, and contributions corresponding to "secondary" diffracted body-wave rays. The latter are generated by rays travelling via the crack faces. Thus, the diffracted displacement field $\underline{u}^d$ at Q can be represented by

$$u^d_\beta = \sum_\alpha u^\alpha_\beta + \sum u^\alpha_{\beta\gamma} \tag{4.2}$$

where u^α_β and $u^\alpha_{\beta\gamma}$ represent the primary and secondary diffractions, respectively. In u^α_β the symbol α defines the incident ray, i.e., $\alpha = L$ or $\alpha = T$, while β defines the diffracted ray, $\beta = L$ or $\beta = T$. In $u^\alpha_{\beta\gamma}$ the symbol β defines the crack-face ray, i.e., $\beta = RS$ (surface-symmetric), $\beta = RA$ (surface-antisymmetric) or $\beta = TH$ (horizontally polarized transverse). The symbol γ defines the body-wave rays generated by diffraction of a crack-face ray; thus $\gamma = L$ or $\gamma = T$. If needed the summations in Eq.(4.2) are carried out over all rays of a particular type passing through Q. The number of relevant rays can be determined on the basis of arrival times in the time domain.

Results obtained on the basis of Eq.(4.2) have been presented in Refs. [3]-[5] and they have been compared with results obtained by numerical solutions of a governing singular integral equation.

In Refs.[4] and [5] a hybrid method has been explored. In this method the crack-opening-displacement (COD) is computed on the basis of elasto-dynamic ray theory, and the diffracted field is subsequently obtained by the use of a representation theorem. The advantage of this approach is that the trouble with ray theory at shadow boundaries and boundaries of zones of specular reflection is eliminated, and caustics only need to be dealt with on the faces of the crack.

V. EXPERIMENTAL RESULTS. Experimental results in the high-frequency range that are suitable for comparison with theoretical results have been reported in Ref.[6]. The sample was a circular disk (2.5 x 10 cm) of titanium alloy which contained a penny-shaped crack of radius 2500μ parallel to the flat faces, and located at the center of the disk. The disk was immersed in water. A transmitter launched a longitudinal wave to the water-titanium interface under normal incidence. This wave was transmitted into the solid, diffracted by the crack, and the diffracted waves were transmitted back into the fluid, where they were received by a second transducer. The experimental set-up and the processing of the data are discussed in some detail in Ref.[6].

In the experimental work the nature of the diffracted signals is determined by their arrival times. Since the first arriving signals are related to longitudinal waves in the solid, it is possible to gate out and separate the purely longitudinal diffracted signals from subsequent signals. By appropriate processing of the experimental data, as discussed in Ref.6, the amplitude-spectrum is obtained for the longitudinal diffracted waves only. Thus for the present comparison of analytical and experimental results we need to consider only the primary diffracted body-wave rays in our analytical work.

9
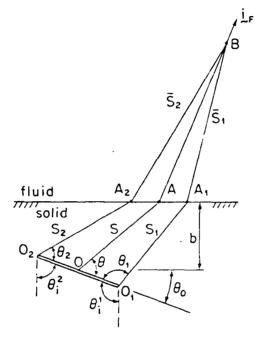
Fig. 3: Geometry in the plane of symmetry of a penny-shaped crack.

The interference patterns for the first arriving longitudinal waves in the fluid are generated by phase differences and amplitude differences on the direct rays from the two crack tips, see Fig. 3. Adding the primary diffracted longitudinal fields from the point 0_1 and 0_2 we obtain in the far-field.

$$u_L \sim F(\theta,\theta_o) \exp[i\omega(S/c_L + \bar{S}/c_F) + i\pi/4] \, U_o \underline{i}_F \tag{5.1}$$
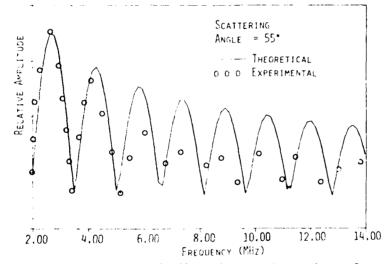
$$F(\theta,\theta_o) = H_1\exp[-i(\omega a/c_L)(\cos\theta - \sin\theta_o)] + H_2\exp[i(\omega a/c_L)(\cos\theta - \sin\theta_o)] \tag{5.2}$$
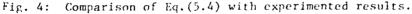
$$H_j = \frac{\text{sgn}(\cos\theta_j) \, T(\theta_L^j) \, |D_L^L(\theta_j;\theta_1^j)|}{(\omega S_j/c_L)^{1/2}(1+S_j/C)^{1/2}(1+\bar{S}_j/E)^{1/2}(1+\bar{S}_j/C)^{1/2}} \qquad j = 1,2 \tag{5.3}$$

Here ω is the circular frequency, a is the crack radius, S = AB, U_o represents the incident wave at point O, and c_L and c_F are the velocities of longitudinal waves in solid and fluid respectively. The geometrical quantities are indicated in Fig. 3. In Eq.(5.3) $T(\cdot)$ is the transmission coefficient at the solid fluid interface, and $D_L^L(\cdot;\cdot)$ is the diffraction coefficient. For details of the derivation of Eqs.(5.1)-(5.3) we refer to Ref.6. It should be noted that one of the terms H_j is imaginary, since the ray has crossed a caustic. Of particular interest is the absolute magnitude of F,

$$|F| = \{|H_1|^2 + |H_2|^2 + 2|H_1||H_2|\sin 2(\omega a/c_L)(\cos\theta - \sin\theta_o)\}^{\frac{1}{2}} \qquad (5.4)$$

Here we have taken into account that either H_1 or H_2 is imaginary.

Fig. 4:   Comparison of Eq.(5.4) with experimented results.

Theoretical results obtained from Eq.(5.4) have been plotted together with experimental data in Fig. 4. The frequency varies from 2 MHz to about 14 MHz. The angle in the solid is $\theta (=\pi/2 - \theta) = 55°$. The amplitudes of the first few cycles agree well. At higher frequencies (above 6 MHz) the experimental results are lower than predicted by theory. One possible explanation is the effect of attenuation which is not accounted for in the theory. In all cases the positions of maxima and minima of the spectra agree well. The locations of the maxima are significant for the inversion process. Additional comparisons with experimental data have been reported in Ref.6.

VI.   ELEMENTARY CONSIDERATIONS FOR THE INVERSE PROBLEM. The theoretical expression for the amplitude spectrum given by Eq.(5.4) implies that the amplitude of the primary diffracted field is modulated with respect to ω/c_L with period

$$P = \pi/a \; |\cos\theta - \sin\theta_o| \tag{5.5}$$

It is of interest to apply Eq.(5.5) to the experimental measurements. Since we know that $\theta_o = 0$, each amplitude spectrum will give a number for a from the periodicity of the modulation. We have

$$a = \frac{c_L}{2 \sin(\theta_L') \; \Delta f_{ave}} \tag{5.6}$$

where $\theta_L' = \frac{1}{2}\pi - \theta_L$ and Δf_{ave} is the average frequency spacing between two consecutive maxima.

The results of the size determination are given in Table 1. The agreement between actual crack radius (a - 2500μ) and the predicted values is excellent.

$\theta_L' = \pi/2 - \theta_L$	Δf_{ave}	computed a in μ
35°	2.18	2530
40	1.87	2630
45	1.83	2450
50	1.68	2460
55	1.60	2410
60	1.47	2500
65	1.39	2510

Table 1:   Crack radius a computed from Eq.(5.6) for a penny-shaped crack in titanium (c_L = 6330 m/s)

VI. SURFACE-BREAKING CRACKS. A surface-breaking crack is one of the most harmful crack configurations. It is, therefore, not surprising that considerable efforts have been devoted to their detection. In this section we review two-dimensional solutions to the direct scattering problem for incident surface waves. The geometry is shown in Fig. 5.
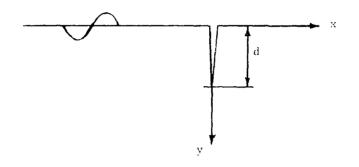

Fig. 5:   Surface-breaking crack.

The easiest problem for the geometry shown in Fig. 5 is concerned with scattering of incident body waves of anti-plane strain. The problem has been considered as a specific and separate problem by some authors. This is, however, completely unnecessary. Referring to Fig. 5, the solution can be obtained from the results for a crack of length 2d ($x = 0, -d \leq y < d$) in an unbounded medium by taking a system of two incident waves which are symmetric with respect to the plane $y = 0$.

Unfortunately, the simple symmetry considerations that hold for the case of anti-plane strain are not valid for the in-plane case. Symmetry considerations do not work because of mode coupling of longitudinal and transverse waves at a traction-free plane. Thus, it is not possible to construct a system of incident waves in an infinite solid with an interior crack, so that the conditions for a surface-breaking crack are automatically satisfied. Hence the problem of scattering by a surface-breaking crack must be considered as a completely separate problem.

Exact solutions for the two-dimensional geometry of a crack of depth d in an elastic half-plane were given in Refs.[10] and [11]. In Ref.[11] the scattered displacement fields due to either a time-harmonic surface wave or a plane time-harmonic longitudinal or transverse body wave incident upon the crack from infinity are investigated. The total field in the half-plane is taken as the superposition of the specified incident field in the uncracked half-plane and the scattered field in the cracked half-plane generated by suitable surface tractions on the crack faces. These tractions are equal and

13

opposite to the tractions generated by the incident wave in the uncracked
half-plane when evaluated in the plane of the crack. By decomposing the
scattered field into symmetric and anti-symmetric fields with respect to
the plane of the crack, a pair of boundary value problems for the quarter-
plane is obtained. These two boundary value problems are reduced by integral
transform techniques to two uncoupled singular integral equations, which are
solved numerically using a collocation scheme. The derivation of the symmetric
equation has been presented in Ref.[10], and the derivation of the anti-symmetric
integral equation is presented in Ref.[11]. The crack-opening displacements are
then easily calculated from the solutions of the singular integral equations.
The exact representations of the diffracted displacement fields are subse-
quently obtained in the form of finite integrals over the crack length, which
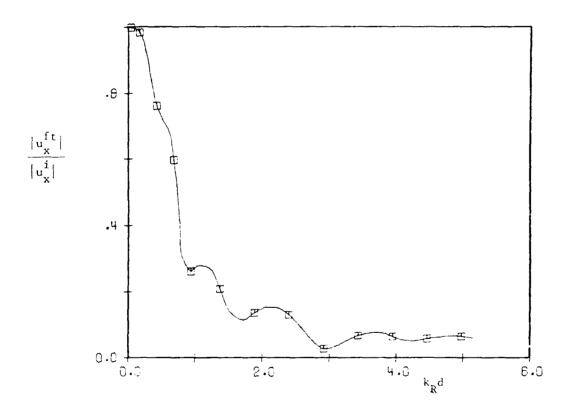are evaluated numerically.

Fig. 6:   The total field ahead of the crack (forward transmitted
            field).

Figure 6 shows the forward transmitted field, and the back-scattered field is
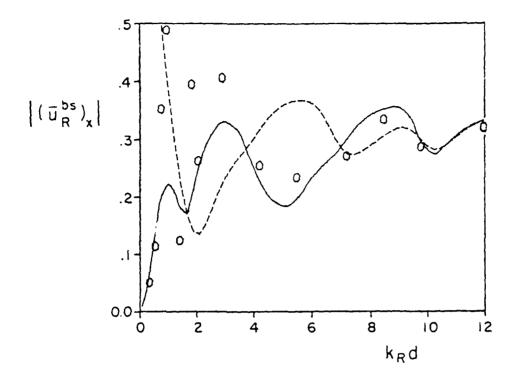shown in Fig. 7. Apparently most of the incident wave is backscattered.

Fig. 7. Comparison of exact and approximate dimensionless x-components of the displacement fields for the back-scattered surface waves, where $|(\overline{u}_R^{bs})_x| = |(u_R^{bs})_x|/A(2c_T^2/c_R^2 - 1)$.
— — — Ray theory of Ref. [12]. O-Exact, see [11]. ———
Asymptotic evaluation of exact integrals of [11].

VII. INVERSION INTEGRALS. We will consider a first approach to the inverse problem for a crack in an unbounded solid. In this approach it is assumed that the scattered field can be adequately represented by ray theory. Thus, the field at a point of observation Q is assumed to consist of the summation of the contributions from each of the rays passing through Q. The nature of these rays depends on the location of Q relative to the crack and relative to the source point S. There can be direct rays, reflected L- and T- rays and diffracted L- and T-rays. If Q is in the shadow zone, only diffracted rays can pass through Q. The magnitudes of the signals carried by diffracted rays is $0[(\omega a/c_I)^{-\frac{1}{2}}]$ as compared to the signals of the direct and reflected rays, where a is a characteristic dimension of the crack.

In practical examples we can include primary diffracted rays which are generated by the incident rays, and secondary diffracted rays generated by rays travelling via the crack faces. In this discussion we will just consider the primary diffracted rays, which correspond to the first arriving diffracted signals. For a flat crack with a smooth edge there are generally two primary diffracted rays. The points of diffraction on the crack edge are called the "flash" points. The locations of the flash points are determined by relatively simple geometrical considerations, based on the rule that the point Q must lie on a cone of diffracted rays.

15

For an incident ray of longitudinal motion the displacement field on a diffracted ray of longitudinal motion is of the general form

$$\underset{\sim}{u}_L^L = [k_L S_Q (1+S_Q/\rho_L)]^{-\frac{1}{2}} \underset{\sim}{d}_L^L \, D_L^L(\phi_L, \theta_L; \theta) \, U_L \, \exp(ik_L S_Q) \tag{7.1}$$

Here $k_L = \omega/c_L$, S_Q is the distance from the point of diffraction D along the diffracted ray, ρ_L is the distance along the ray from the point of diffraction to the caustic, $\underset{\sim}{d}_L^L$ defines the direction of displacement, and $D_L^L(\phi_L, \theta_L; \theta)$ is the diffraction coefficient. The angles ϕ_L and θ_L define the direction of the incident ray relative to a coordinate system at the point of diffraction, θ defines the point of observation, and U_L defines the incident ray at the point of diffraction:

$$U_L = A \, S_D^{-1} \, \exp(ik_L S_D) \tag{7.2}$$

where S_D is the distance from the source point to the point of diffraction.

It is now assumed that we know a point 0 in the vicinity of the crack, while the source point S and the point of observation Q are far from the crack. Let $\underset{\sim}{x}_S$, $\underset{\sim}{x}_Q$ and $\underset{\sim}{x}_D$ denote the position vectors of S, Q and the flash point D relative to 0. Let $x_S = |\underset{\sim}{x}_S|$, $x_Q = |\underset{\sim}{x}_Q|$ and $x_D = |\underset{\sim}{x}_D|$, then x_S, $x_Q \gg x_D$. Defining the unit vectors $\hat{\underset{\sim}{x}}_Q = \underset{\sim}{x}_Q/x_Q$ and $\hat{\underset{\sim}{x}}_S = \underset{\sim}{x}_S/x_S$, we can write

$$S_Q \sim x_Q - (\hat{\underset{\sim}{x}}_Q \cdot \underset{\sim}{x}_D) \tag{7.3}$$

$$S_D \sim x_S - (\hat{\underset{\sim}{x}}_S \cdot \underset{\sim}{x}_D) \tag{7.4}$$

Equation (7.1) may then be expressed in the form

$$\frac{\underset{\sim}{u}_L^L}{G_L(x_S) G_L(x_Q)} = F \, \underset{\sim}{d}_L^L \, k_L^{-\frac{1}{2}} \, \exp(-ik_L \underset{\sim}{q} \cdot \underset{\sim}{x}_D) \tag{7.5}$$

where the bisector vector q is defined as

$$\underset{\sim}{q} = \hat{\underset{\sim}{x}}_S + \hat{\underset{\sim}{x}}_Q , \tag{7.6}$$

and

$$G_L(x) = (4\pi x)^{-1} \, \exp(ik_L x) \tag{7.7}$$

16

The function F follows from (7.1).

The form the diffracted field given by (7.5) suggests simple Fourier-type inversions integrals to recover the size, shape and orientation of a crack from the far-field data. The following inversion integrals have been investigated in some detail in Ref.[13].

(i) $\quad f_1^*(\lambda) = \int\limits_{-\infty}^{\infty} \exp(ik_L g \cdot \underset{\sim}{\lambda}) \, f(k_L) \, dk_L$ \hspace{2cm} (7.8)

(ii) $\quad f_2^*(\lambda) = \int\limits_{-\infty}^{\infty} k_L'^2 \exp(ik_L q \cdot \underset{\sim}{\lambda}) \, f(k_L) \, dk_L$ \hspace{2cm} (7.9)

where λ defines a test point in the mediums.

Suppose we now take the experimentally obtained amplitude spectrum of the early-arriving longitudinal diffracted signal, and divide it by $G_L(x_S)$ and $G_L(x_Q)$. It may then be assumed that the result is of the general form given by the right-hand-side of (7.5). We then apply the inversion integral given by (7.9) to this result. By virtue of the relation

$$\int\limits_{-\infty}^{\infty} \exp(ik_L \rho) \, dk_L = \delta(\rho) \hspace{2cm} (7.10)$$

we obtain a Dirac delta function when

$$q \cdot (\lambda - \underset{\sim}{x}_D) = 0 \hspace{2cm} (7.11)$$

Thus, the application of (7.9) to the right-hand side of (7.5) will give a delta-function behavior when the test point λ is located in a plane through the flash point normal to the known bisector vector q. For convenience λ can be taken along q. By taking many points of observation Q, the crack edge can, in principle, be constructed. For further discussions we refer to Ref.[13].

It should of course be realized that the scattered field is generally only known over a finite frequency range. In that case the application of the inversion integral yields a function of the form $\sin(k_L x)/x$ rather than a Dirac delta function, and the position of the plane corresponds to the principal peak.

REFERENCES

1.   J. D. Achenbach and A. K. Gautesen, "Geometrical Theory of Diffraction for Three-D Elastodynamics," J. Acoust. Soc. Amer. 61 pp. 413-421 (1977).

2.   A. K. Gautesen, J. D. Achenbach and H. McMaken, "Surface Wave Rays in Elastodynamic Diffraction by Cracks," J. Acoust. Soc. Amer. 63, p. 1824 (1978).

3.   J. D. Achenbach, A. K. Gautesen and h. McMaken, "Diffraction of Point-Source Signals by a Circular Crack," Bull. Seism. Soc. Amer. 68, pp. 889-905 (1978).

4.   J. D. Achenbach, A. K. Gautesen and H. McMaken, "Diffraction of Elastic Waves by Cracks-Analytical Results," in Elastic Waves and Non-Destructive Testing of Materials (Y. H. Pao, editor), AMD-Vol. 29, American Society of Mechanical Engineers, 1978.

5.   J. D. Achenbach, A. K. Gautesen and H. McMaken, "Application of Ray Theory to Diffraction of Elastic Waves by Cracks," in Recent Developments in Classical Wave Scattering: Focus on T-Matrix Approach, Pergamon Press (in press).

6.   J. D. Achenbach, L. Adler, D. Kent Lewis and H. McMaken, "Diffraction of Ultrasonic Waves by Penny-Shaped Cracks in Metals: Theory and Experiment," J. Acoust. Soc. Amer. 66, p. 1648 (1979).

7.   F. C. Karal and J. B. Keller, "Elastic Wave Propagation in Homogeneous and Inhomogeneous Media," J. Acoust. Soc. Amer. 31, p. 694 (1959).

8.   J. B. Keller, "A Geometrical Theory of Diffraction," Calculus of Variations and Its Applications, McGraw-Hill, 1958.

9.   E. Resende, "Propagation, Reflection and Diffraction of Elastic Waves," Ph.D. Dissertation, New York University, 1963.

10.   J. D. Achenbach, L. M. Keer and D. A. Mendelsohn, "Elastodynamic Analysis of an Edge Crack," J. Applied Mechanics, in press.

11.   D. A. Mendelsohn, J. D. Achenbach and L. M. Keer, "Scattering of Elastic Waves by a Surface-Breaking Crack, WAVE MOTION, in press.

12.   J. D. Achenbach, A. K. Gautesen and D. A. Mendelsohn, "Ray Analysis of Surface-Wave Interaction with an Edge Crack," IEEE Transactions on Sonics and Ultrasonics SU-27, p. 124(1980).

13.   J. D. Achenbach, K. Viswanathan and A. Norris, "An Inversion Integral for Crack-Scattering Data," WAVE MOTION, 1, p. 299 (1979).

SOLITARY WAVES AND SHOCK PROFILES IN
THE THREE-DIMENSIONAL LATTICE

John D. Powell and Jad H. Batteh*
Applied Physics Branch
Ballistic Modeling Division
Ballistic Research Laboratory, USAARRADCOM
Aberdeen Proving Ground, MD 21005

ABSTRACT. The propagation and interaction of solitary waves in a
three-dimensional, monatomic, face-centered-cubic lattice are investi-
gated. The atoms which constitute the lattice are assumed to interact
through a Morse-type interatomic potential. A sequence of solitary
waves is generated by subjecting the lattice to shock compression at a
steady rate and, from the numerical solution of the atomic equations of
motion, the stability of the solitary waves is studied. It is pointed
out that in general the pulses are not so stable as in similar one-
dimensional models and, in particular, are rather unstable when encount-
ering oscillations transverse to their propagation direction. It is
also observed that under some conditions coupled longitudinal and trans-
verse solitary waves can propagate in phase at the same velocity through
the lattice. The long-wavelength, continuum limit of the equations of
motion is then derived and it is demonstrated analytically that these
equations also predict the existence of the coupled-wave profiles ob-
served numerically. The way in which solitary waves may affect the
shock profile and conventional assumptions regarding it in solids is
also discussed.

1. INTRODUCTION. In some recent calculations we have investigated
the propagation of shock waves in both one-dimensional [1,2] and three-
dimensional [3] discrete, crystal lattices. Our efforts have been moti-
vated to some extent by the early computer-molecular-dynamic calculations
of Tsai and coworkers [4] which revealed a number of anomalous effects
in the shock profile. Our work has tended to substantiate these findings
and has suggested that the existence in the lattice of solitary waves,
or rather well-defined, fairly stable pulses, could account for the un-
expected results. It has, therefore, been of some interest to us to
study the properties of a solitary waves, particularly in three dimen-
sions, since this problem has received relatively little attention in
the literature.

In this paper we will discuss the results of our investigation of
the properties of solitary waves in a three-dimensional lattice. After
defining the model, we begin by demonstrating how solitary waves can be

*Present address: Science Applications, Inc., Atlanta, GA 30339

generated in the lattice and how their stability can be investigated using a computer-molecular-dynamic technique. The continuum limit of the equations of motion will then be taken and it will be shown that these equations are capable of predicting analytically many of the same effects revealed in the numerical study. Finally, we discuss briefly the anomalous effects in the shock-wave calculations, and suggest how the properties of the solitary waves account for these effects. The discussion in this paper is intended to be more abbreviated and qualitative than that presented elsewhere. For greater detail, the reader is referred to the literature [3,5].

II. MODEL AND EQUATIONS OF MOTION. The three-dimensional model which we have employed in the calculations is shown schematically in Figure 1. It consists of a face-centered-cubic lattice which is made as long as necessary in the z direction to complete the calculation and which is periodic in the x and y directions. A typical cross section of the lattice is shown on the left-hand side of the figure and contains eight unique atoms; we have, however, in many calculations employed as many as 32 atoms in the cross section. Planes of atoms normal to the z axis are numbered consecutively, beginning with the first located at z=0, and atoms within a given plane can be numbered any convenient manner.
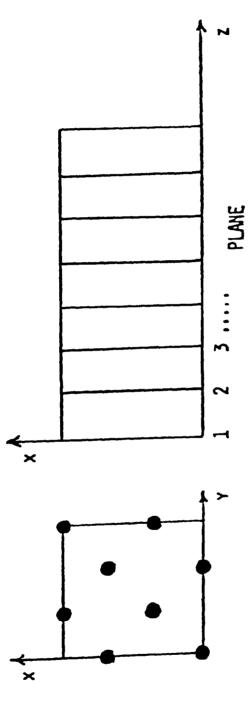
The atoms within the lattice are assumed to interact through a Morse-type interatomic potential. Thus, the equation of motion satisfied by the αth atom in the ith plane can be written

$$\frac{d^2 \vec{r}_{i,\alpha}}{d\tau^2} = 2 \, RA_o \sum_{\beta,j} \left[e^{-2R(A_o|\vec{r}_{i,\alpha}-\vec{r}_{j,\beta}|-1)} - e^{-R(A_o|\vec{r}_{i,\alpha}-\vec{r}_{j,\beta}|-1)} \right]$$

$$\times \frac{\vec{r}_{i,\alpha}-\vec{r}_{j,\beta}}{|\vec{r}_{i,\alpha}-\vec{r}_{j,\beta}|} \, . \tag{1}$$

All quantities have been made dimensionless: $\vec{r}_{i,\alpha}$ represents the position vector to αth atom in the ith plane, and is normalized by the lattice constant; A_o is the lattice constant, normalized by the separation of an isolated atom pair at minimum potential; τ represents the time, normalized by $(m/D)^{\frac{1}{2}} a_o$, where m is the atomic mass, D the dissociation energy, and a_o the lattice constant; and R is a parameter indicating the degree of nonlinearity in the Morse potential. The sums over j and β in Eq. (1) go over all atoms in the vicinity of the (i,α)th for which an appreciable interaction occurs. Equation (1) just represents

Figure 1. Model for three-dimensional calculations. The drawing on the left shows a typical cross section of the lattice. The vertical lines in the drawing on the right indicate planes of atoms normal to the z axis.

21

Newton's second law for the atoms in the lattice and we are concerned
with solving this equation numerically for each atom and, from the solu-
tion, inferring the response of the lattice to any excitation. For the
calculations discussed, the equations were solved under the assumption
that R was given by 6.29, a value which is appropriate for a lattice of
nickel atoms [6].

III. GENERATION OF SOLITARY WAVES AND NUMERICAL STUDY OF THEIR
STABILITY. A sequence of solitary waves can be generated in the lattice
by having each atom initially at rest in its equilibrium position and
subjecting the lattice to shock compression. To do so we drive the
end-most plane of atoms, located at $z=0$, along the positive z axis at a
constant compression velocity. The formation of the solitary waves can
be seen most easily by looking at a series of velocity-time trajectories
of various planes of atoms in the lattice subsequent to their being ex-
cited by the shock front. A typical set of such trajectories is shown
in Figure 2. Each plot is begun at time τ_o which corresponds to the
time at which the plane in question is first encountered by the shock,
and v_i denotes the (common) velocity of atoms in the ith plane, nor-
malized by $\sqrt{D/m}$.

We see that the shock front introduces an oscillatory wave profile
at the second plane which is very similar to that of a harmonic, one-
dimensional chain. As the shock propagates farther, however, and the
atoms become farther displaced from their equilibrium positions, non-
linear effects become increasingly important. These effects tend to
steepen the profile as can be seen in the trajectory of the 20th plane.
Furthermore, it is found that the higher-amplitude pulses propagate at
a higher velocity and, consequently, the pulses tend to spread apart as
they form. The spreading effect can be seen by comparing the separation
of peaks at the 20th and 40th plane. Asymptotically, which for practical
purposes occurs by about the 40th plane, the pulses approach the same
height and the spreading ceases to occur. At this point a sequence of
solitary waves has formed in the vicinity of the front and they will
propagate indefinitely into the lattice without changing their shapes.
Physically, they represent a balance between the dispersion in the
lattice, which tends to spread the pulses out, and the nonlinearity in
the lattice which tends to steepen them. In the event that the solitary
waves are stable to various types of perturbations, they are called
solitons.

A single solitary wave can be isolated from the sequence near the
shock front and its properties studied. The first question that might
be investigated is to ask the extent to which the solitary waves are
stable to mutual collisions in three dimensions. To investigate this
problem we launched two solitary waves, having equal but oppositely
directed velocities at opposite ends of a lattice that was 48 planes
long. Shown at the top of Figure 3 is the rightward-moving solitary
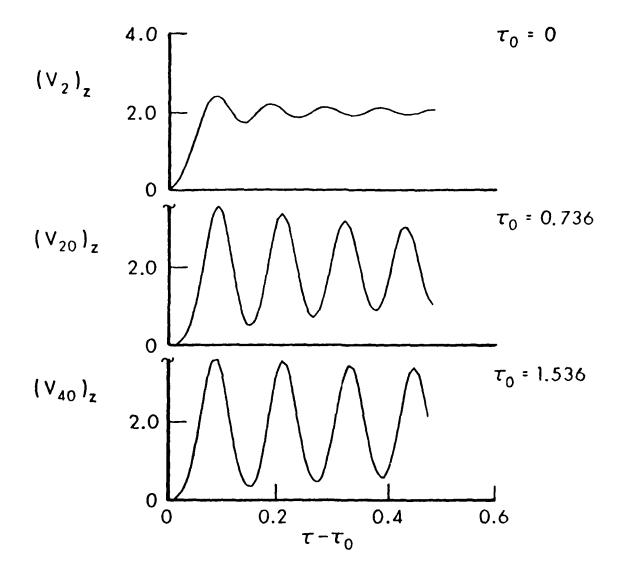wave as it encounters the 13th plane in the lattice. The leftward-

22

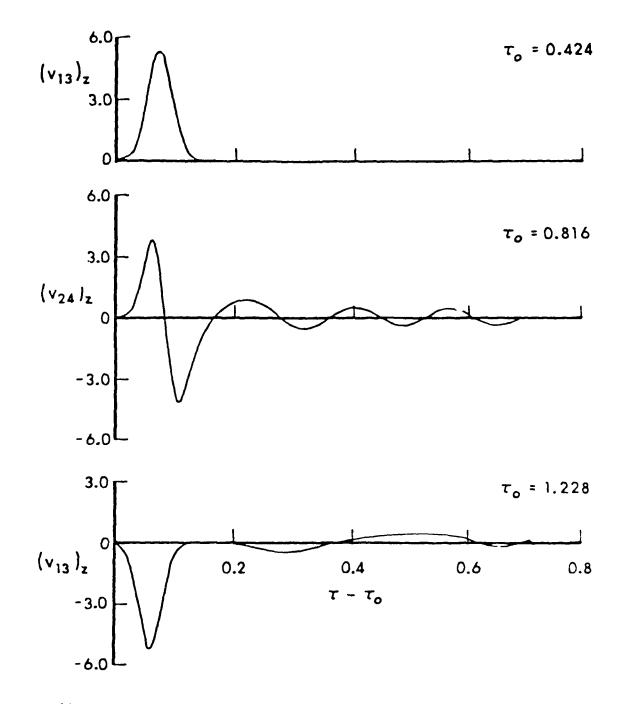Figure 2. Evolution of solitary waves near the shock front.

23

Figure 3. Collision of two solitary waves.

moving pulse, not shown, is now at the 35th plane. The two pulses collide in the vicinity of the 24th plane and a very nonlinear inter- action occurs as can be seen in the center of the figure. Finally, at a much later time, the collision has occurred and the negative-velocity solitary wave has reached the 13th plane whose trajectory is shown on the bottom of the figure.

It is evident from the latter trajectory that some substantial oscillations are left behind after the collision, and the pulses are not completely stable. This case might be compared with the correspond- ing case in one dimension where previous calculations [1,2] indicated that the pulses were, for the Morse lattice, stable to within the accu- racy of the numerical data. Even in the three dimensions, however, the integrity of the pulses is maintained fairly well.

We have also investigated the stability of the solitary waves to small-amplitude, longitudinal and transverse planar oscillations and to random thermal oscillations in the lattice by sending a solitary wave through a region of the lattice which contained one or more of these particular types of oscillations. Without belaboring the details, we found essentially that the pulses were fairly stable both to small longi- tudinal oscillations and to thermal background, but were rather unstable to planar oscillations in the transverse direction. In fact, a rather interesting effect occurred when we sent a solitary wave of sufficiently high amplitude through a region of the lattice which contained transverse planar oscillations in the y direction. The incident pulse, shown at the top of figure 4, was longitudinal and had an amplitude of about 3.60. If the pulse were stable to the transverse oscillations, we would expect the pulse to emerge from the oscillatory region with its shape unchanged. What we actually observed, however, shown on the lower half of the figure, was that two pulses emerged, one longitudinal and one transverse. We have followed their propagation some distance into the lattice and found that they propagate exactly in phase and at the same propagation veloc- ity. We have therefore called the emergent pulses coupled solitary waves. As might be expected, the emergent longitudinal pulse has an amplitude which is smaller than the incident pulse (3.1 here compared to 3.60). Evidently, the longitudinal solitary wave accentuates the transverse planar oscillations initially present in the lattice, thereby producing the coupled pulses observed in the figure.

IV. RESULTS PREDICTED BY CONTINUUM EQUATIONS. Since the results of the preceding section are completely numerical, it is of interest to ask whether we can obtain approximate analytic solutions for the soli- tary wave profiles and predict the existence of coupled solitary waves. In an effort to do so we made a number of simplifying assumptions in the equations of motion represented by Eq. (1). First, we took the continu- um limit of the equation which should be valid for excitations whose wavelength is long compared to the interparticle separation. We re- tained in the limit not only the usual harmonic terms, but also the lowest order terms in the dispersion and in the nonlinearity of the lat-

INCIDENT SOLITARY WAVE
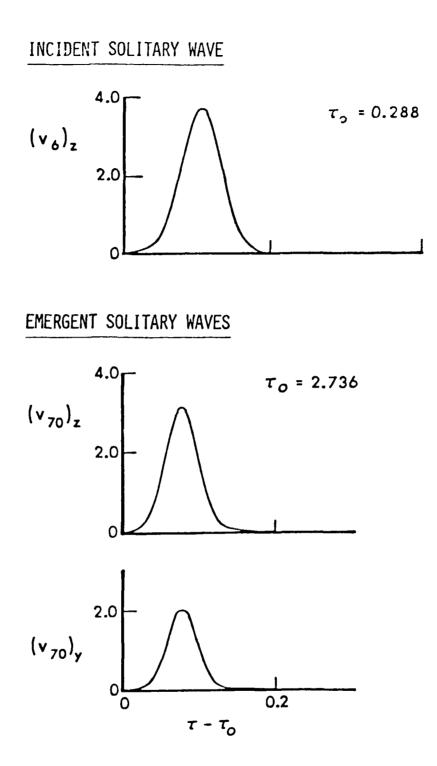

EMERGENT SOLITARY WAVES

Figure 4. Generation of coupled solitary waves. The upper part of the figure shows a longitudinal solitary wave incident on a region of the lattice containing transverse planar oscillations. The lower part shows the two emergent solitary waves.

tice. Second, we restricted ourselves to only planar oscillations in
the longitudinal (z) direction and one transverse (y) direction. There-
fore, each atom in a particular plane normal to the z axis had, at any
time, a velocity identical to every other atom in the plane. The other
transverse direction could have been included also, but doing so greatly
complicates the algebra without really clarifying any essential physics
of the problem. Third, we assumed that each atom in the lattice interacts
with only its nearest neighbors, of which there are 12 in a face-centered-
cubic lattice. Finally, since the solitary waves represent steady,
travelling-wave solutions to the equations of motion, we assumed solutions
for the y and z components of the planar velocities of the form

$$v_y = v_y(z-C\tau) = v_y(\xi)$$

(2)

$$v_z = v_z(z-C\tau) = v_z(\xi) \ .$$

The unknown parameter C represents the propagation speed of the solitary
waves. Identical arguments were chosen for the two functions since the
coupled pulses were found to propagate in phase.

The assumed forms of the solutions indicated in Eq. (2) were then
substituted into the continuum equations derived in the manner discussed
above (see Ref. 5 for details) and two coupled, nonlinear, second-order
differential equations were obtained for the planar velocities:

$$v_y{''} = \alpha v_y - 4\beta v_y v_z$$

(3a)

$$v_z{''} = \gamma v_z - \delta v_z^{\ 2} - \beta v_y^{\ 2} \ .$$

(3b)

The primes represent differentiation with respect to ξ and the parameters
α, β, γ, and δ are given by

$$\alpha = 12(C^2/ C_t^2 - 1)$$
$$\beta = 3(3R-1)/C$$
$$\gamma = 12(C^2/ C_\ell^2 - 1)$$
$$\delta = 18(R-1)/C.$$

(4)

Here $C_\ell = 2\sqrt{2} \ R$ is the long-wavelength longitudinal sound speed in the
crystal, and $C_t=2R$ is the corresponding transverse speed. The solution

27

of Eqs. (3) should predict the profiles of the coupled solitary waves in the continuum limit. Of course, for the case $v_y=0$, the solution of the equation for v_z corresponds to the profile of an isolated longitudinal solitary wave.

Unfortunately Eqs. (3) are still too difficult to solve analytically, but are obviously easier to solve numerically than is Eq. (1). Their numerical solution can be effected by assuming an amplitude v_{zo} for the longitudinal pulse and deriving from Eqs. (3) the corresponding amplitude, v_{yo}, of the transverse pulse. One finds [5]

$$v_{yo} = v_{zo} \left[\frac{2\delta v_{zo}/3-\gamma}{\alpha/2-2\beta v_{zo}} \right]^{1/2} . \tag{5}$$

The equations are then integrated numerically assuming that their maximum values are attained at $\xi=0$. It is found that solutions which remain finite at infinity are obtained only if the appropriate value of C, found by trial and error, is used in the numerical solution.

For the limiting case in which $v_{yo} \ll v_{zo}$, however, it is possible to obtain an approximate analytic solution to Eqs. (3). The solution may be viewed as the first step in an iterative procedure. We proceed by noting that since v_y is small, we can, as a first approximation, set it equal to zero in Eq. (3b). The resulting equation can then be reduced to a quadrature and integrated to yield,

$$v_z = \frac{3\gamma}{2\delta} \, \text{sech}^2(\tfrac{1}{2} \sqrt{\gamma} \, \xi) . \tag{6}$$

We now substitute Eq. (6) into Eq. (3a), reducing it to a second-order, linear differential equation which is identical in form to the time-independent Schroedinger equation. The equation can be solved by series solution [5] and, for the case R=6.29, we obtain

$$C = 1.66 \, C_\ell \tag{7}$$

and

$$v_y = v'_{yo} \, \text{sech}^{5.21}(2.3\xi) . \tag{8}$$

v'_{yo} represents the amplitude of the transverse solitary wave in this approximation. Using the value of C represented by Eq. (7), we obtain from Eq. (6)

$$v_z = 9.8 \, \text{sech}^2(2.3\xi) . \tag{9}$$

We should emphasize that Eqs. (8) and (9) represent only a very approximate solution to Eqs. (3). The solution predicts, for instance, that the amplitude of the longitudinal solitary wave in the coupled-wave profile is identical to that for an isolated longitudinal solitary wave having the same value of C; actually, the amplitude of the longitudinal pulse is reduced in the coupled configuration and the amount of reduction depends on the amplitude of the transverse pulse. Furthermore, the value of v'_{yo} is not predicted by the analytic solution in lowest order and, if one wishes to compare analytic and numerical solutions, this parameter must be fit to the numerical results. Finally, the lowest-order approximation yields only one acceptable eigenvalue and therefore only one solution to the equations. Other solutions do exist, however, and these are no doubt predicted by the higher-order terms.

Despite these limitations, the solutions represented by Eqs. (8) and (9) do predict the solitary-wave profiles quite well in the limit $v_y \ll v_z$. To demonstrate this agreement explicitly, we have solved Eqs. (3) numerically for a longitudinal solitary wave having an amplitude of about 10.2. The solution was found to diverge at infinity unless C was given by the value, $C = 1.68C_\ell$. From Eq. (5), then, the amplitude of the transverse wave is about 1.5. The results of the numerical calculation are shown by the solid-line curves in Figure 5, in which velocity profiles are plotted as a function of ℓ/C. Since it is apparent that $v_y \ll v_z$ in this case, the analytic solutions represented by Eqs. (8) and (9) should approximate the profiles reasonably well. A graph of the analytic solution for the longitudinal wave [Eq. (9)] is shown by the dashed curve at the top of Figure 5. Furthermore, when we set $v'_{yo} = 1.5$ in Eq. (8) and attempted to plot the transverse pulse on the lower graph in Figure 5, the analytic result was found to be coincident with the numerical result to within the accuracy with which we could plot the data. Obviously the agreement is quite good.

As a final point, we should indicate that we have observed coupled solitary waves in our numerical studies of the discrete-lattice equations only for rather large-amplitude longitudinal solitary waves. Furthermore, we have been unable to obtain convergent numerical solutions to Eqs. (3) whenever we assumed an amplitude for the longitudinal pulse that was smaller than that predicted by the analytic solution, namely, 9.8. Apparently, then, a threshold amplitude exists for the longitudinal solitary wave, below which it cannot support the propagation of a transverse wave coupled to it. Furthermore, in the continuum limit, that amplitude is predicted by the analytic solution in lowest order.
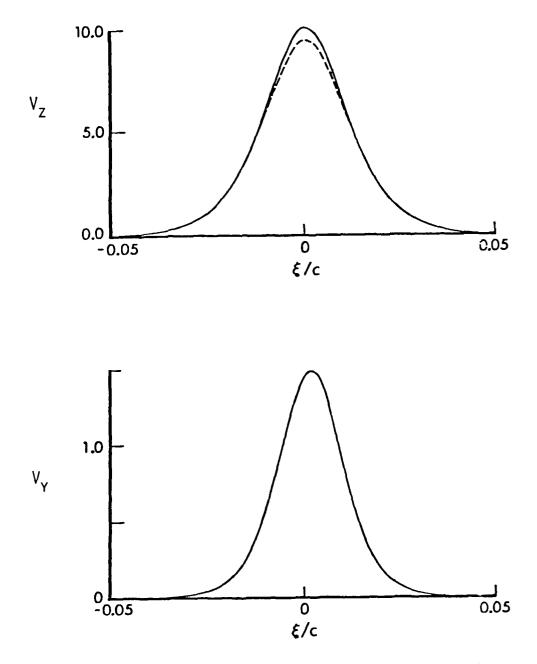
Figure 5. Comparison of numerical and analytic solitary-wave profiles. The solid line represents the numerical solution of Eqs. (3), the dashed line the analytic solution.
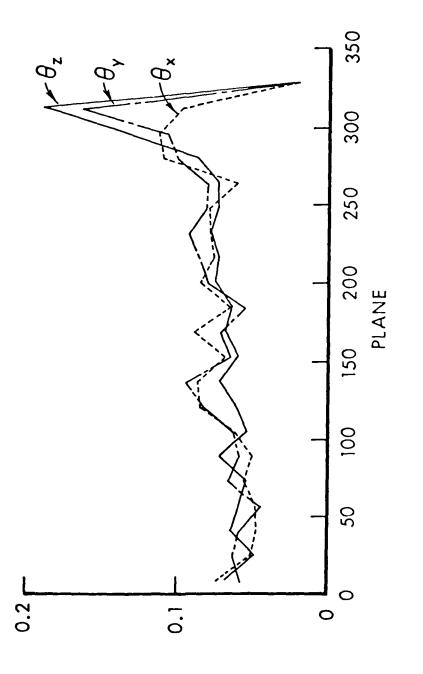
V. EFFECTS OF SOLITARY WAVES ON SHOCK PROFILES.   As indicated previously our interest in the study of solitary waves arise from attempts to explain anomalous effects which occur in the computer-molecular-dynamic simulations of shock propagation in discrete crystal lattices.   Those calculations were carried out using the same model discussed here and the shock wave was again initiated and sustained by driving the end-most plane of atoms at a constant compression velocity. Usually, however, we allowed for some initial thermal motion in the lattice prior to compression in order to simulate an initial ambient temperature.   Nevertheless, we still observed solitary waves, both isolated and coupled, propagating amid and interacting with the thermal background in the lattice.   The existence of the solitary waves in the lattice can account for some rather unexpected results which occur in these shock-wave calculations.
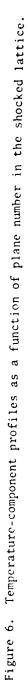
First, because the pulses which are growing into solitary waves propagate at speeds which increase with increasing amplitude, they tend to spread apart as they form.   This spreading effect gives rise, at least at early times and for low ambient temperatures, to a nonsteady shock profile.   Thus, the transition region between the two equilibrated parts of the lattice becomes longer as the shock wave propagates farther into the crystal.   This effect was first noted by Tsai and coworkers [4] in earlier shock-wave calculations.

Second, because of the fair degree of stability of the solitary waves, the approach to thermal equilibrium behind the shock front is rather slow.   It is clear that if the solitary waves were completely stable (solitons), no mechanism would exist for destroying this orderly progression of energy and thermal equilibrium could never be attained. Because the pulses do decay somewhat as they are subjected to various perturbations, however, there is a tendency for the lattice to equilibrate, but only at distances far behind the front.

Finally, one of the more interesting anomalous effects which occurs in shock-wave simulations is the existence of an overshoot in the thermal-energy density directly behind the front.   In particular, if one defines a "temperature" associated with each Cartesian direction, it is found that each temperature overshoots its final equilibrated value behind the front for strong shocks.   The effect is shown in Figure 6, where each of the three temperatures is plotted as a function of position behind the shock when the front is at the 520th plane.   The overshoots in the three Cartesian directions can be accounted for by the existence of high-amplitude, coupled solitary waves behind the front. For weak shock waves, it was observed that the overshoots in the transverse directions disappeared.   In that case, evidently, the amplitudes of the longitudinal solitary waves lie below the threshold for which coupled solutions can exist.

All three of these effects are clearly in contradiction to the usual assumptions and/or results of continuum-mechanical treatments.

31

Figure 6. Temperature-component profiles as a function of plane number in the shocked lattice.

Those calculations predict that the shock wave is steady, that the shock-front thickness is quite small, and that the temperature rises montonically from its ambient value ahead of the shock to its final value behind the shock. In future calculations it would be of interest to make the model more realistic in an effort to determine whether any of these effects is likely to occur in real solids.

REFERENCES

1. J.H. Batteh and J.D. Powell, "Shock Propagation in the One-Dimensional Lattice at a Nonzero Initial Temperature," J. Appl. Phys. 49, 5933 (1978).

2. J.H. Batteh and J.D. Powell, "Soliton Propagation in a One-Dimensional Lattice under Shock Compression," in Solitons in Action, edited by K. Lonngren and A. Scott (Academic, New York, 1978), p. 257.

3. J.D. Powell and J.H. Batteh, "Effects of Solitary Waves Upon the Shock Profile in a Three-Dimensional Lattice," J. Appl. Phys. 51, 2050 (1980).

4. See, e.g., D.H. Tsai, "An Atomistic Theory of Shock Compression of a Perfect Crystalline Solid," in Accurate Characterization of the High-Pressure Environment, edited by E.C. Lloyd, Natl. Bur. Stds. Spec. Publ. No. 326 (U.S. GPO, Washington, DC, 1971), p. 105.

5. J.H. Batteh and J.D. Powell, "Solitary-Wave Propagation in the Three-Dimensional Lattice," Phys. Rev. B 20, 1398 (1979).

6. F. Milstein, "Applicability of Exponentially Attractive and Repulsive Interatomic Potential Functions in the Description of Cubic Crystals," J. Appl. Phys. 44, 3825 (1973).

TRAVELING WAVE SOLUTIONS OF A MODEL SYSTEM

FOR FLAME PROPAGATION

SHAO-SHIUNG LIN

Mathematics Research Center, University of Wisconsin

ABSTRACT

A simplified model for flame propagation is derived from the lowest order terms of the asymptotic expansion (in terms of a suitable length scale) of the full set of nonlinear equations for gaseous combustion in an open infinite tube. It is shown the simplified model system supports unique traveling wave solutions determined by the unburned gas state. The problem of the "cold boundary difficulty" is analyzed.

1. INTRODUCTION. It is well-known that the flame fronts in a combustible gas mixture are generated as a balance of the energy release from chemcial reactions and transport processes such as heat conduction or chemical species diffusion. A model for the flame propagation based on the assumption that the density of the gas be constant during the combustion process has been studied in detail [5], [2]. In this paper, we intend to improve the model by taking account of the fact that the gas expands after the combustion, and hence induces motion of the gas. We will discuss the existence, uniqueness, and properties of the flame fronts. We will, in particular, take note of the "cold boundary difficulty".

The model under study is derived from the lowest order terms of an asymptotic expansion of the complete set of equations governing the dynamics of gaseous combustion. The asymptotic expansion is done with respect to a

PRECEDING PAGE BLANK-NOT FILMED

typical length scale of the reaction zone. Thus, the pressure across the
flame front necessarily remains constant. If we also assume that the density
is constant, our model reduces to the reaction-diffusion model studied in [5].

II. THE GOVERNING EQUARIONS. Gaseous combustion obeys the conservation
laws of mass, momentum, energy and chemical species. In one space dimension
$(-\infty < x < \infty)$ and time $t > 0$ the corresponding set of nonlinear pde is:

$$\rho_t + (\rho u)_x = 0 \tag{1}$$

$$(\rho u)_t + (\rho u^2 + p)_x = \nu\, u_{xx} \tag{2}$$

$$(\tfrac{1}{2}\rho u^2 + \rho e)_t + (u(\tfrac{1}{2}\rho u^2 + \rho e) + u\, p)_x \tag{3}$$
$$= (\lambda T_x)_x + \nu(u\, u_x)_x + Q\,\frac{\varphi(T,\varepsilon)}{\tau}$$

$$(\rho\varepsilon)_t + (\rho u\varepsilon)_x = \delta(\rho\varepsilon_x)_x - \frac{\varphi(T,\varepsilon)}{\tau} \tag{4}$$

$$p = p(T,\rho) , \qquad e = e(T). \tag{5}$$

For the notation in these equations, see the nomenclature at the end of the
paper. The identities in (5) are the equations of states. In particular, for
a polytropic ideal gas,

$$p = RT\rho , \quad e = \frac{R}{\gamma-1}\, T$$

where $\gamma \equiv c_p/c_\nu$ is the specific heat ratio of the gas, and is assumed to be
constant.

The chemical reaction in the combustion process is assumed to be of the
form

$$\text{reactant} \;\rightarrow\; \text{product},$$

and exothermic with a heat release of quantity Q per reaction. Since ε is
the mass fraction of the reactant, the law of mass action demands that the
reaction term in (3) and (4) be of the form

$$\varphi(T,\varepsilon) = z\,\varepsilon A(T) ,$$

36

where the reaction rate $A(T)$ is assumed to be of Arrhenius type with an ignition temperature T_i :

$$A(T) = \begin{cases} 0 & , \text{ if } 0 < T < T_i \\ \exp\left(-\dfrac{E}{RT}\right), & \text{ if } T > T_i \, . \end{cases} \qquad (6)$$

In general, T_i is obtained from actual experiment.

The gas mixture is also assumed to be in exact stoichiometric ratio so that, when the chemical reaction has completed, only the products remain.

A discussion of these model equations and other ommitted mechanisms can be found in [6].

III. THE SIMPLIFIED MODEL EQUATIONS FOR FLAME PROPAGATION. It is well-known that a flame front in a typical (hydrocarbon) gas mixture propagates into the quiet unburned gas with a speed having order of magnitude 100 cm/sec. This speed of propagation is highly subsonic. This fact is consistent with an analysis of weak deflagrations in the ZND model [2]. It is also well-known that, in order to predict correct flame speed, it is necessary to take account of the internal mechanisms of the reaction zone. That is, the effects of heat conduction or chemical species diffusion are essential to the formation of the flame fronts. Therefore, to obtain a simplified model for flame propagation from (1)-(5), we shall assume that

   (a) the gas is non-viscous ($\nu = 0$ in (2), (3)),

   (b) the gas is incompressible, and

   (c) the typical length scale for the reaction zone $L_0 \equiv (\delta\tau)^{1/2}$ is much smaller than a typical length scale of the environment.

Then, if we introduce the dimensionless time and length scale

$$\hat{t} = \frac{t}{\tau} \, , \quad \hat{x} = \frac{x}{L_0} = \frac{x}{\sqrt{\delta\tau}} \, ,$$

37

and form an asymptotic expansion with respect to the inner scale L_0, the equations (1)-(4) become

$$\rho_{\hat{t}} + (\rho u)_{\hat{x}} = 0$$

$$p_{\hat{x}} = 0$$

$$\rho c_p T_{\hat{t}} + \rho u c_p T_{\hat{x}} = \left(\frac{\lambda}{\delta} T_{\hat{x}} \right)_{\hat{x}} + Q \, \varphi(\varepsilon, T)$$

$$\rho \varepsilon_{\hat{t}} + \rho u \varepsilon_{\hat{x}} = (\rho \varepsilon_{\hat{x}})_{\hat{x}} - \varphi(\varepsilon, T).$$

Thus, within the reaction zone, the pressure p remains constant.

In this report, we will study this simplified system. We will from now on write x for $\hat{x}$ and t for $\hat{t}$, and will rewrite the simplified equations in the form:

$$\rho_t + (\rho u)_x = 0 \tag{7}$$

$$\rho c_p T_t + \rho u c_p T_x = \left(\frac{\lambda}{\delta} T_x \right)_x + Q \, \varphi(\varepsilon, T) \tag{8}$$

$$\rho \varepsilon_t + \rho u \varepsilon_x = (\rho \varepsilon_x)_x - \varphi(\varepsilon, T). \tag{9}$$

As for the equations of state in (5), the assumption (6) implies that the gas can expand only due to temperature increase. Thus

$$\rho = \rho(T) , \qquad \rho'(T) < 0 . \tag{10}$$

IV. TRAVELING WAVE SOLUTIONS INTERPRETED AS FLAME FRONTS. The flame fronts travel through the gas mixture with a definite speed and burn the unburned gas to a definite burned gas state. Therefore we shall interpret the flame fronts as traveling wave solutions of (7) - (10). Thus we look for solutions of the form

$$\begin{pmatrix} u \\ T \\ \varepsilon \end{pmatrix} (x,t) = \begin{pmatrix} u \\ T \\ \varepsilon \end{pmatrix} (\xi), \qquad \xi = x + Vt , \qquad (11)$$

satisfying the boundary conditions

$$\begin{pmatrix} u \\ T \\ \varepsilon \end{pmatrix} (-\infty) = \begin{pmatrix} u_0 \\ T_0 \\ 1 \end{pmatrix} , \qquad \begin{pmatrix} u \\ T \\ \varepsilon \end{pmatrix} (\infty) = \begin{pmatrix} u_1 \\ T_1 \\ 0 \end{pmatrix} , \quad T_0 < T_i < T_1 .$$

$$(12)$$

V in (11) is an unknown quantity which represents the speed of the flame. That $\varepsilon = 1$ and $T_0 < T_i$ in (12) shows that the gas is unburned at $\xi = -\infty$. Similarly, the gas is completely burned at $\xi = \infty$. Thus we may take $V > 0$. The main result is:

Theorem 1: Assume that $\lambda(T)$ is bounded. Then, given any unburned gas state at $\xi = -\infty$, there is a unique $V > 0$ and a corresponding unique burned gas state at $\xi = \infty$ such that equations (7)–(10) have a unique traveling wave solution of the form (11), (12).

Furthermore, $T_1 = T_0 + \dfrac{Q}{c_p}$ and there exists a unique number $m > 0$ such that

$$m = \rho(T(\xi))(u(\xi) + V) \quad \text{for all} \quad \xi .$$

Thus, the burned gas state is specified by the relations

$$V = \frac{m}{\rho_0} - u_0 , \qquad u_1 = \frac{m}{\rho(T_1)} - V .$$

The uniqueness in the theorem depends very much on the assumption that $T_i > 0$. See the remarks in Section VI.

The flame profile obtained in this theorem represents the internal structure of the reaction zone. Outside the reaction zone, the flame structure is mainly gas dynamical. This fact can be used to prove the

existence of weak deflagrations for the whole system (1)-(6).

The proof of this theorem will be published in detail elsewhere. However, we will illustrate the method of proof in a special case.

V. PROOF FOR THE CASE OF THE LEWIS NUMBER EQUAL TO ONE. The Lewis number in (7)-(10) is defined to be

$$L = \frac{\lambda}{\rho\, c_p\, \delta} \; .$$

In this section, we shall assume that $L(T) \equiv 1$ for all T. This assumption will lead to the conservation of the total enthalpy.

Substituting (11) into (7)-(9), we obtain that

$$\rho(T(\xi))(u(\xi) + V) = m = \text{constant} \tag{13}$$

$$c_p m\, T' = \left(\frac{\lambda}{\delta} T' \right)' + Q\, \psi(\epsilon,T) \tag{14}$$

$$m\, \epsilon' = (\rho\epsilon')' - \psi(\epsilon,T). \tag{15}$$

where $T' = \frac{dT}{d\xi}(\xi)$, etc. Application of maximum principles to (14) and (15) implies that, if (13)-(15) and (12) have a solution, then

$$\epsilon'(\xi) < 0 \; , \quad T'(\xi) > 0$$

and $\epsilon(\xi), T(\xi)$ are bounded. Thus,

$$\epsilon'(\pm\infty) = T'(\pm\infty) = 0 \; . \tag{16}$$

Now, from (14) and (15),

$$m(c_p T' + Q\epsilon') = \left(\frac{\lambda}{\delta} T' + Q\rho\epsilon' \right)' \; . \tag{17}$$

Integrating (17) from $\xi = -\infty$ to general ξ yields

$$m(c_p T + Q\epsilon) = \left(\frac{\lambda}{\delta} T' + Q\rho\epsilon' \right) + m(c_p T_0 + Q). \tag{18}$$

Similarly, integrating (17) from $\xi = \infty$ to general ξ gives

$$m(c_p T + Q\epsilon) = \left(\frac{\lambda}{\delta} T' + Q\rho\epsilon' \right) + m\, c_p T_1 \; . \tag{19}$$

Thus if (12)-(15) have a solution, a comparison of (18) and (19) gives

$$c_p T_0 + Q = c_p T_1 \; . \tag{20}$$

T_1 is uniquely determined by T_0 ; (20) is true no matter whether $L = 1$ or not. Note that we have used (16) in the derivation of (17) and (18).

Suppose that $L = 1$, i.e.

$$\frac{\lambda}{\delta} = c_p \rho \; ;$$

then the quantity

$$H = c_p T + Q\epsilon$$

will satisfy

$$mH = \rho H' + mH(-\infty) \; ,$$

$$H(-\infty) = H(\infty) \; .$$

This follows from (18) and (20). Obviously,

$$H(\xi) = H(\infty) \quad \text{for all} \; \xi \; . \tag{21}$$

Thus $H(\xi)$ is conserved during the combustion process. The quantity H is the total enthalpy of (7)-(10).

Using (21), (14) and (15) can be combined into a simple nonlinear eigenvalue problem

$$c_p m T' = c_p (\rho(T)T')' + Q \psi \left(\frac{c_p}{Q} (T_1 - T), T'\right) ,$$

$$T(-\infty) = T_0 , \qquad T(\infty) = T_1 ,$$

where m appears as the eigenvalue. That this problem has a unique solution m and $T(\xi)$ follows from a well-known phase plane analysis first rigorously discussed by Gel'fand [3]. Also, see [5].

For arbitrary Lewis number L, the Schauder fixedpoint theorem is used to prove the existence of solution to (12)-(15). The main estimate needed to establish the applicability of Schauder fixed point theorem is to estimate the total enthalpy H in terms of L; the latter is only constant when $L = 1$.

VI. COLD BOUNDARY DIFFICULTY AND OTHER REMARKS. The determination of the ignition temperature T_1 is somewhat arbitrary. Strictly speaking, the gas is not in stable chemical equilibrium even if $T_0 < T_1$. The gas is always in a "metastable" state even at low temperature. However, without the

assumption that $T_i > 0$, the problem (12)-(15) would not be well-posed. This "cold boundary difficulty" is well-discussed in [6].

In the model (7)-(10), the cold boundary difficulty can be discussed as follows. Assuming that $T_0 = 0$ in (12), one can establish

Proposition: If the unburned state in (12) is fixed, then

$$\lim_{T_i \to 0} V(T_i) = V_0 \qquad .$$

exists.

Thus, no matter how small the ignition temperature is, the gas can always support a flame front with a definite speed. It seems that the assumption $T_i > 0$ is immaterial.

However, if we don't assume the existence of an ignition temperature, i.e. instead of (6), we assume

$$A(T) = \exp\left(-\frac{E}{RT} \right) \qquad \text{for all} \quad T ,$$

then one can show

Theorem 2: Fix the unburned state in (12) with $T_0 = 0$; then there exists V_0 such that (7)-(12) have a solution iff $V > V_0$.

Thus, without the assumption of an ignition temperature, the observed flame front tends to be unstable. Its speed tends to fluctuate. The cold boundary difficulty actually occurs.

Mathematically, the cold boundary difficulty is due to the extremely singular behavior of the function $\exp\left(-\frac{E}{RT}\right)$ around $T = 0$. This fact also leads to difficulty in computing the flame speed for certain gas mixture.

VI. CONCLUSION. The model equations discussed in this report take care of the combined effects of gas expansion due to the temperature increase after combustion and the transport processes. We show that the flame fronts exist in this model, and we discussed some of their properties.

This model is exclusively used to discuss flame propagation (deflagrations); the model system (7)-(10) is not appropriate for a discussion of detonation waves. We shall show in a future paper that it leads to a discussion of flame propagation for the full set of equations (1)-(5) where all the effects of gas dynamics are incorporated.

NOMENCLATURE

ρ	Gas density
u	Gas velocity
p	Gas pressure
e	specific internal energy
T	Gas Temperature
ε	mass fraction of the reactant
ν	coefficient of shear viscosity
λ	coefficient of heat conduction
δ	coefficient of chemical species diffusion
τ	typical reaction time of the chemical reaction
Q	heat content of the chemical reaction
c_ν	specific heat at constant volume
c_p	specific heat at constant pressure
R	universal gas constant
E	activation energy of the chemical reaction.

References

1. S. Z. Burnstein, P. D. Lax, G. A. Sod [ed.], _Lectures on Combustion Theory_, Courant Mathematics and Computing Laboratory, New York University (1978).

2. R. Courant and K. O. Friedrichs, _Supersonic Flow and Shock Waves_, Springer-Verlag, New York (1948).

3. I. M. Gel'fand, _AMS Translations_, 29, 295 (1963).

4. I. Glassman, _Combustion_, Academic Press, New York (1977).

5. S. S. Lin, _Ph. D. Thesis_, University of California, Berkeley (1979).

6. F. A. Williams, _Combustion Theory_, Addison-Wesley, Reading, Masachusetts (1965).

DEVELOPMENT OF DEFLAGRATION ON INITIALLY COLD COMBUSTIBLES

A. K. Kapila
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, New York   12181

ABSTRACT Deflagration waves may be generated in com-
bustible materials in several different ways, which include
self-heating, application of external thermal stimulus or
increasing the Damkohler number above the critical. The
corresponding transients are compared, with special emphasis
on the third mode of combustion initiation.

I. Introduction Burning can be initiated in a variety
of ways in materials that combust as a result of thermally
accelerated exothermic reactions. Self-induced burning
(mode I) can occur if chemical heating overcomes heat loss
to the environs. Otherwise, an external stimulus, such as
heat flux at the surface, can be applied (mode II). Alter-
natively, the reaction rate can be enhanced by increasing
the Damkohler number above the critical (mode III), by
raising the pressure, for example.

Recently, for modes I and II, Kapila ([1] and [2]) has
described the transients that lead to the establishment of
combustion waves in nondeformable materials. For mode I the
evolutionary process consists of a mildly reactive induction
stage which ends in thermal runaway, followed by a brief ex-
plosion period in which a rapidly intensifying hot spot
develops. Upon maturity, the hot spot is transformed into a
propagating wave. The situation is essentially the same for
mode II, except that there is an initial period of inert
heating and the weak reaction responsible for thermal runaway
occurs in a thin surface layer. In this paper we discuss how
thermal runaway occurs in mode III; events subsequent to run-
away follow the same course as in modes I and II. Only a
brief description is given here, since details can be found
in [3].

II. Formulation Let a combustible be confined to the
region between the planes $x = \pm 1$. Let the boundaries of
the region be maintained at constant levels of temperature
and reactant concentration (i.e. heat, fresh mixture and com-
bustion products are allowed to cross the boundaries). Taking
unit Lewis number and invoking symmetry about $x = 0$, the
mathematical problem to be considered is

$$z = (1+\beta-y)/\beta, \tag{1}$$

$$y_\tau = y_{xx} + [D/(\beta\gamma)] (1+\beta-y) \exp(\gamma-\gamma/y), \quad 0<x< 1, \tag{2}$$

$$y_x(0,\tau) = 0, \quad y(1,\tau) = 1, \tag{3}$$

with appropriate initial conditions. This dimensionless system describes a single, one-step Arrhenius reaction. Here y is the temperature, z the reactant mass fraction, β the chemical heat release, γ the activation energy and D the Damkohler number.

The relevant static problem has been studied by Kapila and Matkowsky [4] in the limit $\gamma\to\infty$. The steady-state response diagram, i.e. a plot of y at $x=0$ against D, is seen to be the S-shaped curve of Fig. 1. The upper and lower branches are found to be asymptotically stable and the middle branch unstable, the exchange of stabilities occuring precisely at the turnaround points of the S. An analytical description of the entire response is given in [4] where it is shown, in particular, that on the lower branch AC, the appropriate expansions for y and D are

$$y = 1+\gamma^{-1}y_1 + O(\gamma^{-2}), \quad D = D_1+O(\gamma^{-1}). \tag{4}$$

Furthermore, the solution y_1 is given by

$$y_1(x) = H(x;D_1) \tag{5}$$

where H has the parametric representation

$$H = 2 \ln [\cosh \alpha \; \text{sech} (\alpha x)], \quad D_1 = 2\alpha^2 \text{sech}^2\alpha \tag{6}$$

and the parameter α increases from $\alpha=0$ at A to $\alpha=\alpha_c$ at C, where

$$\alpha_c \tanh \alpha_c = 1 \quad (\alpha_c \approx 1.2)$$

and correspondingly,

$$D_{1_c} \approx 0.88, \quad y_{1_c}(0) \approx 1.187.$$

The analysis in [4] also shows that at the point F (vertically above C) on the upper branch in Fig. 1,

$$y_F(0) = 1+\beta - \text{est} \tag{7}$$

where est stands for exponentially small terms in the limit $\gamma\to\infty$.

Before proceeding further we observe that the exchange of stabilities at C occurs due to the passage of the largest eigenvalue through zero. The corresponding eigenfunction of the linearized problem, given to leading order by

$$E(x) = 1 - \alpha_c x \tanh(\alpha_c x), \qquad (8)$$

plays an important role in the sequel.

III. The Dynamic Response

Of interest is the dynamic behavior of y as D varies slowly through D_c. This slow variation can be characterized explicitly by introducing a slow time t, where

$$t = \delta\tau, \qquad \delta \ll 1,$$

thereby transforming (2) into

$$\delta y_t = y_{xx} + [D(t)/(\beta\gamma)] (1+\beta-y) \exp(\gamma-\gamma/y) . \qquad (9)$$

At an initial time t_B let the system be at the state corresponding to the point B in Fig. 1. Following (4) we let

$$D(t) = D_1(t) + O(\gamma^{-1}) \qquad (10)$$

and assume that $D_1(t)$ is a smooth, monotonically increasing function which has the power series representation

$$D_1(t) = D_{1_c} [1+t+O(t^2)] \text{ as } t \to 0. \qquad (11)$$

This specifies $t=0$ to be the point at which the critical point C is reached. The goal is to obtain an asymptotic solution to (9) and (3), with the initial condition specified above, in the limit $\delta \to 0$, $\gamma \to \infty$. We shall concentrate on the limit $\delta \gg \gamma^{-1}$. Several different regimes need to be distinguished.

III.A Precritical Solution

Expecting y to stay close to unity prior to criticality, we let

$$y = 1 + \gamma^{-1}z(x,t) \qquad (12)$$

whence (9) and (3) reduce to

$$\delta z_t = z_{xx} + D_1 e^z + O(\gamma^{-1}), \quad z_x(0,t) = z(1,t) = 0. \qquad (13)$$

The expansion

$$z = z_1 + \delta z_2 + o(\delta) \qquad (14)$$

then leads, to leading order, to the pseudo-steady solution

$$z_1 = H(x; D_1(t)). \qquad (15)$$

In view of (11), it can be shown that

$$z_1 = z_{1_c}(x) - \frac{2}{\alpha_c} (-t)^{1/2} E(x) + O(t) \text{ as } t \to 0^-, \qquad (16)$$

where

$$z_{1_c}(x) = 2 \ln[\cosh \alpha_c \text{ sech } (\alpha_c x)],$$

and $E(x)$ has been introduced in (8).

The asymptotic behavior (16) indicates that the expansion (14) is not valid beyond t=0. It can be shown that z_2 has the asymptotic behavior

$$z_2 = \frac{1}{3\alpha_c^2} (t)^{-1} E(x) + O[(-t)^{1/2}] \text{ as } t \to 0^-,$$

which merely confirms the breakdown of (14).

III.B Transition Solution

Further development of the solution occurs on a new time scale s defined by the stretching

$$t = \delta^{2/3} b^{-1} s \qquad (17)$$

where the O(1) constant b will be chosen later in a way that simplifies algebra. The transition expansion is taken to be

$$z = z_{1_c}(x) + \delta^{1/3} v_1(x,s) + \delta^{2/3} v_2(x,s) + \delta v_3(x,s) + o(\delta). \qquad (18)$$

We find that v_1 satisfies the homogeneous problem

$$L(v_1) \equiv v_{1_{xx}} + (D_{1_c} \exp z_{1_c}) v_1 = 0, \quad v_{1_x}(0,s) = v_1(1,s) = 0 \qquad (19)$$

while the v_i (i≥2) satisfy the nonhomogeneous problems

48

$$L(v_i) = w_{i-1}, \quad v_{i_x}(0,s) = v_i(1,s) = 0, \tag{20}$$

with w_i linear in v_{i_s} and depending, in addition, upon x, s and v_j ($1 \le j \le i$). The problem (19) has a nontrivial solution

$$v_1 = f_1(s) \, E(x) \tag{21}$$

where the "amplitude function" $f_1(s)$ is determined by requiring that the problem (20) have a solution for $i=2$. The requisite orthogonality condition provides the Riccati differential equation

$$bf_1'(s) = 3b^{-1}s - \frac{3}{4}\alpha_c^2 f_1^2(s), \quad s < \infty,$$

for which the initial condition

$$f_1 \to -\frac{2}{\alpha_c} b^{-1/2} s^{1/2} - \frac{b}{3\alpha_c^2} s^{-1} \quad \text{as } s \to \infty$$

is the result of matching with (14). The choice $b = (9\alpha_c^2/4)^{1/3}$

leads to the solution

$$f_1(s) = 3(9\alpha_c^2/4)^{-2/3} \, Ai'(s)/Ai(s), \tag{22}$$

where $Ai(s)$ is the Airy function. Thus, v_1 is determined completely. The higher-order v_i can be computed in an analogous way. Since $f_1(s)$ has a pole at s_o, where $s_o = -2.3381$ is the first zero of $Ai(s)$, the solution (21), and therefore the expansion (18), is valid only for $s > s_o$.

III.C Post-critical Solution

The breakdown of (18) suggests the stretching

$$s = s_o - 2^{-1/3} \delta^{1/3} (\rho - \rho_o) \tag{23}$$

where the new time scale ρ, is of the same order as the fast time τ and the shift ρ_o, assumed to be $o(\delta^{-1/3})$, is to be determined. We now let z have the expansion

$$z = w_1(x, \rho) + o(1) \tag{24}$$

where w_1 can be shown to satisfy

$$w_{1_\rho} = w_{1_{xx}} + D_{1_c} \exp w_1, \quad 0 < x < 1, \quad \rho > -\infty,$$

$$w_{1_x}(0,\rho) = w_1(1,\rho) = 0 \qquad\qquad (25)$$

$$w_1 = z_{1_c}(x) - \frac{4}{3\alpha_c^2 \rho} E(x) + o(\rho^{-1}) \quad \text{as } \rho \to -\infty.$$

The initial condition in (25) comes from matching with the transition expansion (18). (In order to determine the shift ρ_o and to fix the origin of ρ in (25), higher-order matching with (18) is needed.)

We note that w_1, which evolves at a constant value D_{1_c} of D_1, measures the departure of y from unity on the $O(\gamma^{-1})$ scale (see (12) and (24)). Thus, w_1 is entirely analogous to the induction-period solution for the self-induced combustion case (mode I), discussed in [1]. In fact, the numerical solution of (25) leads to a graph much like that in Fig. 3 of [1]. In other words, the solution develops slowly in the initial stage, but there w_1 begins to rise rapidly near x=0 while variations continue to be leisurely elsewhere. Eventually, at a definite $\rho = \rho_\infty$, $w_1(0,\rho)$ becomes unbounded, signalling thermal runaway and the birth of a hot spot.
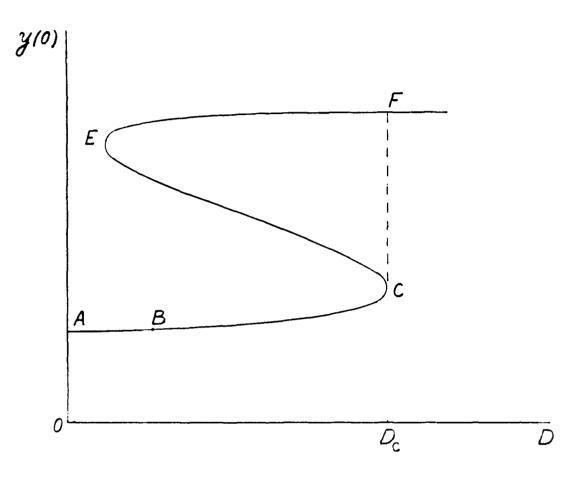
Further development occurs precisely as in [1]. The hot spot intensifies, reaches maturity when the temperature in it has reached the value 1+ε, detaches from x=0 and propagates into the domain. Eventually, the combustion wave comes to rest near x=1 to accommodate the boundary condition there, thereby completing the jump from C to F in Fig. 1. Further movement along the upper branch will again be governed by the slow variable t, much in the manner of the precritical solution.

IV Concluding Remarks

The asymptotic analysis has concentrated on the case when combustion is initiated by the slow passage of the Damkohler number through the critical. Details of the transient upto thermal runaway are given, and it is pointed out that subsequent evolution of the combustion wave is analogous to the case of self-induced burning.

REFERENCES

1. A. K. Kapila, Reactive-diffusive system with Arrhenius
    kinetics: dynamics of ignition, SIAM J. Appl. Math.
    (1980), to appear.

2. A. K. Kapila, Evolution of deflagration in a cold com-
    bustible subjected to a uniform energy flux, Intnl. J.
    Engin. Sci., (1980), to appear.

3. A. K. Kapila, Arrhenius Systems: dynamics of jump due
    to slow passage through criticality, MRC Technical
    Summary Report #2059 (1980), Mathematics Research Center,
    University of Wisconsin.

4. A. K. Kapila and B. J. Matkowsky, Reactive-difusive
    systems with Arrhenius kinetics: multiple solutions,
    ignition and extinction, SIAM J. Appl. Math, 36 (1979),
    pp. 373-389.

Fig. 1. The Steady Response

MATHEMATICAL QUESTIONS FROM COMBUSTION THEORY[*]

G.S.S. Ludford and D.S. Stewart
Cornell University, Ithaca NY 14853

ABSTRACT. Residual mathematical questions from combustion theory are
presented, in particular those relating to a problem discussed by Liñán
(1974). After illustrating the main ideas by means of an exactly integra-
ble model, known results about Liñán's problem are summarized, with indi-
cations of how they are obtained by numerical and asymptotic methods. The
hope is to stimulate further, purely mathematical, work on these questions.

I. INTRODUCTION.

With the advent of activation-energy asymptotics as an effective analy-
tical tool in combustion theory, a host of residual mathematical questions
have appeared. These questions are typically concerned with a differential
problem governing the behavior of a thin reaction-diffusion zone. Existence
of solutions and the determination of various parameters associated with
the differential problem are critical questions in the overall analysis
of the combustion phenomenon.

In such diverse combustion questions as the burning of monopropellant
drops, detonations, fast deflagration waves and counterflow diffusion flames
the structure of the reaction-diffusion zone is governed by Liñán's problem
(1974), i.e. the differential equation

$$(1a) \qquad 2y'' = y \, \exp(\alpha x - y)$$

subject to the boundary conditons

$$(1b) \qquad y' = \begin{cases} -1 + o(1) & \text{as} \quad x \to -\infty \ , \\ \\ o(1) & \text{as} \quad x \to +\infty \ , \end{cases} \qquad (1c)$$

where α is a real constant. In general, the solution must be computed numerically by taking x_0 large and negative, and varying y_0 in the initial conditons

$$(2a,b) \qquad y'(x_0) = -1 \ , \quad y(x_0) = -x_0 + y_0$$

until the condition (1c) is satisfied for some large x. The corresponding y_0 approximates the constant

$$(3) \qquad \lim_{x \to -\infty} (y+x) - y_{-\infty}(\alpha) \ ,$$

which is thereby found. The numerics indicate that there is a unique solution for $\alpha > -1/2$ but none for smaller values of α. Moreover, the constant

$$(4) \qquad \lim_{x \to \infty} y = y_{+\infty}(\alpha)$$

thereby approximated has the properties

$$(5) \qquad y_{+\infty}(\alpha) \begin{cases} = 0 & \text{for} \quad \alpha \geq 0 \ , \\ \\ > 0 & \text{for} \quad 0 > \alpha > -\frac{1}{2} \ , \\ \\ \to \infty & \text{as} \quad \alpha \to -1/2 \ . \end{cases}$$

Existence, uniqueness and properties such as those mentioned above have been established numerically and in some instances by asymptotic analysis. While the combination of these two approaches is adequate in the context

of a specific investigation, the problem is of sufficient frequency and importance to justify deeper mathematical treatment. In the present paper we shall summarize known results about this particular problem in the hope of encouraging further work. We shall also give additional examples of residual problems from combustion theory that are worthy of further analysis.

II. A MODEL PROBLEM.

Here we examine the linear problem

$$(6a,b,c) \qquad 2y'' = y \exp(\alpha x) \ , \quad y' = -1 \ \text{at} \ x = 0 \ , \quad y' = o(1) \ \text{as} \ x \to \infty$$

to illustrate the ideas behind Liñán's. Under the transformation

$$(7) \qquad t = (\sqrt{2}/\alpha)\exp(\alpha x/2)$$

(6) becomes the modified Bessel's equation

$$(8) \qquad \frac{d^2 y}{dt^2} + \frac{1}{t}\frac{dy}{dt} - y = 0 \ ,$$

from which we find the unique solution

$$(9) \qquad y = \begin{cases} -\dfrac{\sqrt{2}K_0(t)}{K_0'(\sqrt{2}/\alpha)} & \text{for} \ \alpha > 0 \ , \\[2em] \sqrt{2}e^{-x/\sqrt{2}} & \text{for} \ \alpha = 0 \ , \\[2em] -\dfrac{\sqrt{2}I_0(t)}{I_0'(\sqrt{2}/\alpha)} & \text{for} \ \alpha < 0 \end{cases}$$

of the problem (6).

In contrast to Linán's problem there is no limitation on α. On the other hand the constants $y_o = y(0)$ and $y_{+\infty} = y(\infty)$ play the role that $y_{-\infty}$ and $y_{+\infty}$ play in Linán's problem, being determined by the solution. We find

$$(10) \qquad y_o = \begin{cases} -\dfrac{\sqrt{2}K_o(\sqrt{2}/\alpha)}{K_o'(\sqrt{2}/\alpha)} & \text{for } \alpha > 0 \ , \\[2em] \sqrt{2} & \text{for } \alpha = 0 \ , \\[2em] -\dfrac{\sqrt{2}I_o(\sqrt{2}/\alpha)}{I_o'(\sqrt{2}/\alpha)} & \text{for } \alpha < 0 \ ; \end{cases}$$

and

$$(11) \qquad y_{+\infty} = \begin{cases} 0 & \text{for } \alpha > 0 \ , \\[1em] 0 & \text{for } \alpha = 0 \ , \\[1em] -\sqrt{2}/I_o'(\sqrt{2}/\alpha) & \text{for } \alpha < 0 \ . \end{cases}$$

For future reference we cite the asymptotic behavior of $y_{+\infty}$ as $\alpha \to 0^-$.

$$(12) \qquad y_{+\infty} \sim -\frac{2^{5/4}\pi^{1/2}}{|\alpha|^{1/2}} \exp(-\sqrt{2}/|\alpha|) \quad \text{as } \alpha \to 0^- \ .$$

Thus the model problem has features similar to Linán's. The given boundary conditions are sufficient to determine the solution completely and hence the constants y_o and $y_{+\infty}$ as functions of α. The constant $y_{+\infty}$, as in Linán's problem, vanishes for $\alpha \geq 0$ and is non-zero for $\alpha < 0$, being controlled by the behaviors of the corresponding Bessel functions for large x.

III. PROPERTIES FOR $\alpha > 0$, ESPECIALLY $\alpha \to +\infty$.

In the following a unique solution is assumed to exist. Of most inter-
est is the determination of the two constants $y_{+,c}(\alpha)$ for a given α.
Unless otherwise stated the analysis in the following sections is essential-
ly due to Liñán (1974).

For $\alpha > 0$ and $\alpha \to \infty$, (1) is effectively replaced by

(13) $2y'' = y \exp(\alpha x)$.

Again using the transformation (7) we reach the modified Bessel's equation
(8). The only bounded solutions for $t \to \infty$ are proportional to $K_o(t)$
so we conclude that

(14) $\lim_{x \to \infty} y = y_{+} = 0$ for $\alpha > 0$.

The result is otherwise obvious by inspection of the problem itself. The
boundary conditions require the solution to tend to a constant as $x \to \infty$,
and that constant must be zero if the differential equation is to be satisfied.

If we consider the limit $\alpha \to \infty$, and in particular focus on the behav-
ior of y for x large and negative, then (13) still governs the behavior
of y. The solution to (13) is again a multiple of $K_o(t)$ and in partic-
ular as $x \to -\infty$, i.e. $t \to 0$, y is approximated by

(15) $A[\ln(t/2) + \gamma)(1+t^2/4+...) + t^2/4 + ...]$

where γ is the Euler constant. The boundary condition (1b) then shows
that $A = 2/\alpha$. Hence

(16) $y = (2/\alpha)K_o(t)$

and thus

(17) $y \sim -x - (\gamma - \ln(\ldots))/\alpha$ as $x \to -\infty$,

from which follows

(18) $y_{-\infty} \sim (\ln\alpha^2 - \gamma)/\alpha$ as $\alpha \to +\infty$.

[The result (17) can be made into a formal asymptotic expansion.]

IV. PROPERTIES FOR $|\alpha| \ll 1$

Expanding y in powers of α shows that the leading-order solution satisfies

(19) $y_0'' = y_0 \exp(-y_0)$

and the boundary conditions (15) and (16). This is the standard structure problem under the combustion approximation (Buckmaster & Ludford, 1982). An integration under the boundary conditions at $x = -\infty$ yields

(20) $(y_0')^2 = 1 - (y_0+1)\exp(-y_0)$.

Another integration now gives

(21) $x = y_{-\infty} - y_0 + \int_{y_0}^{y} \left[\frac{1}{\sqrt{1 - (\eta+1)\exp(-\eta)}} - 1 \right] d\eta$

where $y_{-\infty}$ is an integration constant which may be identified as $y_{-\infty}(0)$ according to the definition (5). Its determination requires consideration of the perturbation, the indeterminacy at this stage being due to the ineffectiveness of the condition at $x = -\infty$; the function (21) satisfies it for any value of $y_{-\infty}$.

In fact the perturbation satisfies

$$(22a,b,c) \qquad 2y_1'' = \exp(-y_o)[y_1 + y_o(x - y_1)] \quad , \quad y_1' = o(1) \quad \text{as} \quad x \to \bar{\mp} \infty \quad ,$$

a problem which, for large positive x, reduces to

$$(23) \qquad 2y_1'' = y_1 + x e^{(\beta - x)/\sqrt{2}} \quad , \quad y_1' = o(1) \quad \text{as} \quad x \to +\infty$$

where $\qquad \beta = y_{-\infty} + \int_0^\infty \left[\dfrac{1}{\sqrt{1 - (1+u)e^{-u}}} - 1 - \dfrac{\sqrt{2}}{u(1+u)} \right] du \quad .$

The general solution of (23) is

$$(24) \qquad y_1 = -\frac{1}{4} \left(\frac{x^2}{\sqrt{2}} + x \right) e^{(\beta - x)/\sqrt{2}} + A e^{-x/\sqrt{2}} + B e^{x/\sqrt{2}} \quad .$$

One would expect that, by choosing $y_1(-\infty)$ appropriately, the increasing exponential could be suppressed, so as to satisfy the boundary condition as $x \to \infty$. However, for general $y_{-\infty}$ this apparently is not the case, the only exception being $y_{-\infty} = 1.344$.

Since $y_{+\infty}(\alpha)$ is identically zero for $\alpha > 0$ its asymptotic expansion for $\alpha \to 0$ has all zero terms. This behavior is completely analogous to the model problem given in Sec. 2 [cf. equation (12)] where $y_{+\infty}$ was shown to be exponentially small. Nevertheless, the exponential behavior of $y_{+\infty}(\alpha)$ as $\alpha \to 0$ in Liñán's problem can be determined by matching. The result is due to Joulin (1979), although he does not give details.

For sufficiently large values of x, the term αx in equation (1) is not a perturbation, as was supposed above. Accordingly we introduce the new variable

$$(25) \qquad \chi = \alpha x - 2 \ell n |\alpha| \quad \text{for} \quad \alpha < 0$$

(the constant being needed for balance) and the expansion

(26) $y = g(\alpha)Y_0(\chi) + \ldots$

Here $g = o(1)$ is a gauge function which is to be determined. The differ-
ential equation (1) then yields

(27) $2d^2Y_0/d\chi^2 = Y_0\exp(\chi)$

and, since $\chi \to -\infty$ as $x \to \infty$, we must have

(28) $Y_0 = I_0(\sqrt{2}\exp(\chi/2))$

when the constant factor is absorbed into g. Thus

(29) $y = g(\alpha)I_0(t) + \ldots$

where

(30) $t = \sqrt{2}\exp(\chi/2)$

is the analog of the variable (7).

Now we match to determine $g(\alpha)$. If the first term of the expansion
(29) is written in terms of the variable x and expanded for $\alpha \to 0$, the
leading term is

(31) $\dfrac{g|\alpha|^{1/2}}{2^{3/4}\pi^{1/2}} \exp(\sqrt{2}/|x|)\exp(-x/\sqrt{2})$.

Likewise, if y_0 [given implicitly by the result (21)] is written in terms
of χ and expanded for $\alpha \to 0$, the leading term is

(32) $\exp(\beta-x)/\sqrt{2}$

(when written in terms of x again). These expressions are identical if

60

$$(33) \qquad g(\alpha) = y_{+\infty} = e^{\delta/\sqrt{2}} \frac{2^{3/4}\pi^{1/2}}{|\alpha|^{1/2}} \exp(\frac{-\sqrt{2}}{|\alpha|})$$

V. PROPERTIES FOR $\alpha < 0$.

Liñán found numerically that no solution exists for $\alpha \leq -1/2$ but did not give a proof. Ludford, Yannitell & Buckmaster (1976) were apparently the first to supply one, at least for $\alpha > -1$; the argument goes as follows.

Let α lie in the range $(-1,0)$ and suppose that a solution exists. Then multiplying (1a) by $\alpha - y'$ and integrating by parts twice, using the boundary conditions, gives

$$(34) \qquad 2\alpha + 1 = (y+1)e^{\alpha x - y} \Big|_{-\infty}^{+\infty} - \alpha \int_{-\infty}^{\infty} e^{\alpha x - y} dx$$

which, under the hypothesis and boundary conditions, becomes

$$(35) \qquad 2\alpha + 1 = -\alpha \int_{-\infty}^{\infty} e^{\alpha x - y} dx \quad .$$

For $\alpha \leq -1/2$ there is a contradiction (in sign for $\alpha < -1/2$) and hence the assumption that a solution exists cannot be correct.

We may expect singular behavior of $y_{+\infty}$ for $|2\alpha+1| = \epsilon \ll 1$ and Liñán has given the analysis. First assume that $|y_{+\infty}| \gg 1$ (it will turn out to be $O(\epsilon^{-1})$), and consider the shift of origin

$$(36) \qquad \chi = x + F(y_{+\infty}) \quad ,$$

where F is to be determined. The expansion

$$(37) \qquad v = y_{+\infty} + y_0 + o(1)$$

then leads to the balanced equation

$$(38) \qquad 2d^2 y_0/d\chi^2 = \exp(-\chi/2 - y_0)$$

if we make the choice

$$(39) \qquad F = 2(y_{+\infty} - \ell n\, y_{+\infty})$$

Setting $\sigma = y_0 + \chi/2$ now gives the differential equation

$$(40) \qquad 2d^2\sigma/d\chi^2 = \exp(-\sigma)$$

subject to the boundary conditions

$$(41a,b) \quad d\sigma/d\chi = -1/2 + o(1) \quad \text{as} \quad \chi \to -\infty \ , \quad d\sigma/d\chi = 1/2 + o(1) \quad \text{as} \quad \chi \to +\infty \ .$$

The solution is

$$(42) \qquad \sigma = 2\ell n[2\cosh(|\chi - \chi_0|/4)]$$

Expanding for $\chi \to -\infty$ shows that y behaves like

$$(43) \qquad y_{+\infty} + y_0 = -x + y_{+\infty} - F + O(1) \ ,$$

so that matching with the boundary condition (1b) requires

$$(44) \qquad y_{-\infty} = -y_{+\infty} + 2\ell n\, y_{+\infty} + O(1) \ .$$

Expanding for $\chi \to \infty$ shows that

$$(45) \qquad y \sim y_{+\infty} + O(1) \ .$$

Equation (42) leads to a uniformly valid approximation for $v(x)$.

To complete the analysis we need only determine $v_{+\infty}$ as a function of ϵ. This is done simply from the formula (35) by using the approximation in the integral; we find $v_{+\infty} = 1/\epsilon$.

VI. OTHER PROBLEMS.

One of the first (and few) questions treated analytically before the advent of activation-energy asymptotics was the diffusion flame of Burke & Schumann (1928). While it is unnecessary to discuss the structure of the flame sheet to obtain the main results, there is always the possibility that one may not exist (which would vitiate the whole disucssion).

The differential equation is

$$(46) \qquad y'' = x^2 - y^2$$

and the boundary conditions are

$$(47) \qquad y = \pm x + o(1) \quad \text{as} \quad x \to \mp\infty \ .$$

One would expect the weaker boundary conditions

$$(48) \qquad y' = \pm 1 + o(1) \quad \text{as} \quad x \to \mp\infty$$

to be sufficient to determined the solution, supposing it exists. If so, does that solution satisfy the stronger conditions? Again one would expect so: the only linear functions that satisfy the differential equation are $y = \pm x$. We are left with the questions of existence and uniqueness.

Numerics leave little doubt about existence. The computation is started at $x = -x_o$, where x_o is large, with $y = -x_o - \epsilon$ and $y' = 1$. The small positive number ϵ is then adjusted until $y' = -1$ at $x = +x_o$. Moreover, linearization of the differential equation makes plausible that there is a family of solutions having the asymptote $y = x$; presumably one of them (at least) satisfies the right boundary condition. Uniqueness is in doubt, however; since the Conference, Professor Alexander has apparently found a second numerical solution.

Another problem concerns the response of a steady combustion process, which is often the only feature of interest. One parameter (e.g. burning rate) is determined as a function of another (e.g. pressure). Such a response can sometimes be multivalued and, to decide which of the possibilities occurs in practice, the stability of the steady states is often invoked.

One of the first analytical discussions of such stability has recently been made by Matalon & Ludford (1980), in the context of a chambered diffusion flame. The steady states $y_s(x)$ near the so-called ignition point are solutions of the differential problem

$$(49a,b,c) \qquad y_s'' + y_s' + Qe^{-x}(1-e^{-x})e^{y_s} = 0 \quad , \quad y_s(0) = y_s(\infty) = 0 \quad .$$

Here Q is a positive parameter and the problem is found (numerically) to have two solutions for Q less than some value Q_o, one for $Q = Q_o$ and none for $Q > Q_o$.

While a proof of these results is of interest in itself (and would be a necessary preliminary) the more important question concerns the differential problem

$$(50a,b,c) \qquad y'' + y' + [\lambda + Qe^{-x}(1-e^{-x})e^{y_s}]y = 0 \quad , \quad y_s(0) = y_s(\infty) = 0$$

governing the stability. For each $Q < Q_o$ and each of the two steady states $y_s(x)$ associated with that Q, we wish to know whether there is a negative eigenvalue λ (implying instability). Using a Galerkin method is open to question because the spectrum is known to be complex; and a major step was to show that the portion of the spectrum with negative real part was in fact real, so validating the method. Numerically it was found that

one of the steady states, for each $Q < Q_o$, is stable while the other is unstable.

The problem obtained on replacing the boundary condition (50c) by

$$(52) \qquad \int_0^\cdot y^2(x)\,dx < \infty$$

has been investigated mathematically (Coddington & Levinson, 1955), although the type of information we seek does not seem to be available. The eigenvalue problem (51) has apparently not been treated; and ones of similar form will undoubtedly arise as the stability of combustion processes is pursued further. [Cf. Taliaferro, Buckmaster & Nachman (1981).]

REFERENCES.

Buckmaster, J. & Ludford, G.S.S. 1981 Theory of Laminar Flames. Cambridge University Press.

Burke, S.P. & Schumann, T.E.W. 1928 Diffusion flames. Indust. Engrg. Chem. 20, 998-1004.

Coddington, E.A. & Levinson, N. 1955 Theory of Ordinary Differential Equations, p. 231. Oxford University Press.

Joulin, G. 1979 Existence, Stabilité et Structuration des Flammes Premelangées. Doctoral Thesis: University of Poitiers.

Liñán, A. 1974 The asymptotic structure of counterflow diffusion flames for large activation energy. Acta Astronaut. 1, 1007-1039.

Ludford, G.S.S., Yannitell, D.W. & Buckmaster, J.D. 1976 The decomposition of a cold monopropellant in an inert atmosphere. Combust. Sci. Tech. 14, 133-145.

Matalon, M. & Ludford, G.S.S. 1980 On the near-extinction stability of diffusion flames. To appear in Int. J. Engrg. Sci.

Taliaferro, S., Buckmaster, J. & Nachman, A., Submitted for publication.

POSTSCRIPT.

Professor Liñán has recently sent us the following analytical determination of the constant $y_{-\infty}$ in Sec. IV.

The equation

$$(y_0' y_1')' = (y_0'' y_1)' + x y_0' y_0''$$

is easily seen to be a consequence of equations (19) and (22a); the boundary conditions (1b,c) and (22b,c) therefore imply

$$\int_{-\infty}^{\infty} x y_0' y_0'' \, dx = 0 \quad ,$$

a restriction on y_0 for there to be a solution of the problem (22). Now y_0 depends on $y_{-\infty}$, so that this restriction is actually an equation for $y_{-\infty}$.

To obtain its solution explicitly, integrate by parts to find

$$\int_{-\infty}^{\infty} x y_0' y_0'' \, dx = \left[\tfrac{1}{2}(x y_0'^2 + y_0) \right]_{-\infty}^{\infty} - \tfrac{1}{2} \int_{-\infty}^{\infty} (y_0' + 1) y_0' \, dx$$

$$= - \tfrac{1}{2} y_{-\infty} + \tfrac{1}{2} \int_0^{\infty} (y_0' + 1) \, dy_0$$

(cancelling terms have been added to ensure finiteness). Here the exponential smallness of the integral in the solution (21) at $x = -\infty$ and of y_0 itself at $x = +\infty$ have been used. Setting the last expression equal to zero now gives

$$y_{-\infty} = \int_0^{\infty} [1 - \sqrt{1 - (y_0 + 1)\exp(-y_0)}] \, dy_0 \quad .$$

Numerical quadrature shows that the integral is 1.344, in agreement with Sec. IV.

NOTE ON THE STABILITY OF STOCHASTIC
REACTION-DIFFUSION EQUATIONS[*]

P. L. Chow
Department of Mathematics, Wayne State University
Detroit, Michigan 48202

ABSTRACT. We consider a class of initial-boundary value problems for reaction-diffusion equations, subject to random parametric excitations. If the unperturbed system is assumed to be stable in the sense of Liapunov, the effects of random perturbation on the stability of such system is examined. By the theory of random evolution equations, the stochastic stability of equilibrium solution will be discussed. The stability criteria are based on the construction of appropriate Liapunov functionals. The theory will be applied to several concrete examples.

I. INTRODUCTION. In a recent paper [1], we introduced a Liapunov method for studying the stability of nonlinear stochastic evolution equations. It was pointed out that the method is applicable to reaction-diffusion systems under random perturbation. Here we shall briefly review the general stability theory for stochastic equations, and then apply it to several randomly perturbed reaction-diffusion equations arising from chemical reaction and biological system, taken from our papers [1] and [2].

Let D be a domain in R^n for $n = 2$ or 3, with a smooth boundary ∂D. Denote the concentration of i-th chemical or biological species at the instant t and the position $x \in D$, by $u_i(t,x)$, $i = 1,2,\ldots,m$.

Consider the specific reaction-diffusion system with a random drift velocity as follows

$$\frac{\partial u_i}{\partial t} = \nu_i \, \Delta \, u_i + N_i(u_1, u_2, \ldots, u_m) - \sum_{j=1}^{} \xi_j(t, x, w) \frac{\partial u_i}{\partial x_j} \quad \text{in} \quad D$$

$$u_i(0, x) = f_i(x) \, , \tag{1}$$

$$a_i \frac{\partial u_i}{\partial n} + b_i u_i \Big|_{\partial D} = 0$$

where ν_i's are the diffusion coefficients; N_i's the reaction functions; ξ_j's the random (turbulent) velocity components, and f_i's are given functions.

We shall try to answer the following question: Suppose that $u_i = \tilde{u}_i(x)$ is a stable equilibrium solution of the system (1) when $\xi_j \equiv 0$. What is the effect of the random perturbation ξ_j on the stability of $\tilde{u}_i$?

II. LIAPUNOV's STABILITY CRITERIA. Let the unperturbed system have a solution belonging to a subspace X of the Hilbert space H of square-integral, m-vector valued functions on D. Suppose that $\phi(u)$ is a smooth functional on H with locally bounded (Frechet) derivatives $\phi'(u)$, $\phi''(u)$ among other properties (see [1]). If $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is a H-valued white noise $\dot{W}(t, x)$ with a covariance operator Q on H, then we define, for $v \in X$

$$\mathcal{L}\phi(v) = (A(v), \phi'(v)) + \frac{1}{2} \text{Trace}\{\phi''(v)B(v)QB^*(v)\} \tag{2}$$

where $A(v) = \Delta \, \text{diag}(\nu_1 v_1, \ldots, \nu_m v_m) + (N_1(v), \ldots, N_m(v))^T$,

$$B(v) = - \left(\frac{\partial v_i}{\partial x_j}\right)^{m \times n}, \tag{3}$$

and $(\cdot, \cdot)$ means the pairing between X and its dual X^*. A functional

ϕ on H is said to be a Liapunov functional for the system (1) if $\phi(v)$ is positive-definite and

$$\mathscr{L} \, \phi(v) \leq 0 , \quad \text{for all} \quad v \quad \text{in} \quad X . \tag{4}$$

The following stability criteria will be useful, and the proof can be found in [1]:

<u>Stability Theorem</u>: Suppose that $u = 0$ is an equilibrium solution of the reaction-diffusion system (1). (i) If there exists a Liapunov functional ϕ satisfying the property (4), the null solution is (almost surely) a.s. stable, that is, for every initial state $u(o,x)$ in H ,

$$\text{prob.}\{\sup_{t>0} \|u(t,\cdot)\| < \infty\} = 1 .$$

(ii) If, in addition, ϕ satisfies $\lim_{\|u\| \to \infty} \phi(u) = \infty$ and

$$\mathscr{L} \, \phi(v) \leq k \, \phi(v) \quad \text{for all} \quad v \quad \text{in} \quad X \quad \text{and some} \quad k > 0 ,$$

then the null solution of (1) is a.s. asymptotically stable,

$$\text{prob.}\{\sup_{t>0} \|u(t,\cdot)\| = 0\} = 1 .$$

Here $\|\cdot\|$ denotes the H-norm:

$$\|v(x)\|^2 = \int_D |v(x)|^2 dx . \tag{5}$$

III. <u>STABILITY OF REACTION-DIFFUSION EQUATIONS</u>. We shall apply the stability theorem stated above to three specific problems as illustrative examples, though they are of independent interest.

(Example 1). Consider the scalar random diffusion problem arising from population biology [2] in R^3:

$$\frac{\partial u(t,x)}{\partial t} = \nu \, \Delta \, u - \alpha \, \frac{u}{1+|u|} - \sum_{j=1}^{3} \xi_j(t,x) \frac{\partial u}{\partial x_j} \;, \quad x \; \text{in} \; D \;,$$

$$u(o,x) = u_o(x) \;, \tag{6}$$

$$u(t,x)\big|_{\partial D} = 0 \;.$$

In this case, we have $A(v) = \nu \, \Delta \, v - \alpha \, \dfrac{v}{1+|v|}$, and $B(v) = (\dfrac{\partial v}{\partial x_1}, \dfrac{\partial v}{\partial x_2}, \dfrac{\partial v}{\partial x_3})..$

Let $\phi(v) = \|v\|^2$. Then by (2),

$$\mathcal{L} \, \phi(v) = -2\int_D \{\nu |\nabla v(x)|^2 + \alpha \frac{v^2(x)}{1+|v(x)|}$$
$$- \frac{1}{2} \sum_{i,j=1}^{3} q_{ij}(x,x) \frac{\partial v(x)}{\partial x_i} \frac{\partial v(x)}{\partial x_j} \} dx \;, \tag{7}$$

where $q_{ij}(x,y)$ are the kernel for the covariance operator Q of the random functions $\xi_i(t,x)$. Let q_{ij} be bounded continuous for x,y in D, with

$$q_o = \sup_{\substack{1 \le i,j \le 3 \\ x,y \in D}} |q_{ij}(x,y)| \;. \tag{8}$$

Then, from (7), we have

$$\mathcal{L} \, \phi(v) \le -2\int_D \{(\nu - q_o/2)|\nabla v(x)|^2 + \alpha v^2(x)\} dx$$
$$\le -2[\lambda_o(\nu - q_o/2) + \alpha] \, \phi(v) \tag{9}$$

where

$$\lambda_o = \inf_{v \in X} \frac{\|\nabla v\|^2}{\|v\|^2} \;. \tag{10}$$

Thus, by the stability theorem (ii), we have

70

$$\lim_{t \to \infty} \int_D |u(t,x)|^2 dx = 0$$

with probability one, if $\lambda_o(\nu - q_o/2) + \alpha > 0$.

(Example 2). In this case, we assume that the system (1) is linear, that
is, the reaction function $N(u) = \Gamma u$, where Γ is a constant $m \times m$ matrix:

$$\Gamma = [r_{ij}]^{m \times m} \tag{11}$$

Let X be the solution space in $L^2(D)$ with

$$\|v\|_X = \int_D \{|v(x)|^2 + |\nabla v(x)|^2\} dx \quad . \tag{12}$$

Again the obvious choice for the Liapunov functional is $\phi(v) = \|v\|^2$. It is
not difficult to compute

$$\mathcal{L} \phi(v) = -2 \int_D \{ \sum_{i=1}^{m} \nu_i |\nabla v_i(x)|^2 - \sum_{i,j=1}^{m} r_{ij} v_i(x) v_j(x)$$

$$- \frac{1}{2} \sum_{i,j=1}^{m} \sum_{k,\ell=1}^{3} q_{k\ell}(x,x) \frac{\partial v_i(x)}{\partial x_k} \frac{\partial v_j(x)}{\partial x_\ell} \} dx \tag{13}$$

$$\leq -2[\lambda_o(\nu_o - q_o/2) - r_o] \phi(v) ,$$

where λ_o , q_o are defined as before, $\nu_o = \min_{1 \leq i \leq m} \{\nu_i\}$, and r_o is the largest
eigenvalue of Γ . According to the Stability Theorem (ii), the null solution
of the linearized equation (1) is asymptotically stable almost surely,
provided that

$$\lambda_o(\nu_o - q_o/2) > r_o .$$

(Example 3). The following reaction-diffusion system occurs in the problem
of two competing species [6]:

71

$$\frac{\partial u_1}{\partial t} = \nu_1 \, \Delta \, u_1 - (1+u_1)u_2 - (\xi \cdot \nabla)u_1 \, ,$$

$$\frac{\partial u_2}{\partial t} = \nu_2 \, \Delta \, u_2 + (1+u_2)u_1 - (\xi \cdot \nabla)u_2 \, , \quad \text{in} \quad D \, , \tag{14}$$

$$\frac{\partial u_i}{\partial n} \bigg|_{\partial D} = 0 \, , \quad i = 1,2$$

Without the random drift, the unperturbed system was treated in [3]. Note that

$$E(v_1, v_2) = (v_1 + v_2) - \ell n(1+v_1)(1+v_2) \tag{15}$$

is a first integral for the reduced ordinary different equation for (14) with $\nu_1 = \nu_2 = 0$ and $\xi \equiv 0$. Introduce the functional ϕ defined by

$$\phi(u) = \int_D E[u(t,x)]dx \tag{16}$$

where $u(t,x)$ is a solution of the system (14). By a direct computation, invoking (15) and the divergence theorem, we have

$$
\begin{aligned}
\mathcal{L} \, \phi(u) = -\int_D \Big\{ & \sum_{i=1}^{2} \sum_{j=1}^{3} \nu_i \Big(\frac{\partial u_i}{\partial x_j}\Big)^2 \Big(\frac{1}{1+u_i}\Big)^2 \\
& - \frac{1}{2} \sum_{i,j=1}^{2} \sum_{k,\ell=1}^{3} q_{k\ell}(x,x) \frac{\partial u_i}{\partial x_k} \frac{\partial u_j}{\partial x_\ell} \Big(\frac{1}{1+u_i}\Big)\Big(\frac{1}{1+u_j}\Big) \Big\}dx \\
\leq & -\int_D (\nu_0 - \frac{1}{2} q_0) \sum_{i=1}^{2} \sum_{j=1}^{3} \Big[\frac{\partial}{\partial x_j} \ell n(1+u_i)\Big]^2 \, .
\end{aligned}
\tag{17}
$$

Then $\phi(u)$ becomes a Liapunov functional if

$$\nu_0 > \frac{1}{2} q_0$$

where $\nu_0 = \min(\nu_1, \nu_2)$. Thus, in view of the Stability Theorem (i), the null solution of the system (14) is stable almost surely if (17) holds, but

72

need not be asymptotically stable.

(Example 4). As a last example, we consider the stability of the nonlinear diffusion equations:

$$\frac{\partial u_1}{\partial t} = \nu_1 \Delta u_1 - au_1 + bf(u_2) - (\xi \cdot \nabla)u_1 \ ,$$

$$\frac{\partial u_2}{\partial t} = \nu_2 \Delta u_2 + cu_1 - df(u_2) - (\xi \cdot \nabla)u_2 \ , \quad \text{in} \quad D \ , \tag{18}$$

$$\left.\frac{\partial u_i}{\partial n}\right|_{\partial D} = 0 \ ,$$

where a,b,c,d are positive constants, and $f \geq 0$ with $f(0) = 0$ and $|f'(s)| \leq M$ for $-\infty < s < \infty$.

The associated ordinary differential equations form a Lure's system [4] in the feed-back control theory. A Liapunov function for such a system is

$$E(v) = \frac{1}{2} r^2 v_1^2 + F(v_2) \tag{19}$$

with $F(_2) = \int_0^{v_2} f(s)ds$ and $r = \frac{\sqrt{ad} + \sqrt{ad - bc}}{b}$, $ad > bc$.

Let

$$\phi(u) = \int_D E[u(t,x)]dx \ . \tag{20}$$

Then

$$\mathcal{L}\,\phi(u) = -\int_D \{\nu_1 r^2 |\nabla u_1|^2 + \nu_2 f'(u_2)|\nabla u_2|^2$$

$$+ \frac{1}{2} \sum_{k,\ell=1}^{3} q_{k\ell}(x,x)[r^2 \frac{\partial u_1}{\partial x_k}\frac{\partial u_1}{\partial x_\ell} + f'(u_2)\frac{\partial u_2}{\partial x_k}\frac{\partial u_2}{\partial x_\ell}]\}dx \tag{21}$$

$$\leq -\int_D (\nu_0 - \frac{1}{2} q_0)\{r^2 |\nabla u_1|^2 + M|\nabla u_2|^2\}dx \ .$$

73

Thus we have an almost surely stable null solution if $\nu_o > \frac{1}{2} q_o$, by the Stability Theorem (i). For details, see [2].

REFERENCES

[1] P. L. Chow and E. Pardoux, "Stability of Nonlinear Stochastic Evolution Equations", submitted to J. Math. Analy. and Appl. for publication.

[2] P. L. Chow, "Stability of Reaction-Diffusion Equations under Random Perturbation", in preparation.

[3] S. A. Williams and P. L. Chow, "Nonlinear Reaction-Diffusion Models in Interacting Populations", J. Math. Analy. and Appl., 62 (1978), pp. 157-169.

[4] J. P. LaSalle and S. Lefschetz, Stability by Liapunov's Direct Method, Academic Press, New York, 1961.

DIFFERENTIATION OF TABULAR DATA

Česlovas Masaitis and George C. Francis
Ballistic Modeling Division
US Army Ballistic Research Laboratory/ARRADCOM
Aberdeen Proving Ground, MD 21005

ABSTRACT. A method based on an autoregressive model is derived for estimating the derivative of a function from its values at discrete points. The method is applied to synthetic data and compares favorably with moving polynomial arc, Butterworth filter, and B-spline smoothing.

I.   INTRODUCTION. Various methods for numerical differentiation have been considered. Some of these simply differentiate polynomial interpolation formulas[1], while others use least squares fits of the data by trigonometric[2] or algebraic[3] polynomials. Still others use polynomial splines.[4] This last method imposes certain smoothness conditions on the fitted functions. Similar conditions are introduced by applying Tikhonov's[5] regularization procedure.

Since a derivative, i.e., the limit of a ratio cannot be obtained from a finite set of tabular data, the data must be supplemented by suitable assumptions. The most common assumption is that tabular data are approximate values of a certain function which can be identified by the data and subsequently differentiated. For instance, the moving polynomial arc method assumes that data can be adequately represented by a polynomial, at least locally. Spline smoothing and Tikhonov's regularization procedure suppose that data are numerical values of an element in a Sobolev space. This paper assumes that tabular data are measured values of a function whose derivative can be represented by a linear combination of successive functional values, i.e., $y'(t)$ is an inner product $(a^{(p)}, Y)$ of a constant vector $a^{(p)}$ dependent on an integer p and the vector $Y = (y(t+p\Delta), y(t+(p+1)\Delta), \ldots, y(t+(p+k)\Delta))$, for several choices of the integer p and a positive integer k dependent on the case at hand. As is shown below this assumption implies that the underlying function is an element of the algebra A generated by algebraic, trigonometric, and exponential polynomials of a real variable. An appropriate element is selected by observing that $y \in A$ satisfies a certain family of linear difference equations with constant coefficients dependent on step size. Thus, an approximating function is obtained by constructing an appropriate difference equation, i.e., an autoregressive model. A certain cost functional of an approximating element is defined. Minimization of this functional yields the order of the autoregressive model, the optimal multiple of the data spacing as a step size, and an estimate of the variance of the measuring error. This estimate is obtained by assuming that measuring error is white noise with zero mean.

The coefficients of the autoregressive model determine the analytic structure of the approximating function, and this in turn determines weighting coefficients $a^{(p)}$ in the representation of the derivative:

$$(1.1) \qquad y'(t) = (a^{(p)}, Y).$$

II. FUNCTIONS WITH DERIVATIVES AS LINEAR COMBINATIONS OF FUNCTIONAL VALUES. The basic assumption of this paper is that the derivative of a tabular function is a linear combination of the functional values. A procedure for identifying such a function appropriate to the tabular data is obtained by observing that the function satisfies a certain difference equation. This follows from the following two propositions.

Proposition 1. Let $\tilde{A}$ be the set of functions differentiable on the inverval $I = [0,T]$ such that for every integer p satisfying $-t+k\Delta \leqslant p\Delta \leqslant T-t+\Delta$ there exists a constant vector $a^{(p)}$ of dimension k dependent on y (t) and the relation

$$(2.1) \qquad y'(t) = (a^{(p)}, Y)$$

is satisfied for every $t \in I$. Here $Y = (y(t+(p-1)\Delta), y(t+(p-2)\Delta), \ldots, y(t+(p-k)\Delta))$. Let A be the algebra generated by algebraic, trigonometric, and exponential polynomials on I. Then $A = \tilde{A}$.

Proof. First we show that $\tilde{A} \subset A$. Let the components of $a^{(p)}$ be $a_i(p)$, i = 1,2,...,k. Let p = k+s+1, j = k+s+1-i, and $a(j,s) = a_{-j+k+s+1}^{(k+s+1)}$. Then (2.1) can be written as follows:

$$(2.2) \qquad y'(t) = \sum_{j=s+1}^{k+s} a(j,s) \, y(t+j\Delta).$$

With s = -1 (2.2) becomes:

$$(2.3) \qquad y'(t) = \sum_{j=1}^{k} b_j^{(1)} \, y[t+(j-1)\Delta],$$

where $b_j^{(1)} = a(j-1, -1)$.

With the aid of (2.2) and (2.3) it can be shown by induction that

$$(2.4) \qquad y^{(q)}(t) = \sum_{j=1}^{k} b_j^{(q)} \, y[t+(j-1)\Delta]$$

for $q \leq k$ and some constants $b_j^{(q)}$. Let B be the $k \times k$ matrix with the q-th row $(b_1^{(q)}, b_2^{(q)}, \ldots, b_k^{(q)})$. If B is non-singular then (2.4) implies

$$(2.5) \qquad y(t) = \sum_{j=1}^{k} c_j y^{(j)}(t),$$

where the c_j's are elements of B^{-1}. If the rank of B is less than k, then there exist constants $d_1, d_2, \ldots, d_k$, not all zero, such that

$$\sum_{q=1}^{k} d_q b_j^{(q)} = 0, \quad j = 1, 2, \ldots, k.$$ Hence it follows from (2.4) that

$$(2.6) \qquad \sum_{q=1}^{k} d_q y^{(q)} = 0,$$

i.e., $y(t) \varepsilon \hat{A}$ satisfies either (2.5) or (2.6) and hence $\hat{A} \subset A$.

The inverse inclusion, $A \subset \hat{A}$ follows from a formal substitution of $y(t) \varepsilon A$, i.e., of

$$(2.7) \qquad y(t) = \sum_{j=1}^{m} \sum_{i=0}^{n_j} c_{ji} t^i \lambda_j^t$$

with

$$(2.8) \qquad \sum_{j=1}^{m} (n_j + 1) = k$$

into (2.2). By equating the coefficients of the similar terms on both sides of (2.2) after substitution of (2.7) the following system of equations is obtained:

$$(2.9) \qquad Lb = \Lambda,$$

where $\Lambda^T = (\log^{1-i} \lambda_j)$, $L = (\Lambda_{10}, \Lambda_{11}, \ldots, \Lambda_{1n_1}, \Lambda_{20}, \Lambda_{21}, \ldots, \Lambda_{2n_2}, \ldots, \Lambda_{mn_m})^T$,

$\Lambda_{ij}^T = ((s+1)^i \lambda_j^{s+1}, (s+2)^i \lambda_j^{s+2}, \ldots, (s+k)^i \lambda_j^{s+k})$, $i = 0, 1, 2, \ldots, n_j$,

$j = 1, 2, \ldots, m$, and $b^T = (a(s+1,s), a(s+2,s), \ldots, a(s+k,s))$. Thus, the inclusion $A \subset \hat{A}$ is established by showing that there exists a constant

vector b satisfying (2.9). This is implied by the fact that

$$(2.10) \qquad \det L \neq 0.$$

The relation (2.10) follows from

$$(2.11) \quad \det L = \prod_{j=1}^{n} \lambda_j^{(s+1)(n_j+1)} \prod_{i=1}^{m} n_i!! \prod_{j=1}^{m} \lambda_j^{n_j:}$$

$$\prod_{j=2}^{m} \prod_{i=1}^{j-1} (\lambda_j - \lambda_i)^{(n_i+1)(n_j+1)}.$$

Here $n!! = \prod_{j=1}^{n} j!$ and $n: = \sum_{i=1}^{n} i$. The identity (2.11) can be established

with the aid of the operator $T(m)$ defined as follows. Let

$$T_i = \mu_i \frac{\partial}{\partial \mu_i}, \quad T_i^{(0)}(\lambda_j) \, f(\mu_i) = f(\lambda_j), \quad T_i^{(q)}(\lambda_j) \, f(\mu_i) = T_i^q \, f(\mu_i) \big|_{\mu_i = \lambda_j},$$

and $m_j = \sum_{i=1}^{j} (n_i+1)$; then $T(m) = \prod_{j=1}^{m} \prod_{i=m_{j-1}-1}^{m_j} T_i^{(i-m_{j-1}-1)}(\lambda_j)$.

Let $L*$ be the $k \times k$ determinant with the i-th row $(\mu_i^{s+1}, \mu_i^{s+2}, \ldots, \mu_i^{s+k})$.

Then

$$(2.12) \qquad L* = \prod_{i=1}^{k} \mu_i^{s+1} \prod_{u=2}^{k} \prod_{v=1}^{u-1} (\mu_u - \mu_v).$$

It can be shown by induction on m that application of the operator $T(m)$
to both sides of (2.12) yields (2.11). This completes the description
of a method for proving the proposition.

An algorithm for determining the structure of an approximating
function can be derived from

Proposition 2. Let $y(t) \in A$, i.e., $y(t)$ is given by (2.7). Let

$\Delta > 0$, $P_\Delta (\lambda) = \prod_{j=1}^{m} (\lambda - \lambda_j^\Delta)^{n_j+1}$ and B_Δ be the operator defined by

$B_\Delta y(t) = y(t-\Delta)$. Then $y(t)$ satisfies the difference equation

(2.13) $P_k(B_\Delta)\, y(t) = 0.$

Proof. The relation (2.7) can be written in the form

$$(2.14) \qquad y(t) = \sum_{j=1}^{m} \sum_{i=0}^{n_j} c'_{ji}\left(\frac{t}{\Delta}\right)^i \left(\lambda_j^\Delta\right)^{\frac{t}{\Delta}},$$

where $c'_{ji} = c_{ji}\,\Delta^i$. Now (2.13) follows from the properties of linear difference equations with constant coefficients.

III. STRUCTURE OF AN APPROXIMATING FUNCTION.

The preceding results show that a function with derivative expressible as a linear combination of functional values satisfies the difference equation (2.13). Thus, it is representable by an autoregressive model. Parameters of such a model can be determined by a procedure developed for time series analysis[6], provided that the time series or its differences of order, say, d are stationary. This assumption need not hold for tabular data that must be differentiated. For instance, if the underlying function is exponential, the differences of any order are also exponential and hence non-stationary. An example of this type of data being differentiated[7] is pharmacokinetic data representing the concentration of an injected drug as a function of time after injection. In view of this, instead of attempting to determine the order of differences that may produce stationary series and at the same time considering possible periodicity ("seasonal" variation) a direct method for estimating the coefficients of the autoregressive model is adopted as described below.

Let $x(n)$, $n=1,2,\ldots,N$ be the tabular data and let $x(n,p,q) = x(p+qn)$ where q is a positive integer and $p=0,1,2,\ldots, q-1$, $n=1,2,\ldots,N_p$. Here $N_p = [(N-p)/q]$. Let $y(t)\in A$ be an approximating function of the data, i.e.,

$$(3.1) \qquad x(n,p,q) = y(r\Delta) + \varepsilon_r,$$

where $r = p+qn$ and ε_r is an observation error assumed to be weakly stationary white noise with zero mean and variance σ^2. The function $y(t)$ satisfies (2.1) for a suitable polynomial P, say, of degree k since $y(t)\in A$. We write this equation as follows

$$(3.2) \qquad y(r\Delta) = \sum_{j=1}^{k} a_j\, y[(r-jq)\Delta],$$

where the a_j's remain to be determined. By substituting (3.1) in (3.2)

we get:

$$(3.3) \qquad x(n,p,q) - \varepsilon_r = \sum_{j=1}^{k} a_j \, [x(n-j,p,q) - \varepsilon_{r-j}].$$

By transposing the terms in (3.3) and squaring both sides we get:

$$(3.4) \qquad \left[x(n,p,q) - \sum_{j=1}^{k} a_j \, x(n-j,p,q) \right]^2 =$$

$$\varepsilon_r^2 + \sum_{j=1}^{k} a_j^2 \, \varepsilon_{r-j}^2 + P_r \, ,$$

where P_r is a linear combination of products $\varepsilon_v \varepsilon_u$ with $u \neq v$. Since by assumption ε_r is white noise, we have $E(P_r) = 0$. Thus, by taking expected values of both sides of (3.4) we get:

$$(3.5) \qquad E\left\{ \left[x(n,p,q) - \sum_{j=1}^{k} a_j \, x(n-j,p,q) \right]^2 \right\} = \sigma^2 + \sigma^2 \sum_{j=1}^{k} a_j^2 .$$

We replace the expected value of the left hand side by its estimate (average) and get:

$$(3.6) \qquad \frac{1}{\tilde{N}} \sum_{p=0}^{q-1} \sum_{n=k+1}^{N_p} \left[x(n,p,q) - \sum_{j=1}^{k} a_j \, x(n-j,p,q) \right]^2 = \sigma^2 + \sigma^2 \sum_{j=1}^{k} a_j^2 \, ,$$

where $\quad \tilde{N} = \sum_{p=0}^{q-1} (N_p - k)$.

Thus, we get from (3.6):

$$(3.7) \qquad \sigma^2 = \frac{\dfrac{1}{\tilde{N}} \displaystyle\sum_{p=0}^{q-1} \sum_{n=k+1}^{N_p} \left[x(n,p,q) - \sum_{j=1}^{k} a_j \, x(n-j,p,q) \right]^2}{1 + \displaystyle\sum_{j=1}^{k} a_i^2} .$$

If M is the matrix of the normal equations of the overdetermined system

$$(3.8) \qquad \sum_{j=1}^{k} a_j \; x(n-j,p,q) = x(n,p,q),$$

$n=k+1,\ldots,N_p$, $p+0,1,\ldots,q-1$. and X is the right-hand side of these normal equations, then the coefficients a_j that minimize σ^2 in (3.6) satisfy the following:

$$(3.9) \qquad (M - \hat{N} \, \sigma^2 \, I) \, a = X,$$

where $a^T = (a_1, a_2, \ldots, a_k)$.

Thus, estimates of the a_j's are obtained by iterating (3.7) and (3.9) with initial value $\sigma^2 = 0$ in (3.9).

In order to compute a_j's by this procedure we have to choose k and q in (3.7) and (3.8). Obviously, a larger number of model parameters, i.e., larger k, yields a model better matching the data. A smaller value of q describes the data structure in a greater detail. However, increasing k as well as reducing q makes the system (3.9) ill-conditioned. Hence the value of k and the data spacing q⁴ must be chosen to minimize σ^2 in (3.7) and at the same time to prevent the matrix in (3.9) from becoming nearly singular. Thus, we have two conflicting criteria for selecting* the optimal pair (k,q). As usual a measure of optimality must be chosen heuristically. Our choice is an index

$$(3.10) \qquad J(k,q) = \sigma^2 (k,q)/[D(k,q)]^{1/k}$$

where $D(k,q)$ is the absolute value of the determinant of the last iteration of (3.9) corresponding to the choice of k and q. Thus, we compute the a_j's and $J(k,q)$ for $k=1,2,\ldots,k_o$ and $q=1,2,\ldots,q_o$ and select the pair (k,q) and the corresponding a_j's that minimize J(k,q).

We impose an additional constraint on (k,q) in order to prevent a choice of a model for which the data are inadequate, i.e., a model that contains terms of higher frequency than can be determined by the frequency of the data points. Thus, if ω_m is the maximum frequency of the selected model

---

*This is similar to the solution of the numerical differentiation problem by regularization where increasing the regularization parameter reduces ill-conditioning of the problem and decreasing the parameter yields a better fit of the data.

then we must at least have

$$(3.11) \qquad q \Delta < \frac{2\pi}{\omega_m} .$$

Suppose further that for some q the coefficients, a_j's, in (3.2) yield a real negative eigenvalue, say, $\lambda_j < 0$. Then the term $c_{jo} \lambda_j^n$ in (2.7) is equal to $c_{jo} |\lambda_j|^n \cos n\pi$ for every n. The frequency of this term is π radians per $q\Delta$ sec or $\pi/q\Delta$ radians per sec, i.e., we have $\omega_m \geqslant \pi/q\Delta$, contrary to requirement (3.11).

If for some k and q the equation (3.2) has an eigenvalue with a negative real part, say, $\lambda_j = -a + ib$ (a > 0), then the corresponding term $c_{jo}\lambda^n$ in (2.1) is expressible as $c_{jo} \exp(\alpha n + i\omega n)$ where $\cos \omega = -a/\sqrt{a^2 + b^2}$, i.e., $\omega > \frac{\pi}{2}$ if expressed in radians per unit time equal $q\Delta$ sec. Therefore this choice of q yields a spacing $q\Delta$ with less than four data points per period of the corresponding term in (2.7). Although theoretically two points per period may be adequate to determine the real and imaginary parts of the corresponding eigenvalue, even three points per period are inadequate when the data contain measuring errors. Furthermore, a negative real part only implies that the corresponding frequency is greater than $\pi/2$ per unit time. It may also be greater than π and less than $3\pi/2$, in which case the spacing $q\Delta$ provides less than two points per period. This is the reason why the pairs (k,q) leading to complex roots with negative real part are rejected.

In summary, the models (3.2) are determined for $q=1,2,\ldots,q_o$, $k=1,2,\ldots,k_o$ and among those with eigenvalues having non-negative real components that one which yields minimum J(k,q) in (3.10) is selected.

When the data is very noisy this selection of q may lead to a rather large step-size $q\Delta$ and, thus, may eliminate high frequency terms present in the data even if the original spacing Δ is adequate to represent this high frequency. This may happen when the amplitudes of high frequency terms are too small relative to the measuring error σ_r to be determined by the data taken at any spacing Δ. The procedure described above is intended to determine only the terms of (2.7) for which both the spacing and also the accuracy of the data are adequate, and the selected index J(k,q) is satisfactory in applications.

IV. WEIGHTING FACTORS FOR DERIVATIVES. With the autoregression coefficients determined as described above the weighting factors b_j in (2.3) for the first derivative can be obtained by solving the equation (2.9). Similar systems of equations define the weighting factors for higher order derivatives as in (2.4). The system (2.9) and its equivalent for higher derivatives are completely defined by the eigenvalues, λ_j's, of the autoregressive model.

It is very seldom that an autoregressive model obtained from experimental data yields multiple eigenvalues. Thus, the case of simple eigenvalues is of special interest. In this case the equations defining weighting factors can be solved in a closed form by applying Cramer's rule since the corresponding determinants can be expressed in rather simple form. Indeed, when the roots are simple the determinant of L is proportional to the Vandermondian of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. The numerator in an expression for, say, $b_j^{(r)}(s)$, the j-th weighting factor for the r-th derivative, is the $k \times k$ determinant B_j with the i-th row equal $(\lambda_i^{s+1}, \lambda_i^{s+2}, \ldots, \lambda_i^{s+j-1}, \log^r \lambda_i, \lambda_i^{s+j+1}, \ldots, \lambda_i^{s+k})$.

It can be shown that the minor of the $(n+1) \times (n+1)$ Vandermondian of $x_0, x_1, \ldots, x_n$ corresponding to the (j+1)-th element of the first row is equal to $S_{n-j} V$, where S_{n-j} is the symmetric function of $x_1, x_2, \ldots, x_n$ of order n-j and V is the Vandermondian of the same variables. Therefore expanding B_j with respect to its j-th column yields

$$(4.1) \qquad B_j = (-1)^{j-s} \sum_{p=1}^{k} (-1)^p \log^r \lambda_p S_k^{s+1} \lambda_p^{-s-1} S_{k-j+s}^{(p)} V_{k-1}^{(p)},$$

where $S_{k-j+s}^{(p)}$ is the symmetric function of order k-j+s of the variables $\lambda_1, \lambda_2, \ldots, \lambda_{p-1}, \lambda_{p+1}, \ldots, \lambda_k$ and $V_{k-1}^{(p)}$ is the Vandermondian of the same variables.

The symmetric functions in (4.1) can be expressed in terms of the coefficients of the characteristic equation of the autoregressive model. After this, dividing each term of (4.1) by det L and cancelling common factors (i.e., $V_{k-1}^{(p)}$ and others) the following expression for the weighting factor $b_j^{(r)}(s)$ is obtained:

$$b_j^{(r)}(s) = -\frac{1}{(\Delta)^r} \sum_{p=1}^{k} \frac{\lambda_p^{-s} \log^r \lambda_p \sum_{v=0}^{k-j} a_{k-j-v} \lambda_p^v}{\sum_{v=1}^{k} v \lambda_p^v a_{k-v}}.$$

This together with (2.4) yields the value of the r-th derivative y

V.   EXAMPLES OF SYNTHETIC DATA. Table 1 describes the synthet
data used below to illustrate the method of this paper and to compa
it with moving polynomial arc, Butterworth filter, and B-spline smo
procedures. Here column 1 numbers the cases from 1 to 6 for conven
of reference. The corresponding functions x(t) are specified in col
The last case here is the Bessel function of the first kind of order
The values of the functions were computed at points in the intervals
listed in column 4, and column 3 shows the step size for selecting
points in the respective intervals. Pseudorandom white gaussian no
with zero mean and the standard deviation σ given in column 5 was a
to each value, and then various methods for numerical differentiatic
were applied to the noisy data.

Table 1. Synthetic Data

| 1 | 2 | 3 | 4 | 5 |
Case	x(t)	Δ	I	σ
1	sin 2πt	.004	[0,1]	.02
2	sin 2πt	.004	[0,1]	.05
3	e^t	.01	[0,5]	.0.
4	$4^t(1+t+t^2-t^3)$	.004	[0,2]	.32
5	sin 2πt+.1sin 10πt	.004	[0,1]	.05
6	$J_0(t)$	.01	[1,6]	.01

The methods used are listed at the top of Table 2. Here moving polynomial arc corresponds to a cubic polynomial fitted to either 11 17 data points as indicated. The derivatives were evaluated at the midpoint of this span. Thus, derivatives at a few points at the beg ning and the end of the data sequence are not available.

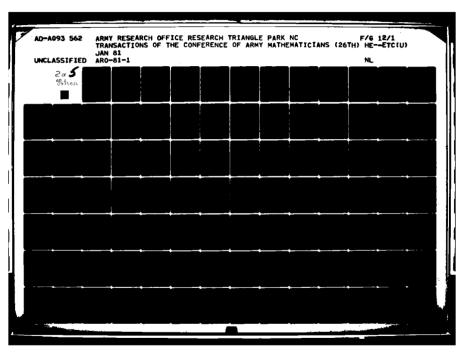The Butterworth filter applied here corresponds to the transfer function[8] $\pi/(s^3+2\pi s^2+2\pi^2 s+\pi^3)$. This method does not provide first derivatives at 65 data points at the end of the data sequence when $\Delta = .01$ and at 163 data points when $\Delta = .004$. Additional points are lost when higher derivatives are calculated.

The method of this report provides derivatives at every data po with appropriate values of s in (2.2) and in the corresponding expre sions for higher derivatives. The bulk of the derivatives are compu at the midpoint of the span of formula (2.2), i.e., for $s = -\left[\frac{k+1}{2}\right]$

Table 2 lists the RMSE of the first and second derivatives x' a x'' expressed in percentage of the RMS of the derivatives obtained by the analytic method. The errors correspond to the data described in Table 1 as indicated by the case numbers in column 1 of Table 2. A seen the current method is much more accurate in all these cases exc the Bessel function where Butterworth filter yields better results.

Table 2. Percentage Error

Case	Method	Moving Polyn. Arc, 11 pts.	Moving Polyn. Arc, 17 pts.	Butterworth Filter	Spline	Current Method
1	X'	25	14	88	6	1.3
1	X''	275	99	98	48	1.4
2	X'	62	33	88	15	3.4
2	X''	688	248	98	120	3.5
3	X'	.46	.25	3.4	.010	.004
3	X''	13	4.7	6.9	2.2	.004
4	X'	9	5	16	2.1	.23
4	X''	259	93	30	45	.80
5	X'	55	30	91	13	8
5	X''	89	255	99	46	18
6	X'	57	31	5.8	13	7.8
6	X''	2341	8384	8.4	403	33

REFERENCES

1. D.B. Hunter, "An Iterative Method of Numerical Differentiation,"
   Comp. J. 3, 270-271, 1960.

2. A. Talmi and G. Gilat, "Methods for Smooth Approximation of Data,"
   Journal of Comp. Physics 23, 93-123, 1977.

3. H.C. Hershey, J. L. Zakin, and R. Simha, "Numerical Differentiation
   of Equally Spaced and Not Equally Spaced Experimental Data," Ind.
   Eng. Chem. Fund. 6, 413-421, 1967.

4. P. Craven and G. Wahba, "Smoothing Noisy Data with Spline
   Functions," Num. Math. 31, 377-403, 1979.

5. A. N. Tikhonov, "Solution of Incorrectly Formulated Problems and
   the Regularization Method," Soviet Mach. Dokl. 4, 1035-1038, 1963.

6. G.E. Box and G.M. Jenkins, "Time Series Analysis: Forecasting and
   Control." Revised Edition. Holden-Day, 1976.

7. S. Wold, "Spline Functions in Data Analysis," Technometrics, Vol.
   16, No. 1, pp 1-11, 1974.

8. J.N. Groff and A.N. Gordon, "The Methodology and Preparatory
   Analysis of Tracking Data for the Antitank Missile Test (ATMT)
   Program. Part I: General Methodology and Shillelagh Analysis."
   Tech Report No. 151, AMSAA, 1976.

COMPUTER AIDED ANALYSIS OF MECHANICAL SYSTEMS WITH INTERMITTENT MOTION*

Edward J. Haug and Roger A. Wehage
Materials Division
College of Engineering
The University of Iowa
Iowa City, Iowa 52242

ABSTRACT. A general method is presented for dynamic analysis of systems with impulusive forces, impact, discontinuous constraints, and discontinuous velocities. This class of systems includes discontinuous kinematic and geometric constraints that characterize backlash and impact within systems. A method of computer generation of the equations of motion and impulse-momentum relations that define jump discontinuities in system velocity for large scale systems is presented. An event predictor working in conjunction with a new numerical integration algorithm efficiently controls its progress and allows for automatic equation reformulation. A weapon mechanism and a trip plow are simulated using the method to illustrate its capabilities.

I.   INTRODUCTION. Dynamic analysis of mechanisms and machines that undergo impulsive loading, impact, and cammed locking of parts is a field that has seen only moderate development. In spite of the technical difficulty of analysis of such systems and their inherently discontinuous behavior, such systems are commonly encountered in manufacturing equipment and in weapon mechanisms and must be carefully analyzed during the design process. The purpose of this paper is to present a computer aided analysis method that is capable of analyzing complex mechanisms and machines that undergo intermittent motion.

Previously used methods of intermittent motion analysis have generally used pieced interval analysis, in which the analysist writes the equations of motion between times at which discontinuous events occur [1]. Momemtum balance equations must be written to account for velocity discontinuities that may occur in a specific system configuration. Numerical integration is halted at the point of discontinuity, new initial conditions on velocity are formulated and integration is continued. A basic limitation of this method of analysis is the effort required to write system equations that are valid in intervals between events whose ordering is not generally known before the analysis is begun. Thus, the analyst is required to write equations and computer code for all ordering of logical events that may conceivably occur.

One method that has been used to alleviate the foregoing difficulty is to use Heaviside step functions that define logic associated

---

with the events occurring during intermittent motion. These disconti-
nuous functions may then be smoothed to provide a set of governing
differential equations of motion [2]. This procedure can be justified
on the basis of distribution theory [3,4] and has been successfully
employed in weapon mechanism dynamics [5]. The distribution theoretic
method has been used in conjunction with a computer code that auto-
matically generates the system equations of motion [6] by defining
"logical spring-dampers" that account for certain aspects of inter-
mittent motion [7]. In this paper, the method of computer generated
equations of motion is employed with the pieced interval analysis method
to treat dynamics of quite general planar systems that undergo inter-
mittent motion.

II. EQUATIONS OF CONSTRAINED PLANAR MOTION. For planar mechanical
systems treated here, constraints between elements are taken as friction
free (workless) translational and rotational joints. Extensions to
include constraints such as cams and prescribed functional relations
are possible by incorporating provisions for nonstandard elements that
are supplied by the user. In addition to standard constraints, standard
spring-damper-actuator combinations connecting any pair of points on
different bodies of the system are included in the model. Allowance is
also made for arbitrary user supplied external forcing functions.

In order to determine the configuration or state of a planar
mechanical system, it is first necessary to define generalized coor-
dinates that specify the location of each body. As shown in Fig. 1,
let the x-y coordinate system be a fixed inertial reference frame.
Define a centroidal body-fixed coordinate system ξ_i - n_i embedded in a
typical body i. The location of body i is specified by the global coor-
dinates (x_i, y_i) of the origin or vector $\vec{R}_i$ and the angle ϕ_i of rotation
of the body fixed coordinate system relative to the global coordinates.
In terms of these generalized coordinates, the kinetic energy of the
i^{th} body is

$$T^i = \frac{1}{2} m_i (\dot{x}_i^2 + \dot{y}_i^2) + \frac{1}{2} J_i \dot{\phi}_i^2 \tag{1}$$

where m_i is the mass of the ith body and J_i is its centroidal polar
moment of inertia.

Figure 1 further depicts an adjacent body j, with body-fixed
coordinate system located by the vector $\vec{R}_j$. Let arbitrary points p_{ij}
on body i and p_{ji} on body j be located by vectors $\vec{r}_{ij}$ and $\vec{r}_{ji}$, respec-
tively. These points are in turn connected by a vector $\vec{r}_p$,

$$\vec{r}_p = \vec{R}_i + \vec{r}_{ij} - \vec{R}_j - \vec{r}_{ji} \tag{2}$$

The vector condition for a rotational joint at points p_{ij} and p_{ji} is
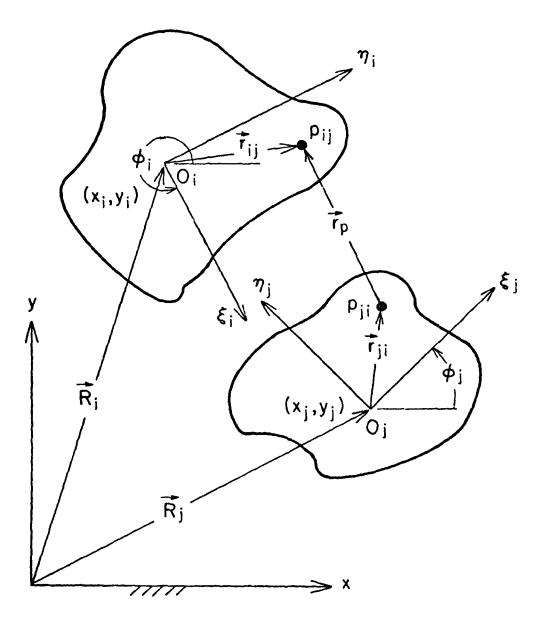$\vec{r}_p = 0$, yielding the following pair of scalar constraint equations:

Figure 1. Body Coordinate and Revolute Joint Definition.

91

$$x_i + \xi_{ij} \cos \phi_i - n_{ij} \sin \phi_i$$

$$- x_j - \xi_{ji} \cos \phi_j + n_{ji} \sin \phi_j = 0 \tag{3}$$

$$y_i + \xi_{ij} \sin \phi_i + n_{ij} \cos \phi_i$$

$$- y_j - \xi_{ji} \sin \phi_j - n_{ji} \cos \phi_j = 0$$

For a translational joint shown in Fig. 2, let points p_{i1} and p_{i2} on body i, and points p_{j1} and p_{j2} on body j lie on some line parallel to the path of relative motion between the two bodies, such that $\vec{S}_i$ and $\vec{S}_j$ are of nonzero magnitude. Since $\vec{S}_i$ and $\vec{S}_j$ are parallel, $\vec{S}_i \times \vec{S}_j = \vec{0}$, with zero z component yielding the scalar equation

$$[(\xi_{i2} - \xi_{i1}) \cos \phi_i - (n_{i2} - n_{i1}) \sin \phi_i]$$

$$[(\xi_{j2} - \xi_{j1}) \sin \phi_j + (n_{j2} - n_{j1}) \cos \phi_j]$$

$$- [(\xi_{j2} - \xi_{j1}) \cos \phi_j - (n_{j2} - n_{j1}) \sin \phi_j]$$

$$[(\xi_{i2} - \xi_{i1}) \sin \phi_i + (n_{i2} - n_{i1}) \cos \phi_i] = 0 \tag{4}$$

Likewise, $\vec{r}_{ji}$ and $\vec{S}_j$ are parallel so $\vec{r}_{ji} \times \vec{S}_j = \vec{0}$ yields a zero z component and the second scalar equation

$$[x_i + \xi_{i1} \cos \phi_i - n_{i1} \sin \phi_i - x_j - \xi_{j1} \cos \phi_j + n_{j1} \sin \phi_j]$$

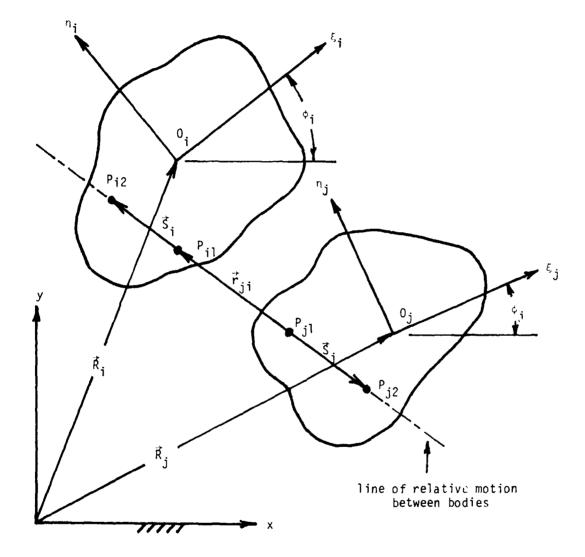$$[(\xi_{j2} - \xi_{j1}) \sin \phi_j + (n_{j2} - n_{j1}) \cos \phi_j]$$

$$- [y_i + \xi_{i1} \sin \phi_i + n_{i1} \cos \phi_i - y_j - \xi_{j1} \sin \phi_j - n_{j1} \cos \phi_j]$$

$$[(\xi_{j2} - \xi_{j1}) \cos \phi_j - (n_{j2} - n_{j1}) \sin \phi_j] = 0 \tag{5}$$

The parameters (ξ_{i1}, n_{i1}) and (ξ_{i2}, n_{i2}) locate points p_{i1} and p_{i2} in body i coordinate system, and (ξ_{j1}, n_{j1}) and (ξ_{j2}, n_{j2}) locate points p_{j1} and p_{j2} in body j coordinate system.

Since springs, dampers, and actuators generally appear together, as shown in Fig. 3, they are incorporated into a single set of force equations. If any are absent, their effect is eliminated by setting the corresponding terms to zero. The equation for spring-damper force is
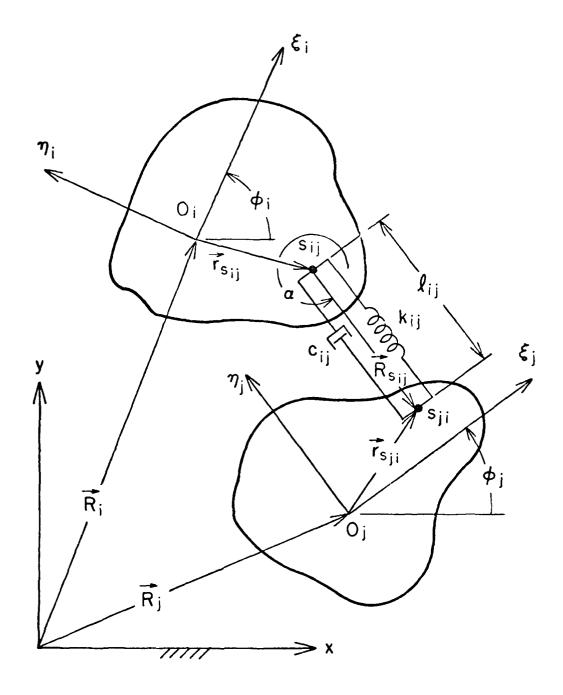
Figure 2. Translational Joint Definition.

Figure 3.   Spring-Damper-Actuator Definition.

$$\vec{F}_{ij} = \left[k_{ij}(\ell_{ij} - \ell_{0_{ij}}) + c_{ij} \dot{\ell}_{ij} + F_{0_{ij}} \right] \frac{1}{\ell_{ij}} \vec{R}_{s_{ij}} \qquad (6)$$

where $\vec{F}_{ij}$ is the resultant force vector $[F_{x_{ij}}, F_{y_{ij}}]^T$ in the spring-damper, $\vec{R}_{s_{ij}}$ is the vector $[\ell_{ij} \cos \alpha, \ell_{ij} \sin \alpha]^T$ between points S_{ij} and S_{ji} of a spring-damper connection, k_{ij} is an elastic spring coefficient that may depend on generalized coordinates and time, c_{ij} is a damping coefficient that may depend on generalized coordinates and time, $\ell_{0_{ij}}$ is the undeformed spring length, ℓ_{ij} is the deformed spring length, $\dot{\ell}_{ij}$ is the time derivative of ℓ_{ij}, and $F_{0_{ij}}$ is an actuator force applied along the spring and may depend upon generalized coordinates and time.

The virtual work of externally applied forces and spring-damper forces acting on body i can now be written as

$$\delta W_i = Q_{x_i}(q,\dot{q},t)\delta x_i + Q_{y_i}(q,\dot{q},t)\delta y_i + Q_{\phi_i}(q,\dot{q},t)\delta\phi_i \qquad (7)$$

where $q = [q^{1^T}, q^{2^T}, \ldots, q^{n^T}]^T$ and $q^i = [x_i, y_i, \phi_i]^T$ are the total system and body i generalized coordinate vectors, respectively, and $\dot{q} = \frac{dq}{dt}$. The vector $Q_i = [Q_{x_i}, Q_{y_i}, Q_{\phi_i}]^T$ of generalized forces on body i is thus defined and $Q = [Q^{1^T}, Q^{2^T}, \ldots, Q^{n^T}]^T$ is the vector of system generalized forces, depending on q and $\dot{q}$.

Let the m-vector equation of all kinematic constraints be denoted simply as

$$\phi(q,t) = 0 \qquad (8)$$

A virtual displacement δq of the system is then consistent with constraints if

$$\phi_q \, \delta q = 0 \qquad (9)$$

where $\phi_q \equiv \left[\frac{\partial \phi_i}{\partial q_j} \right]$ is the Jacobian matrix of the vector constraint function.

95

If the constraints are workless, the variational form of Lagrange's equations of motion is [8]

$$\left[\frac{d}{dt}\,(T_{\dot{q}}) - T_q - Q^T\right]\delta q = 0 \qquad (10)$$

for all δq satisfying Eq. 9. By Farkas Lemma [9], there exists a vector $\lambda \in R^m$ of multipliers such that

$$\frac{d}{dt}\,(T_{\dot{q}})^T - T_q{}^T - Q + \Phi_q{}^T\lambda = 0 \qquad (11)$$

which with the constraint equations of Eq. 8 form the constrained equations of motion of the system. For planar systems treated in this paper, the kinetic energy is $T = \frac{1}{2}\,\dot{q}^T M\dot{q}$, where M is a constant diagonal matrix. Thus Eq. 11 becomes simply

$$M\ddot{q} - Q + \Phi_q{}^T\lambda = 0 \qquad (12)$$

Initial conditions for system motion are given as

$$\begin{aligned} q(0) &= q^0 \\ \dot{q}(0) &= \dot{q}^0 \end{aligned} \qquad (13)$$

where q^0 and $\dot{q}^0$ are consistent with constraints. That is, q^0 satisfies Eq. 8 and $\dot{q}^0$ satisfies the equation obtained by taking the time derivative of Eq. 8,

$$\Phi_q\dot{q}^0 + \Phi_t = 0 \qquad (14)$$

To define the intermittent nature of the motion of a mechanical system, a set of event times t_i, $i = 1, \ldots, k$ at which some discontinuity in system behavior occurs is defined by equations

$$\Omega^i(t,q(t)) = 0 \qquad (15)$$

The ordering of event times is defined by the dynamical system and forcing functions. The equations of motion are integrated forward in time and the values of Ω^i are monitored until one of them becomes zero, defining t_1. The process is continued until a second function Ω^i becomes zero, defining t_2. The process continues until the simulation is completed.

The constraints may be modified at the event times, so the vector $\Phi(q,t)$ of Eq. 8 may be modified as the motion of the system progresses. For example, one of the events may be an impact and subsequent locking

together of two of the bodies in the system. Thus, after the event
occurs additional constraints are added and the equations of motion
and constraint are modified, with additional components of the multiplier
λ introduced.

The foregoing equations of motion and constraints are computer
generated using a computer code, Dynamic Analysis and Design System, DADS
[6] that constructs all of the matrices needed in the simulation. This
computer generation of equations is crucial, since the form of equations
between event times is variable and depends on ordering of the event
times. The remaining task in generation of the complete system of
equations is formulation of velocity jump conditions that must hold at
event times involving impulsive loading and impact between bodies in
the system.

3. Reduced Equations of Motion and Momentum-Impulse Relations.
In order to obtain momentum-impulse relations needed for modelling
intermittent motion, it is helpful to eliminate the multiplier λ from
the equations of motion of Eq. 12. To do this, a partitioning of the
generalized coordinates is introduced that defines dependent generalized
coordinates in terms of independent coordinates, through the constraint
equations.

Beginning with the initial value q^0 of Eq. 13, a Gauss-Jordon
reduction of the Jacobian matrix $\phi_q \equiv \frac{\partial \phi}{\partial q} = \left[\frac{\partial \phi_i}{\partial q_j} \right]$, with double pivoting,
defines a partitioning of $q = [u^T, v^T]^T$ such that ϕ_u is nonsingular. By
the implicit function theorem [10], the constraint equations of Eq. 8
guarantee existence of a twice continuously differentiable function
$f(v,t)$ such that

$$u = f(v,t) \tag{16}$$

satisfies

$$\phi(f(v,t),v,t) = 0 \tag{17}$$

and

$$\frac{\partial u}{\partial v} = - \phi_u^{-1} \phi_v = D(v,t) \tag{18}$$

where the matrix $D(v,t)$ is continuously differentiable.

The matrix ϕ_u is numerically decomposed into lower and upper tri-
angular matrices L and U such that

$$\phi_u = L \cdot U.$$

Forward elimination and back substitution steps replace the less efficient matrix inversion process. For example Eq. 18 is written as

$$L \cdot U \cdot D = - \phi_v$$

which is solved in two steps

$$L \cdot A = - \phi_v$$

and

$$U \cdot D = A$$

for the matrix D.

Given a numerical value of v and a time t, u can be found by Newton iteration, with the increment in u defined by

$$\Delta u = -\phi_u^{-1} \phi \tag{19}$$

Differentiating Eq. 8 with respect to time and partitioning gives

$$\phi_u \dot{u} + \phi_v \dot{v} + \phi_t = 0 \tag{20}$$

Thus, by Eq. 18,

$$\dot{u} = D\dot{v} - \phi_u^{-1} \phi_t \tag{21}$$

Similarly, taking the time derivative of Eq. 20 yields

$$\phi_u \ddot{u} + \phi_v \ddot{v} + V(v, \dot{v}, t) = 0 \tag{22}$$

where

$$V(v, \dot{v}, t) = (\phi_u \dot{u})_u \dot{u} + (\phi_v \dot{v})_u \dot{u} + (\phi_u \dot{u})_v \dot{v} + (\phi_v \dot{v})_v \dot{v}$$

$$+ 2\phi_{tu} \dot{u} + 2\phi_{tv} \dot{v} + \phi_{tt} \tag{23}$$

which can be evaluated explicitly in terms of $\dot{v}$, using Eq. 21.

The equations of motion of Eq. 12 can now be partitioned in the form

$$M^u \ddot{u} - Q^u + \phi_u^T \lambda = 0 \tag{24}$$

$$M^v \ddot{v} - Q^v + \phi_v^T \lambda = 0 \tag{25}$$

where M^u, M^v, Q^u, and Q^v are partitions of M and Q consistent with the partitioning of q. Solving for λ from Eq. 24, substituting into Eq. 25, and noting that

$$\Phi_v^T \Phi_u^{-T} = -D^T$$

yields

$$M^v \ddot{v} + D^T M^u \ddot{u} = Q^v + D^T Q^u \qquad (26)$$

Substituting from Eq. 22 yields an explicit differential equation in the independent variables

$$[M^v + D^T M^u D]\ddot{v} - D^T M^u \Phi_u^{-1} V(v,\dot{v},t) = Q^v + D^T Q^u \qquad (27)$$

Let t_i be a point in time at which a "violent event" occurs, which is to be approximated by a discontinuity. In reality, the event occurs over a time interval $\tau_1 < t_i < \tau_2$, as shown in Fig. 4, and behavior is smooth except possibly at t_i. Integrate Eq. 27 to obtain

$$\int_{\tau_1}^{\tau_2} [M^v + D^T M^u D]\ddot{v}\,dt - \int_{\tau_1}^{\tau_2} D^T M^u \Phi_u^{-1} V\,dt = \int_{\tau_1}^{\tau_2} [Q^v + D^T Q^u]\,dt \qquad (28)$$

Since D^T is differentiable, integration by parts and using the mean value theorem gives

$$[M^v + D^T M^u D]\dot{v} \Big|_{\tau_1}^{\tau_2} - \int_{\tau_1}^{\tau_2} \left\{ -\left[\frac{d}{dt}(D^T M^u D)\right]\dot{v} + D^T M^u \Phi_u^{-1} V(v,\dot{v},t)\right\} dt$$

$$= \int_{\tau_1}^{\tau_2} Q^v\,dt + \hat{D}^T \int_{\tau_1}^{\tau_2} Q^u\,dt \qquad (29)$$

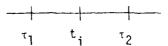where $\hat{D}$ is a matrix whose elements are those of D evaluated in (τ_1, τ_2).

Figure 4, Event Interval

Treating Q as impulsive at t_i, the integrals of generalized force are

"generalized impulse", $P^v = \int_{\tau_1}^{\tau_2} Q^v dt$ and $P^u = \int_{\tau_1}^{\tau_2} Q^u dt$. Taking the

limit in Eq. 29 as $\tau_1 \rightarrow t_i$ and $\tau_2 \rightarrow t_i$, noting that $D^T M^u \phi_u^{-1} V(v,\dot{v},t)$
is bounded, yields the "impulse-momentum" equation at t_i as

$$[M^v + D^T M^u D][\dot{v}(t_i+) - \dot{v}(t_i-)] = P^v + D^T P^u \tag{30}$$

This prescribes the velocity jump in v due to impulsively applied loads.

It is important to note that Eq. 30 involves impulse and momentum of all elements of the mechanical system. This is crucial, since the bodies making up the system interact through constraints, so an impulse-momentum balance relation involving only the bodies on which the impulsive force acts is impossible. Deriving the relation of Eq. 30 by manual calculation would be extremely difficult and time consuming. One of the strongest points of the method presented here is the automatic assembly of the coefficient matrices of Eq. 30.
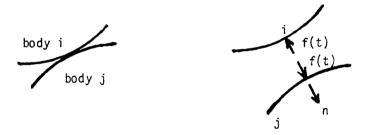
Figure 5 Impacting Bodies

For impact of bodies i and j, as shown in Figure 5, a coefficient of restitution e provides the relative velocity relation in direction n as

$$n^T\left(\dot{q}^i(t_i+) - \dot{q}^j(t_i+)\right) = -en^T\left(\dot{q}^i(t_i-) - \dot{q}^j(t_i-)\right) \tag{31}$$

or with $N \in R^{3n}$, $N = \left[N^{u^T}, N^{v^T}\right]^T$, Eq. 31 may be written

$$N^{v^T}\dot{v}(t_i+) + N^{u^T}\dot{u}(t_i+) = -e\left[N^{v^T}\dot{v}(t_i-) + N^{u^T}\dot{u}(t_i-)\right]$$

or using Eq. 21, this is

$$\left[N^{V^T} + N^{u^T} D \right] \dot{v}(t_i+) = -e \left[N^{V^T} + N^{u^T} D \right] \dot{v}(t_i-) \qquad (32)$$

The generalized impulse of the force $f(t)$ in Fig. 5 is

$$P = \int_{t_i-\epsilon}^{t_i+\epsilon} f(t)N dt = pN \qquad (33)$$

where

$$p = \int_{t_i-\epsilon}^{t_i+\epsilon} f(t) dt \qquad (34)$$

Defining the partitioning $P^V = pN^V$ and $P^u = pN^u$, Eq. 30 gives

$$\dot{v}(t_i+) - \dot{v}(t_i-) = p \left[M^V + D^T M^u D \right]^{-1} \left[N^V + D^T N^u \right] \qquad (35)$$

Premultiplying by $N^{V^T} + N^{u^T} D$ and using Eq. 32, p is determined from the scalar equation

$$-(1 + e)\left[N^{V^T} + N^{u^T} D \right] \dot{v}(t_i-) = p \left[N^{V^T} + N^{u^T} D \right]$$

$$\cdot \left[M^V + D^T M^u D \right]^{-1} \left[N^V + D^T N^u \right] \qquad (36)$$

With p known, $v(t_i+)$ is given by Eq. 35. Equations 35 and 36 thus define jump discontinuities in velocity due to given impulsive loading and impact between bodies.

The above equations are put into matrix form for automatic solution by the DADS computer program. Subtract Eq. 32 from the identity

$$\left[N^{V^T} + N^{u^T} D \right] \dot{v}(t_i-) = \left[N^{V^T} + N^{u^T} D \right] \dot{v}(t_i-)$$

and define

$$\Delta\dot{v}_i = \dot{v}(t_i+) - \dot{v}(t_i-) \qquad (37)$$

The matrix equation thus becomes

$$
\begin{bmatrix}
(M^V + D^T M^u D) & -(N^V + D^T N^u) \\
-(N^{V^T} + N^{u^T} D) & 0
\end{bmatrix}
\begin{bmatrix}
\Delta \dot{v}_i \\
p
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
0 \\
(e+1)\ (N^{V^T} + N^{u^T} D)\ \dot{v}(t_i -)
\end{bmatrix}
\tag{38}
$$

whose solution yields the desired change in velocity and impulse magnitude p.

A numerical integration algorithm for automatic formulation and efficient solution of the reduced system of differential equations of motion is presented in Ref. 6. The algorithm is breifly reviewed here and extensions to include pieced interval analysis and momentum balance are discussed in the following steps.

Step 1. An approximate (from initial estimate or prediction) or exact (from static equilibrium analysis) system configuration is known. Evaluate Φ_q and perform L-U factorization to determine Φ_u, Φ_v, D, and the partitioning of q into u and v. If any constraint equations, Eq. 8, are not satisfied iterate to determine u using Eq. 19. Independent variables v remain fixed at the values provided by the integration algorithm.

Step 2. Evaluate Φ_q and factor as in Step 1. Evaluate $\dot{u}$ by Eq. 21, where independent velocities $\dot{v}$ remain fixed at the values provided by the integration algorithm.

Step 3. Compute independent accelerations $\ddot{v}$ from Eq. 27 and evaluate dependent accelerations $\ddot{u}$ from Eq. 26 if desired.

Step 4. Before advancing the solution ahead in time check Eq. 15 for any $\Omega^i = 0$ in the time interval. This is done by introducing variables $\ell^i = \Omega^i$ and formulating $\dot{\ell}^i = \frac{\partial \Omega^i}{\partial q} \dot{q} + \frac{\partial \Omega^i}{\partial t}$. These differential equations are integrated along with the system equations of motion. The variables ℓ^i are first predicted to the next point in time and if one or more change sign the corresponding Ω^i are zero in the time interval. A new time step is then calculated corresponding to the point in time where the first ℓ^i becomes zero and control then passes to Step 5. If no ℓ^i changes sign, control passes directly to Step 5. (The algorithm also checks for the possibility that a given ℓ^i passes through zero twice in a given step in which case the first root is selected.)

Step 5. Using the explicit Adams PECE algorithm, advance the solution to the next time step. The algorithm requires execution of Steps 1 to 3 each time evaluation of the reduced system of differential equations is called for. Go to Step 6.
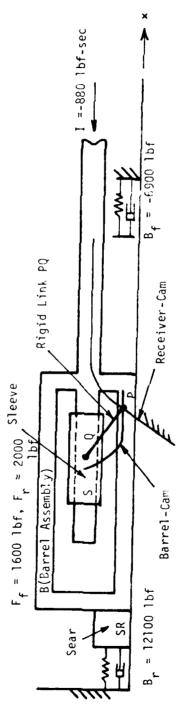
Step 6. If no event is detected return to Step 4. Otherwise determine the appropriate action to be taken such as momentum balance using Eq. 38. Then return to Step 4 to continue the simulation. A more detailed description of the procedures involved in Steps 4 and 6 is presented in the numerical examples of Section 4. The procedure for solving Eq. 38 for $\Delta \dot{v}_i$ and p is basically the same as for solving Eq. 27 for $\dot{v}$.

4. **NUMERICAL EXAMPLES.** Two numerical examples are presented here to illustrate the analysis method. The first example is a 75 mm automatic cannon with three moving masses. Although the system is composed mainly of translating bodies, it does have a number of significant logical events that include discontinuous velocities due to impulsive loading, body impact, mass capture and release, and discontinuous accelerations due to contact with and separation from buffers. The second example is a more complicated trip-plow mechanism. It consists of seven rigid bodies, five of which undergo large angular displacements. These bodies experience multiple impacts as the mechanism progresses through a reset cycle.

4.1. **The 75 mm Cannon System.** The 75 mm automatic weapon mechanism shown in Fig. 6 consists of three main masses: the barrel assembly B, the sleeve S, and the sear SR. A camming action is used to move the sleeve over a telescoped cartridge, so that the cartridge can be safely fired during each cycle of system operation. The B-cam path is fixed in the barrel assembly B, while the R-cam path is fixed in the receiver R, which is rigidly attached to ground. The sleeve S is connected by a rigid bar PQ to a pin at point P that slides without friction along the R and B cam paths.

Two forces, F_f and F_b, drive the barrel during its forward (counter recoil) and rearward (recoil) motion, respectively. A front buffer B_f and a rear buffer B_r slow the barrel assembly during extreme displacement. Both front and rear buffers are designed to produce constant retarding forces.

Logical times t_i at which impact or other irregularities of intermittent motion occur are introduced as an integral element of the dynamic model. Between these times, the motion and acceleration of the system is continuous. At these times, discontinuities in velocities and acceleration, changes in system constraints, and mass capture or release can occur. These logical times are functions of the system state and are determined as the simulation progresses. Logical times will now be defined for the firing from run-out mode of weapon operation:

103

Figure 6. 75mm Cannon System.

(1) $t_0 = 0$: The barrel assembly B, in aut..matic fire, passes
     the sear position with velocity $\dot{x}_2 = 40$ in/sec. (Initial
     starting point, not considered as a logical event). A
     forward driving force $F_f = 1600$ lbf acts on the barrel.

(2) t_1: The barrel B contacts the front buffer, $B_f = -6900$ lbf
     becomes active (restart integration because of discontinuous
     acceleration).

(3) t_2: The charge is ignited. An impulse of -880 lbf-sec acts
     on the barrel B, (perform momentum balance to obtain new
     velocities. F_f is deactivated and drive force $F_b = 2000$ lbf
     is activated. Restart integration.)

(4) t_3: The barrel B breaks contact with the front buffer,
     $B_f = 0$ lbf (restart integration).

(5) t_4: The barrel B impacts and captures the sear SR which was
     locked to the receiver. The rear buffer $B_r = 12100$ lbf acts
     against the sear (release constraint between sear and receiver,
     perform momentum balance with coefficient of restitution
     $e = 0$, activate constraint between barrel B and sear SR, and
     restart integration with new velocities.)

(6) t_5: The barrel B and sear SR come to rest. The barrel drive
     force F_b is deactivated, the drive force F_f is activated, and
     the rear buffer force B_r is deactivated (restart integration).

(7) t_6: If automatic fire is to terminate, the barrel B and sear
     SR return to the initial sear position. The sear impacts
     the receiver, and the sear and barrel are captured by the
     receiver. (Perform momentum balance with coefficient of
     restitution $e = 0$, activate constraint between sear and
     receiver.) The cycle is complete with sear and barrel locked
     to receiver.

(7') t_6': If automatic fire is to continue, the barrel B and sear
     SR return to the initial sear position. The sear impacts the
     receiver and is captured by the receiver, while the barrel
     is released from the sear, (release the constraint between
     sear and barrel, perform momentum balance with coefficient
     of restitution $e = 0$, activate constraint between sear SR
     and receiver, and restart integration with new velocities).
     The cycle is complete and barrel is in the runout configura-
     tion for another round.

     Logical times t_1 to t_6 depend upon the state of the system; the
relative horizontal displacements and relative velocities between bodies
of the system. Since the horizontal position, velocity, and acceleration

of body centers-of-mass are state variables, logical times are expressed as functions of these variables.

The logical events are defined as follows:

(1) t_1: $x_2 - 34.26 = \ell^1 = 0$

(2) t_2: $x_2 - 36.75 = \ell^2 = 0$

(3) t_3: $x_2 - 34.26 = \ell^1 = 0$

(4) t_4: $x_2 - x_3 - 16 = \ell^3 = 0$

(5) t_5: $\dot{x}_2 = \ell^4 = 0$

(6) t_6: $x_2 - x_3 - 16 = \ell^3 = 0$

The six events t_1 to t_6 are thus defined by the four logical variables ℓ^1 to ℓ^4. In order to incorporate these event predictors into the numerical integration algorithm, the derivatives of the above equations, with appropriate initial conditions, are formulated and integrated along with the system equations of motion. Thus

$$\dot{\ell}^1 = \dot{x}_2 , \qquad \ell^1(0) = -18.26$$

$$\dot{\ell}^2 = \dot{x}_2 , \qquad \ell^2(0) = -20.75$$

$$\dot{\ell}^3 = \dot{x}_2 - \dot{x}_3, \qquad \ell^3(0) = C$$

$$\dot{\ell}^4 = \ddot{x}_2 , \qquad \ell^4(0) = 40$$

The procedure for determining the complete system state precisely at logical times t_1 to t_6, identified by logical variables ℓ^1 to ℓ^4, is as follows. An appropriate time step is determined by the numerical integration algorithm based on the previous system state, polynomial predictor order, and error tolerance. Each logical variable in succession is predicted ahead in time, using this time step. If no logical variable is found to have passed through zero, the program advances the solution by the desired time step and the process is repeated. If one or more logical variables have passed through zero, the precise times at which the corresponding logical variables are zero are calculated by interpolation, using the polynomial predictor. A solution is then forced at the earliest logical time, indicating occurance of the first event. Control is then returned to user supplied subroutines so that actions can be taken according to the intent of the active logical variable.
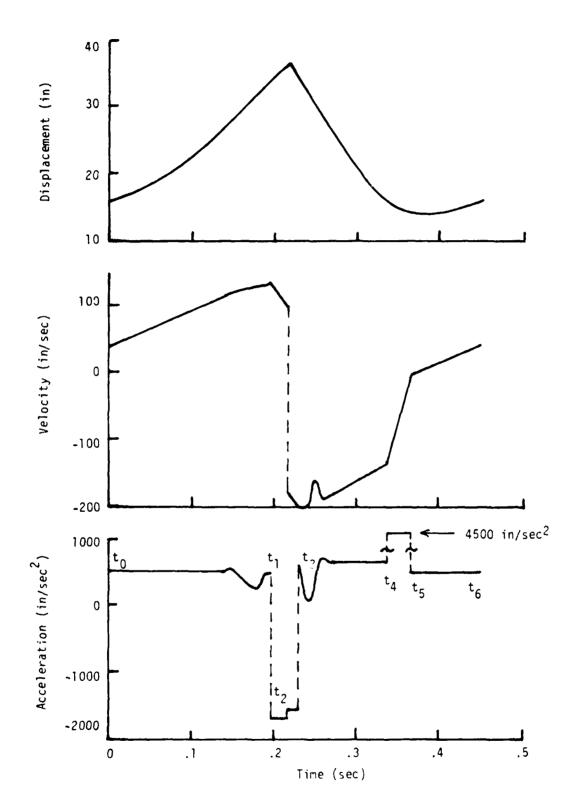
Fig. 7. Dynamics of the 75 mm Gun Tube.

107

Most discontinuous events can be categorized according to the order of increasing difficulty as follows:

(1) Those requiring restart of the integration process only, such as when discontinuous forces act on or within the system. These forces are not considered to be impulsive in nature, thus only discontinuous accelerations result.

(2) Those requiring momentum balance due to impulsive external loads (impact between bodies and mass capture or release is excluded) with no supplemental restitution equations or constraint equation modification.

(3) Those requiring momentum balance due to impact between bodies and mass capture or release. Supplemental coefficient of restitution equations are appended to the momentum balance equations to achieve the desired velocity changes. Constraints are added or deleted, as needed to facilitate mass capture or release.

The six events t_1 to t_6 fall into the following three categories:

(1) t_1, t_3, t_5 - These events define discontinuous forces of relatively small magnitude, therefore only a restart of the integration procedure is required.

(2) t_2 - This event defines an externally applied impulsive load requiring a momentum balance and restart of the integration procedure.

(3) t_4, t_6 - These events define impulsive loading, due to impact between bodies of the system, and mass capture and release. Supplemental equations are required for momentum balance and a restart of the integration procedure is required.

The effects of the various events at logical times t_1 to t_6 on the position, velocity, and acceleration of the barrel are shown in Fig. 7. The DADS computer program required 14 seconds on an Itel AS6 computer to execute one cycle of the weapon system.

4.2. The Spring-Reset Trip-Plow Mechanism. A spring-reset plow-share mechanism model is shown in Fig. 8, in its initial configuration, just prior to impact between the plow-share tip and a rock buried in the ground. The model consists of six moveable rigid bodies, identi-fied as follows: body 1 - plow-share and standard, body 2 - lower link, body 3 - rear toggle link, body 4 - front toggle link, body 5 - u-bolt, and body 6 - combined plow-frame and tractor mass. The bodies are connected by various rotational joints, as illustrated in Fig. 8 and the entire tractor-plow system is initially moving to the right at
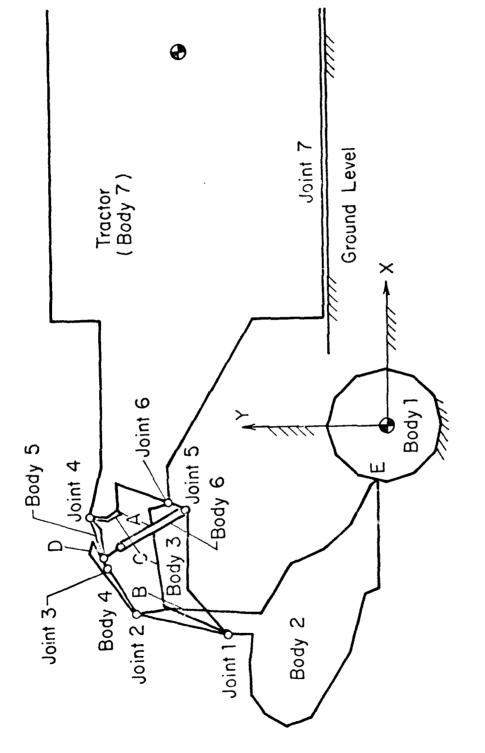
Figure 8. Spring-Reset Trip-Plow Mechanism.

2 meters per second, along a horizontal translational joint between
tractor and ground. A spring-damper-actuator combination is connected
between the u-bolt and rear toggle link, to simulate the spring reset
device. In addition, five potential contact points, where impact
between adjacent bodies will occur, are identified by the letters A-E,
as follows: A - contact between the u-bolt and main frame, B - contact
between the shank and lower link, C - contact between the lower link
and main frame, D - contact between the front and rear toggle links,
and E - contact between the plow-share tip and rock embedded in the
ground.

Contacts are simulated by attaching markers at some distance from
the point of contact on adjacent bodies, such that the nonzero vector
connecting the two points passes through the point of impact and defines
a normal to the surface at that point. These markers are simply
modified spring-damper-actuator combinations, in which provisions are
made to control spring rates, damping coefficients, and actuator forces
as needed. The elements can play various roles in the simulation.
Logical variables ℓ^k defining logical event times t_m of impact are
formulated in terms of spring-damper deformed and undeformed lengths as

$$\ell^k = \ell_{ij} - \ell_{0ij}$$

The constant ℓ_{0ij} is selected so that impact occurs at $\ell_{ij} = \ell_{0ij}$, hence
$\ell^k = 0$. As noted earlier, to facilitate event prediction, the
differential equation

$$\dot{\ell}^k = \dot{\ell}_{ij}, \qquad \ell^k(0) = \ell_{ij}(0) - \ell_{0ij}$$

is formulated and solved for ℓ^k. The system state and event times are
thus determined precisely when $\ell^k = 0$ or $\ell_{ij} = \ell_{0ij}$.

At this point one has several options in continuing the simulation:

(1) Define spring and damper coefficients or actuator forces in
     the element to represent contact characteristics and restart
     the integration. (They remain active until ℓ^k becomes zero
     again, indicating separation, at which time they are set to
     zero.)

(2) Define a coefficient of restitution e; the normal impulse
     direction and location is determined by the element's direc-
     tion and location. Perform momentum balance to determine
     jump discontinuities in velocity and restart integration.

(3) If the coefficient of restitution is zero, and the bodies
     are to be locked together, a constraint equation of the form

$$\phi = \ell_{ij} - \ell_{0ij} = 0$$

     is introduced and appended to other existing constraints.

The impacts encountered in this simulation usually occur between
bodies driven together (and then held together for some time) by large
forces. Relative displacement between the impacting bodies after
impact is usually negligible. Assuming a coefficient of restitution
of zero prevents multiple impacts and simplifies the computer logic.
The procedure for most impacts in the above simulation is then to set
e = 0 in the second option above and perform momentum balance to get
new velocities. Then set spring and damper coefficients as in option
1 to represent physical contact between bodies and restart the
integration.

Logical variables ℓ^1 to ℓ^5 are formulated for the five contact
points A to E. Logical times are defined for the spring-reset plow
model (not necessarily in the order of occurrence) as follows:

(1) $t_0 = 0$: The tip of the plow contacts the rock ($\ell^5 = 0$;
     perform momentum balance with coefficient of restitution
     e = 0.5; restart integration).

(2) t_1: The tip of the plow makes second contact with the rock
     ($\ell^5 = 0$; repeat above.)

(3) t_2: The u-bolt contacts plow frame ($\ell^1 = 0$; perform
     momentum balance with coefficient of restitution, e = 0;
     activate spring and damping coefficients; restart integra-
     tion).

(4) t_3: The lower link contacts plow frame ($\ell^3 = 0$; perform
     momentum balance with coefficient of restitution, e = 0;
     activate spring and damping coefficients; restart integration).

(5) t_4: The lower link and standard separate ($\ell^2 = 0$; set spring
     and damping coefficient to zero and continue).

(6) t_5: Impact between toggle links ($\ell^4 = 0$; perform momentum
     balance with coefficients of restitution, e = 0; activate
     spring and damping coefficients; restart integration).

(7) t_6: Lower link breaks contact with plow frame ($\ell^3 = 0$; set
     spring and damping coefficient to zero and continue).

(8) t_7: u-bolt breaks contact with plow frame ($\ell^1 = 0$); set
     spring and damping coefficient to zero and continue).

(9) t_8: Impact between lower link and standerd (x^2 = 0; perform
momentum balance with coefficient of restitution $e = 0$;
activate spring and damping coefficients; restart integration).

The predicted motion of the plow-share mechanism is shown in Fig. 9 as
follows: The tip of the plow makes contact with the rock at time = 0.0
seconds. The plow tip fails to clear the rock and impacts it again at
0.33 seconds. The impacts impart angular velocity to the plow-share,
causing it to move rearward and upward. This motion drives the toggle
links upward, bringing spring 1 into tension. The u-bolt and lower
link come into contact with the plow frame (contact points A and C) at
0.12 and 0.40 seconds, respectively. Contact at B between the standard
and lower link is broken at C.33 seconds.

Contact at C between the lower link and frame stops upward movement
of the plow-share and the reset cycle begins. Stored energy in the
spring rapidly collapses the toggle links. This action causes a rapid
change in angular displacement of the plow-share, with only a small
effect on its vertical displacement. At 0.55 seconds, the toggle links
have reset (contact at D).

The lower link and u-bolt break contact with the frame at 0.57
and 0.70 seconds, respectively. It is interesting to note that the
toggle action results in the plow-share being brought to within 20° of
horizontal, while its center of mass is still C./5 meters above ground.
The plow-share therefore re-enters the ground at a shallow angle, pre-
venting the mechanism from being tripped again. Finally, at about
0.86 seconds, contact occurs at stop B and the mechanism regains its
approximate initial configuration. The DADS computer program required
38 seconds on an Itel AS6 computer to execute one cycle of the trip-
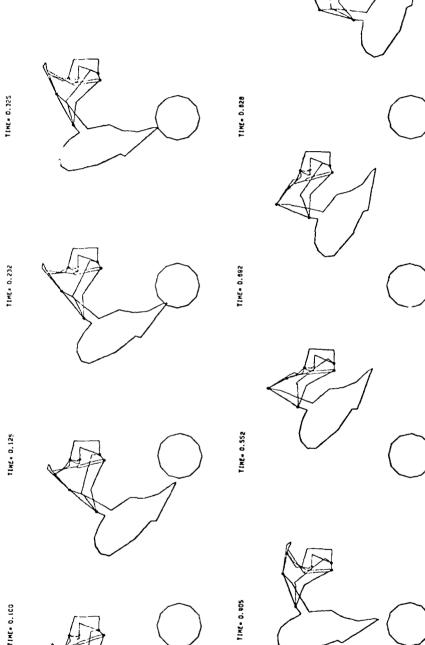plow mechanism.

Figure 9. Spring-Reset Trip-Plow Reset Cycle.

113

REFERENCES

1. Huang, R.C., Haug, E.J., Andrews, J.G., "Sensitivity Analysis and Optimal Design of a Mechanical System with Intermittent Motion," _Journal of Mechanical Design, Transactions of ASME_, Vol. 100, 1978, pp. 492-499.

2. Ehle, P.E., Haug, E.J., "A Logical Function Method for Dynamic and Design Sensitivity Analysis of Mechanical Systems with Intermittent Motion," to appear in _ASME Journal of Mechanical Design_.

3. Kees, W. and Teodorescu, P.P., _Applications of the Theory of Distributions in Mechanics_, Abacus Press, Tunbridge Wells, Kent, England, 1974.

4. Huang, R.C. and Haug, E.J., "A Distribution Theoretic Method for Design Sensitivity Analysis of Mechanisms with Intermittent Motion; Part I: Theory, Part II: Applications," _Developments in Mechanics_, Proceedings of the 16th Midwestern Mechanics Conference, Vol. 10, pp. 115-126.

5. Ehle, P.E., _Dynamic Analysis and Design Sensitivity Analysis of Mechanisms with Intermittent Motion_, Technical Report No. 48, Division of Materials Engineering, College of Engineering, The University of Iowa, Iowa City, IA, February, 1979.

6. Wehage, R.A., Haug, E.J., "Generalized Coordinate Partitioning for Dimension Reduction in Analysis of Constrained Dynamic Systems," to appear in the _ASME Journal of Mechanical Design_.

7. Haug, E.J., Wehage, R.A., and Barman, N.C., "Dynamic Analysis and Design of Constrained Mechanical Systems," to appear in _The ASME Journal of Mechanical Design_.

8. Greenwood, D.T., _Principles of Dynamics_, Prentice-Hall, Englewood Cliffs, N.J., 1965.

9. Haug, E.J., and Arora, J.S., _Applied Optimal Design_, John Wiley & Sons, New York, NY, 1979.

10. Kaplan, W., _Advanced Calculus_, Addison-Wesley, Reading, Massachusetts, 1973.

114

APPLICATIONS OF DELAY FEEDBACK IN CONTROL SYSTEMS DESIGN

N. P. Coleman, E. Carroll, D. Lee and K. Lee
US ARRADCOM
Dover, NJ 07801

ABSTRACT: Necessary and sufficient conditions for exact state recon-
struction using delays are discussed together with an example in which the
technique is implemented in real time using an 8080/8085 microprocessor.
Also, a frequency domain technique for synthesizing certain feedback control
laws with delays is developed and several examples discussed.

I.   INTRODUCTION: In designing a control system using optimal control
theory or classical frequency domain techniques, one often encounters sit-
uations in which certain required signals or states of the system are
unavailable by direct measurement. In modern control design this problem is
usually handled by implementing some form of reduced order or full order
observer which provides an asymtotic estimate of the unmeasured state. In
this paper a technique is developed for exact state reconstruction of unmeasured
system states using values of the measured variables, their delayed values
and the control variables on the maximum delay interval. Several examples
are discussed which demonstrate the application of this technique on a
laboratory servo system using an 8080 microprocessor.

A second application of delay feedback for frequency domain compensation
is also discussed. A frequency domain technique is developed for selecting
appropriate gain and delay parameters for synthesizing a feedback controller
using delays in the output and several applications as discussed.

II.   REAL TIME STATE RECONSTRUCTION USING DELAYS: In this section a
technique is presented for exact state reconstruction using delay feedback
of measured states of a control system and the values of the control input
over the delay interval. A real time application of this technique in a
servo control system using an 8080 microprocessor is also discussed. For
simplicity, consider the linear time invariant system:

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{1}$$

where x is an nx1 state vector, u is an rx1 control vector, A is an nxn
constant matrix, and B is an nxr constant matrix. Let the observation vector
y(t) be given by:

$$y(t) = Hx(t)$$

where y is a mx1 vector, and H is an mxn constant matrix. Let $o \leq h_1 < h_2 < \cdots < h < a$
be time delays.

The problem is to reconstruct the state x(t) from the measurements y(t),
y(t-h), $\cdots$, $y(t-h_\ell)$ and the measureable control vector u(s), $t-h_\ell \leq s \leq t$.

The following argument due to D. H. Chyung, Reference (1) provides the
basis for a real time state reconstruction algorithm discussed in the examples.
This argument makes use of the well known variation of parameter expression

i : the time response $x(t)$ of the system (1) given by:

$$x(t) = e^{A(t-h_i)}x(t-h_i) + \int_{t-h_i}^{t} e^{A(t-s)}Bu(s)ds \tag{2}$$

$$= e^{Ah_i}x(t-h_i) + \int_{-h_i}^{0} e^{-As}Bu(t+s)ds \qquad\qquad i = 1,2,\cdots,\ell \quad .$$

Multiplying both sides of equation (2) by He^{-Ah_i} results in the equation:

$$He^{-Ah_i}x(t) = Hx(t-h_i) + He^{-Ah_i}\int_{-h_i}^{0} e^{-As}Bu(t+s)ds$$

$$= y(t-h_i) + He^{-Ah_i}\int_{-h_i}^{0} e^{-As}Bu(t+s)ds \qquad\qquad i = 1,2,\cdots,\ell \tag{3}$$

in which the right hand side is completely known. Letting C denote the matrix given by:

$$C \overset{\Delta}{=} \begin{bmatrix} He^{-Ah_1} \\ He^{-Ah_2} \\ He^{-Ah_\ell} \end{bmatrix} \tag{4}$$

we can now write equation (3) in the form

$$Cx(t) = z(t) \tag{4}*$$
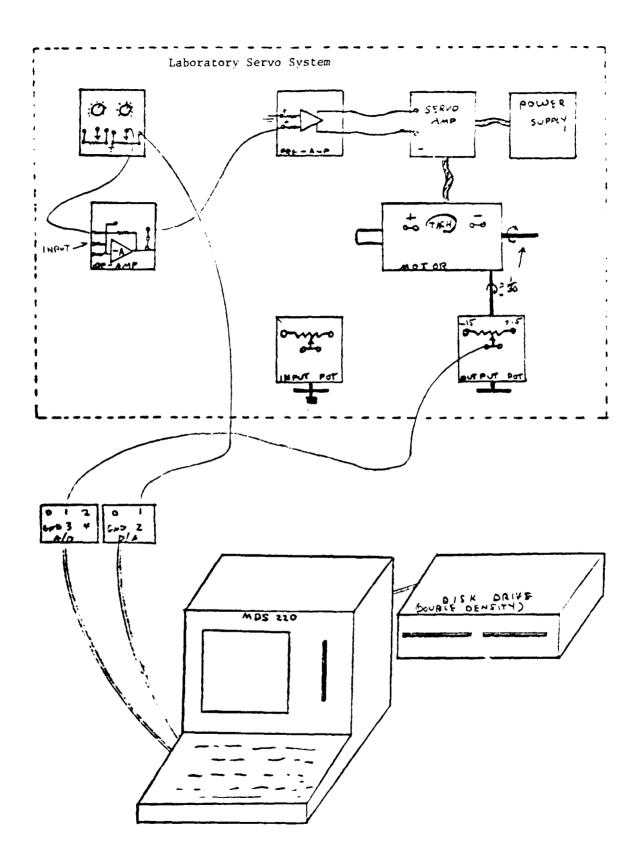
where;

$$z(t) = \begin{bmatrix} y(t-h_1) + He^{-Ah_1}\int_{-h_1}^{0} e^{-As}Bu(t+s)ds \\ \\ y(t-h_2) + He^{-Ah_2}\int_{-h_2}^{0} e^{-As}Bu(t+s)ds \\ \\ y(t-h_\ell) + He^{-Ah_\ell}\int_{-h_\ell}^{0} e^{-As}Bu(t+s)ds \end{bmatrix}$$

is a known $m\ell\times 1$ vector and C is an $m\ell\times n$ constant matrix depending on the parameters h_1, h_2, $\cdots,h_\ell$. If the matrix C has rank n, then equation (4)* can be written as:

$$x(t) = \left[C^T C\right]^{-1} C^T z(t) \tag{5}$$

where C^T denotes matrix transpose.

116

Figure 1

Figure 2

118
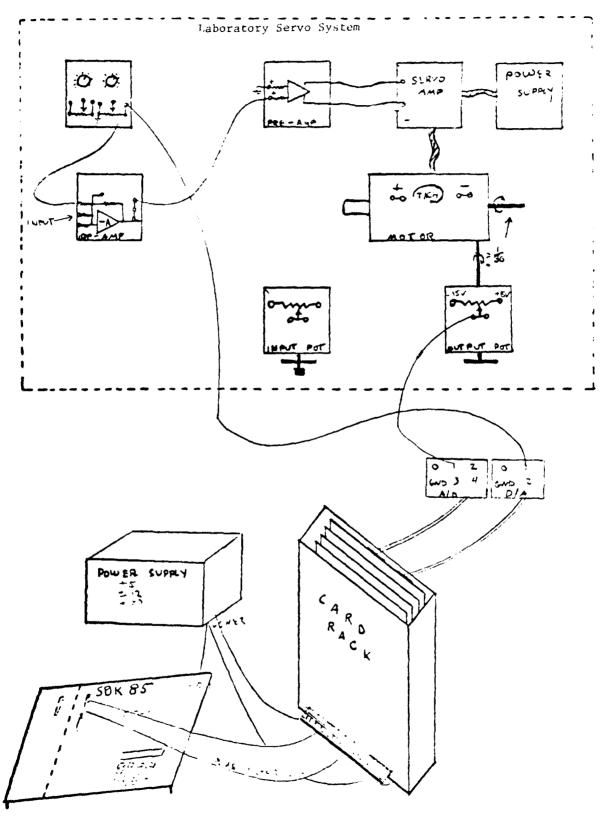
Equation (5) has several important implications; First, if the matrix $\left[C^T C\right]^{-1}$ exists, then the state $x(t)$ can be exactly reconstructed from the measurement $y(t)$, its delayed values and the input signal $u(t)$, $o \leq t \leq h$; secondly, the C matrix depends only on the delays $h_1, \cdots h_2$ so the right hand matrix calculation can be performed completely off line. This leaves only the relatively straight forward calculation of $x(t)$ and a matrix multiplication for on-line microprocessor computation. This latter comment is of particular importance in real time control applications in which relatively low speed microprocessors are utilized for control law implementation. The following result establishes the condition under which the matrix C has rank n.

Result: There exists a set of n delays $o \leq h_1 < h_2 \cdots < h_n \leq a$, for any $a > o$ such that the matrix C has rank n, if and only if rank $(Q) = n$, where
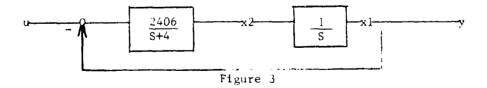
$$Q = \begin{bmatrix} H \\ HA \\ HA^{n-1} \end{bmatrix}$$

Proof: Let $a > o$ and assume rank $(Q) = n$. Then the row vectors of the matrices $He^{-Ah} h \varepsilon [o,a]$ contains n independent vectors since, if not, there exists $b \in R^n$ such that $He^{-Ah}b = O$ for all $h \varepsilon [o,a]$. Repeated differentation with respect to h gives $He^{-Ah}b = HA^{-Ah}b = HA^{n-1} e^{-Ah}b = 0$. This implies that the non zero vector $e^{-Ah}b$ is in the null space of the matrix Q and hence rank $(Q) < n$.

Conversely, assume rank $C = n$, then rank $(Q) = n$ since, if not, there exists $b \neq o \varepsilon R^n$ such that $Hb = HAb = \cdots = HA^{n-1}b = o$. This implies $He^{-Ah}b = o$ for all h and hence rank $C < n$, a contradiction.

Example: Evaluation of the state reconstruction technique given by equation (5) was carried out on an 8080 microprocessor development system which was in turn interfaced with a laboratory servo system as shown in Figure 1. In this example the system state vector is given by $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ where x_1 is the motor shaft output position and x_2 is the motor shaft velocity. The measured signal is x_1 and x_2 is reconstructed using equation (5). Once the software was developed and debugged the program was down-loaded to a single board 8085 microprocessor shown in Figure 2, for faster execution.* The block diagram of the servo system without tack feedback is shown in Figure 3.

Figure 3

* The 8085 configuration shown in Figure 3 is currently being used to evaluate digital control concepts for the XM97 turret system.

119

The state space equation is given by:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2406 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2406 \end{bmatrix} u \qquad (6)$$

$$y = x_1 = \begin{bmatrix} 1, 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The state transition matrix for this system is readily computed to be

$$e^{At} = e^{-2t} \begin{bmatrix} \cos 49t + \frac{2}{49}\sin 49t & \frac{1}{49}\sin 49t \\ \\ -49.1\sin 49t & \cos 49t - \frac{2}{49}\sin 49t \end{bmatrix} \qquad (7)$$

with the associated C matrix of equation (4) being given by:

$$C = \begin{bmatrix} H \\ He^{-Ah} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ e^{2h}(\cos 49h - \frac{2}{49}\sin 49h) & \frac{-e^{2h}}{49}\sin 49h \end{bmatrix} \qquad (8)$$

with $h_1 = 0$ and $h_2 = h$.

For values of $h \neq \frac{2n\pi}{49}$, the matrix C is non singular and we may compute $(C^T C)^{-1} C^T = C^{-1}$ directly as

$$C^{-1} = \begin{bmatrix} 1 & 0 \\ \\ 49\cot(49h) - 2 & -49e^{-2h}\csc 49h \end{bmatrix} \qquad (9)$$

Using equation (5) one obtains the required state reconstruction equation for $x_2(t)$ in terms of the measurements $x_1(t)$, $x_1(t-h)$ and $u(s)$, $k-h \leq s \leq t$.

$$x_2(t) = \begin{bmatrix} 49\cot(49h) - 2 \end{bmatrix} x_1(t) - (49e^{-2h}\csc 49h)x_1(t-h)$$

$$+ 2406 \begin{bmatrix} e^{2h}\cos 49h - \frac{2e^{2h}\sin 49h}{49} \end{bmatrix} \int_{-h}^{0} \frac{-e^{2s}}{49}(\sin 49s)u(t+s)ds \qquad (10)$$

$$+ 2406 \begin{bmatrix} \frac{-e^{2h}\sin 49h}{49} \end{bmatrix} \int_{-h}^{0} (e^{2s}\cos 49s + \frac{2e^{2s}}{49}\sin 49s)u(t+s)ds$$

The implementation of this state reconstruction algorithm was carried out on an 8080 microprocessor with a delay value h = .01 sec. The position output state was sampled at 2.2 millisecond intervals and the accuracy of the A/D and D/A converters was 12 binary bits.

120

Figure 4a compares the actual tach output signal representing the $x_2(t)$ state with the microprocessor output signal which attempts to reconstruct $x_2(t)$ via equation (10) using only the first two terms of this expression. Note: In this case equal weightings must be used for $x_1(t)$ and $x_1(t-n)$. The effects of measurement noise are readily apparent in this figure. Figure 4b again compares measured tach output with the microprocessor output signal, however, in this case the full state reconstruction equation (10) is implemented. This implementation is seen to give a very accurate state reconstruction which is less sensitive to measurement noise.

III.   FREQUENCY DOMAIN CONTROL SYNTHESIS USING DELAY FEEDBACK

Several papers, (see Reference (5) and (6)) have appeared in the recent literature which address the problem of feedback control using delays. Reference (6) develops several feedback control laws using delays in the state and derivative of the state which are shown to drive the full state of the system to zero and keep it there. The constructions, however, have limited utility in servo control applications since they assume first that the control space has the same dimension as the state and all states of the system are accessible for on-line measurement.

In this section we consider a restricted class of delay feedback controllers shown in Figure 5. This configuration has proved quite useful in turret and servo control applications in which G(s) represents the open loop transfer function between the command input and the position output. The two design parameters introduced by delayed feedback are seen to be K, the feedback gain, and T, the feedback time delay. The reason for chosing the two feedback gains in the form K and K-1 differing by unity in the general case, will be made clear below. The equivalent feedback transfer function, H(s), for the system in Figure 5 is:

$$H(s) = K - (K-1)e^{-Ts} \qquad (11)$$

We may represent the e^{-Ts} term by its equivalent Taylor series form as:

$$e^{-Ts} = 1 - Ts + \frac{T^2 s^2}{2} - \frac{T^3 s^3}{6} + \ldots$$
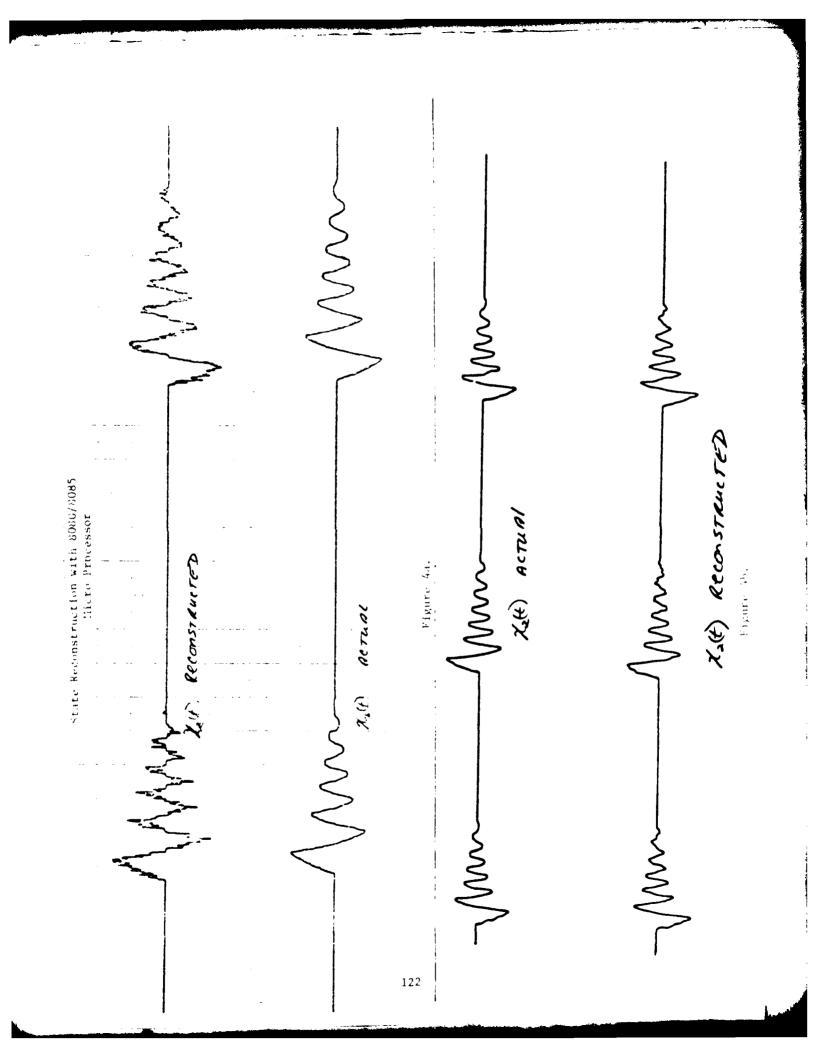
The frequency band of primary interest from a stability and transient response point of view in $[s: \ |G(s)H(s)| \doteq 1]$ or $[s: \ s = w_c]$ where w_c denotes the gain crossover frequency of the compensated open loop system. Setting $s = jw$ and assuming $|wT| << 1$, we may approximate e^{-Ts} by the first two terms of its Taylor series expansion or;

$$e^{-Ts} = 1 - Ts = 1 - jwT \qquad (12)$$
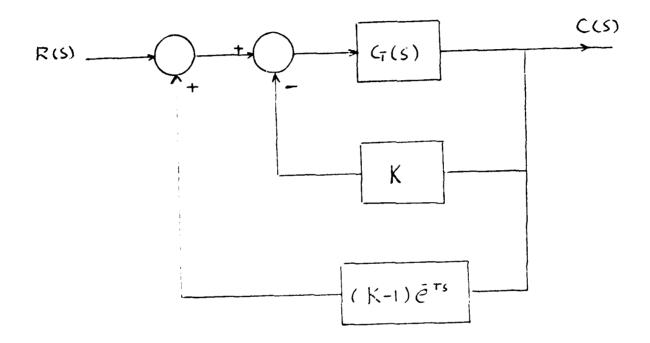
Substituting (12) into equation (11) yeilds;

$$H(s) = K - (K-1)(1-jwT) = 1 + j(K-1)Tw \qquad (13)$$

Since $K > 1$ will be required, this corresponds to a phase lead network (on a first order approximation basis) in the controller feedback path. If this phase lead term is properly positioned in frequency, it will produce a stabilizing effect upon the control systems unit step or impulse response characteristics. As will be seen in the examples, the time delay or feedback gain can be adjusted

State Reconstruction with 8080/8085
Micro Processor

$X_1(t)$ RECONSTRUCTED

$X_1(t)$ ACTUAL

Figure 4a.

$X_2(t)$ ACTUAL

$X_2(t)$ RECONSTRUCTED

Figure 4b.

Feedback Control with Delay

Figure 5

to provide any desired damping in the system response. The procedure for introducing a lead network effect around $w = w_c$ using delayed feedback can now be developed.

First, choose w_c such that $\left| G(jw_c) \right| = \dfrac{1}{\sqrt{2}}$

Second, select the feedback time delay, T, such that $Tw_c < < 1$. The choice of $Tw_c = .1$ is reasonable and is used in the examples. For this choice, the first term disregarded in the series expansion has magnitude .005 at w_c and rapidly becomes smaller for higher frequencies. Third, select the feedback gain parameter, K, such that the lead time constant becomes effective at or near $w = w_c$ i.e. $(K-1)T = \dfrac{1}{w_c}$. Note under this condition using step 1 and equation (8), that;

$$\left| G(w_c)H(w_c) \right| = \left| G(w_c) \right| \left| H(w_c) \right|$$

$$\cong \frac{\sqrt{2}}{\sqrt{2}} = 1 \text{ and } K = 11$$

The delayed feedback design procedure thus is seen to be straight forward in concept. The effect of the particular delayed feedback configuration discussed here is to replace the more standard tach feedback stabilization loop. When the delay time and gain parameters are properly chosen, system response characteristics may be improved substantially.

Example:

We consider first a simple laboratory servo system whose open loop transfer function, $G(s)$, is given by;

$$G(s) = \frac{600}{s(1+\frac{s}{4})} \tag{14}$$

The -3db crossover frequency, w_c, of the open loop transfer function $G(s)$ is 56 rad/sec and the delay time, T, is computed from step 3 and satisfies $10T = \dfrac{1}{56}$ or $T = .0017$ sec. The gain K is fixed and satisfies the relation;

$$K - 1 = \frac{1}{Tw_c} = \frac{1}{.1} = 10$$

Due to limitations of the 8080 microprocessor, the above design using a delay of 1.7 ms could not be implemented. The smallest delay which could be implemented with the 8080 was 2.2 ms. The performance of this design for a step input command is shown in Figure 6b. Figure 6 illustrates that the effective damping introduced by the feedback delay can be further increased by increasing the delay parameter T. The desired damping can also be achieved by adjusting the gain parameter K.

To evaluate the effects of delay parameters which were too small for implementation on an 8080 microprocessor, simulations were run for values of T = .8 msec, 1.7 msec, 2.2 msec, 4.4 msec, 8.8 msec and 17.6 msec, using the servo transfer function (14). These results are shown in Figures 7 - 12. Deficiencies in the linear model of the servo system are readily apparent since the simulations indicate more damping than is evident in the test results of Figure 6 and Figure 12

124

indicates an instability with the 17.6 msec delay in constrant to the over
damped response in the hardware test shown in Figure 6e.
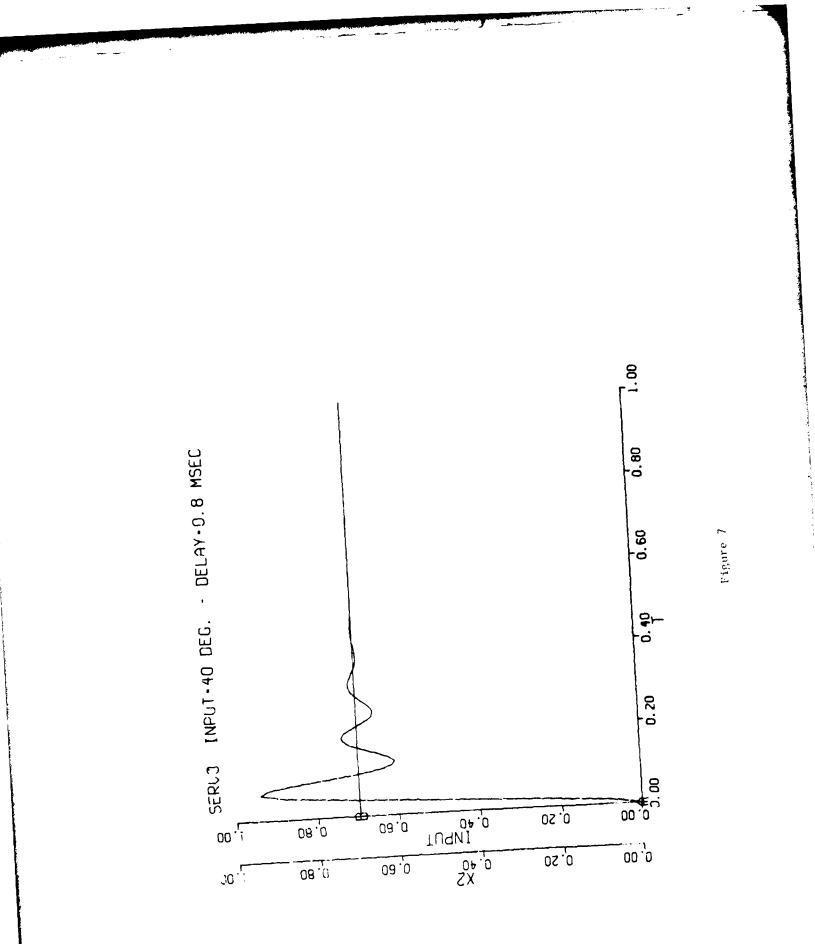
Example:
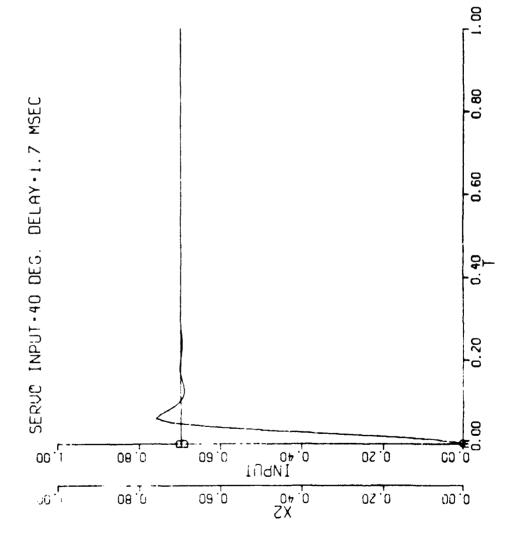
In this example we illustrate the application of the delay feedback control
synthesis technique to the design of a controller for an XM97 helicopter turret
control system shown in Figure 13. The transfer function block diagram of this
system is shown in Figure 14. The -3db crossover frequency for the open loop
system (tach loop open) was computed to be 20 rad/sec resulting in a feedback
time delay of .005 sec. The step response of the original XM97 design is
shown in Figure 15 and that of the delay feedback design in Figure 16. The
latter design exhibits a dramatic improvement with respect to overshoot and
settling time. This improvement can be explained partially by the fact that the
original system uses motor tachometer feedback for stabilization while the delay
feedback design effectively uses actual turret position and rate for feedback
stabilization. Figures 17 - 20 also show the effects of increasing and decreasing
the delay feedback parameter. Saturation, columb friction and deadband non-
linearities are included in the simulation.

IV.   CONCLUSION: Applications of delay feedback for state construction
and feedback control are presented together with simulation results and examples
of actual implementations using Intel 8080 and 8085 microprocessors. These
examples demonstrate the practicality of the ideas and suggest that these tech-
niques may provide a useful adjunct to the more standard frequency domain and
state variable techniques for estimation and control applications.
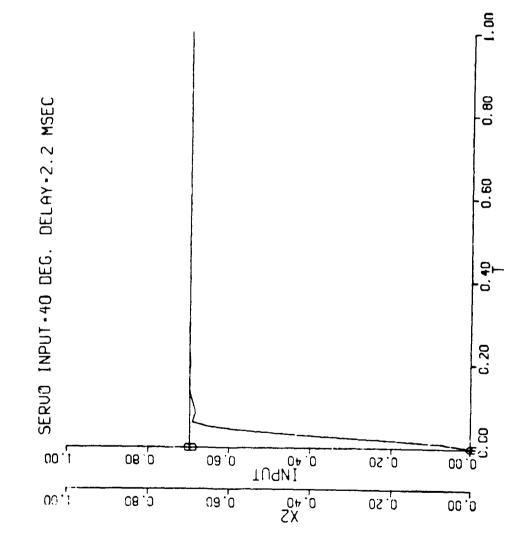
G(s)

Figure 6a

2.2 ms Delay, K=11

Figure 6b

4.4 ms Delay, K=11

Figure 6c

8.8 ms Delay, K=11

Figure 6d

17.6 ms Delay, K=11

Figure 6e

126

SERL3 INPUT·40 DEG. - DELAY·0.8 MSEC

Figure 7

127

SERVO INPUT - 40 DEG. DELAY - 1.7 MSEC

Figure 8

128

SERVO INPUT=40 DEG. DELAY=2.2 MSEC

Figure 9

129

SERVO INPUT=40 DEG. DELAT=4.4 MSEC

Figure 10

130

SERVO INPUT=40 DEG. DELAY=8.8 MSEC

Figure 11

131

SERVO INPUT=40 DEG. DELAY=17.6 MSEC

Figure 12

132

$G_1(s) = \frac{1}{s}$,

$G_2(s) = \frac{1}{15.7s}$,

$G_3(s) = \frac{1}{0.00025s}$,

$G_4(s) = \frac{0.02}{1 + 0.002s}$,

$G_5(s) = \frac{7.5}{1 + 0.002s}$,

$G_6(s) = \frac{535}{1 + 0.005885s}$,

$G_7(s) = \frac{1}{620s}$,

$G_8(s) = \frac{0.00412s}{1 + 0.1s}$

$K_1 = K_s = 1$, $K_b = 0.0192$, $K_c = 500000$, $f_m = 0$, $f_\ell = 50$,

N = 620 (Azimuth)

N = 810 (Elevation)
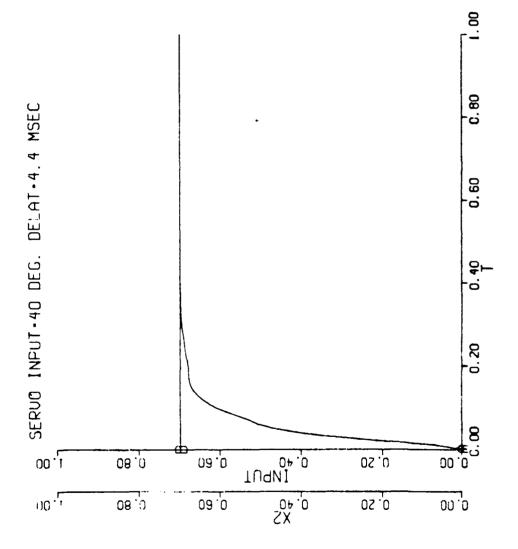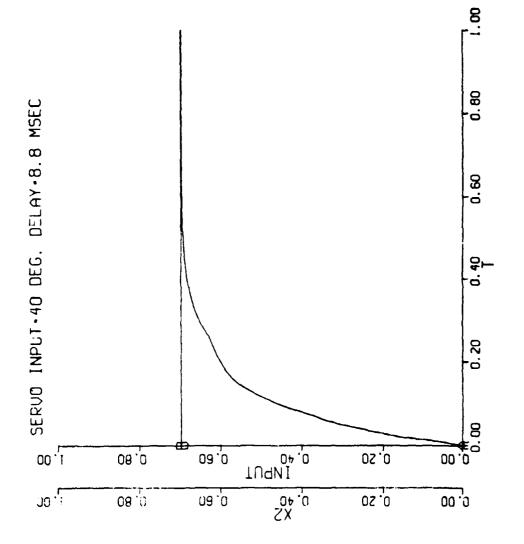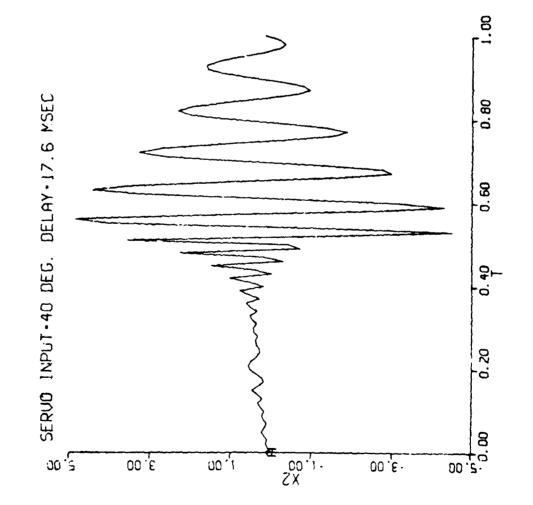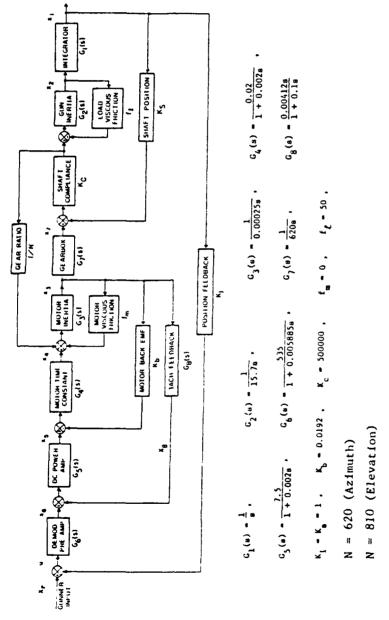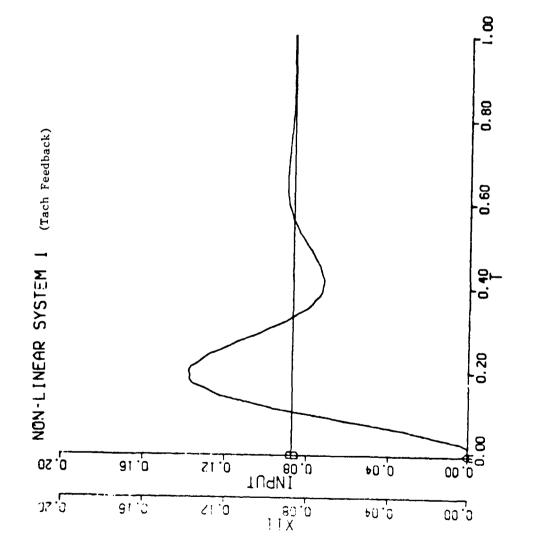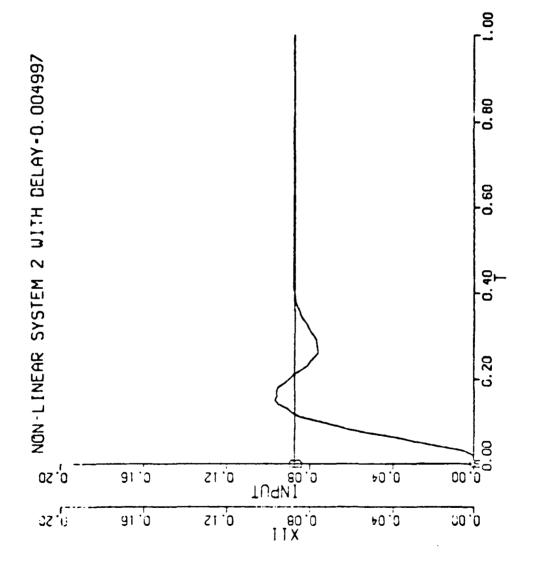
Fig. 14. EXISTING TURRET CONTROL SYSTEMS – AZIMUTH AND ELEVATION CHANNELS

133

NON-LINEAR SYSTEM 1 (Tach Feedback)

Figure 15

134

Figure 16

135

NON-LINEAR SYSTEM 2 WITH DELAY=0.002997

Figure 17

NON-LINEAR SYSTEM 2

Figure 18

137

NON LINEAR SYSTEM 2-DELAY=0.0055

Figure 19

138

NON-LINEAR SYSTEM 2.DELAY-0.005997

Figure 20
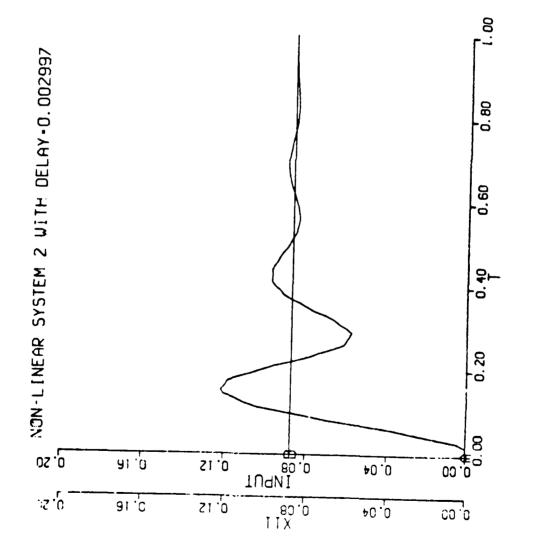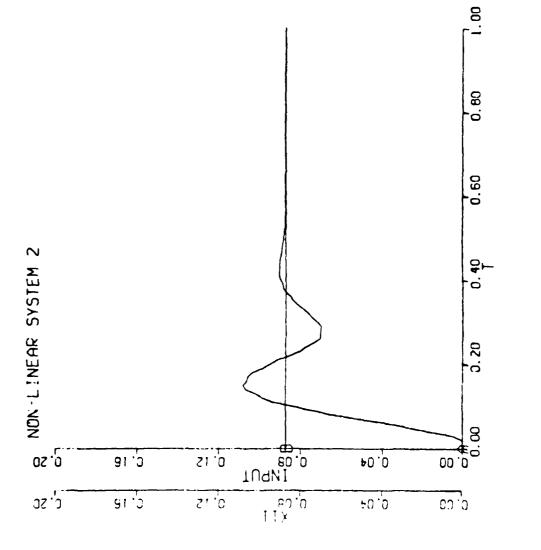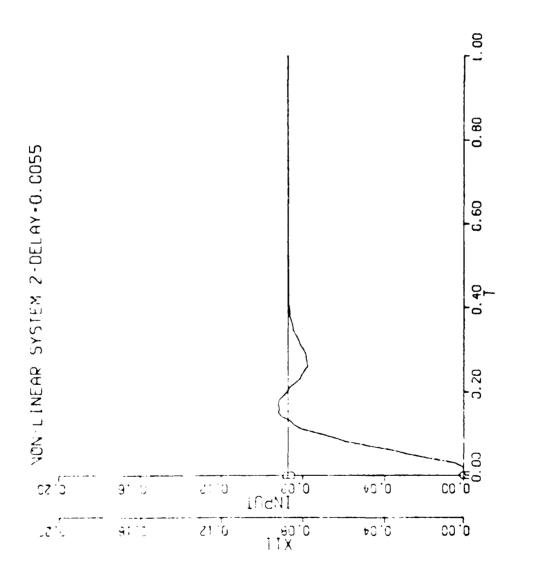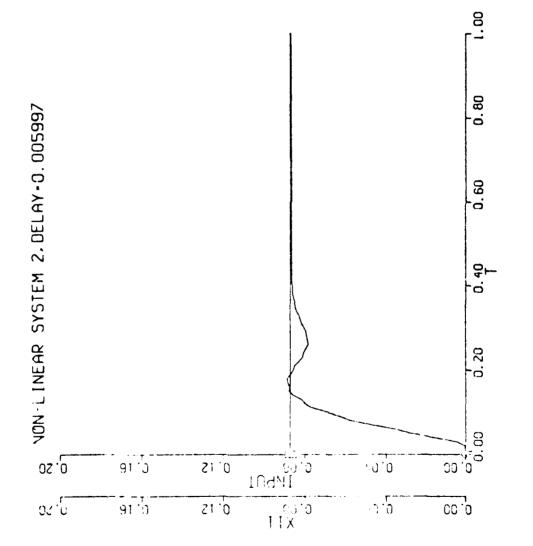
References

1.   D. H. Chyung, "On A Method For Reconstructing Inaccessible State Variables Using Time Delays." Unpublished Paper.

2.   N. K. Loh & D. H. Chyung, "State Reconstruction From Delayed Observations." Proceedings Of The Fifth Annual Pittsburg Conference On Modeling & Simulation, Volume 5, April 1974.

3.   D. G. Luenberger, "An Introduction To Observers." IEEE Trans. Automat. Contr., Volume AC-16, No. 6, December 1971.

4.   J. D. Gilchrist, "N-Observability For Linear Systems." IEEE Trans. Automat. Contr., Volume AC-11, No. 3, July 1966.

5.   A. Thowsen, "On Pointwise Degeneracy, Controllability & Minimal Time Control Of Linear Dynamical Systems With Delays." Int. J. Control, Volume 25, No. 3, 1977.

6.   B. Asner & A. Halanay, "Indirect Delay-Feedback Control Of Linear Systems." Proc. Of 14th Allerton Conference.

7.   A. Thowsen, "Function Space Null Controllability By Augmented Delay Feedback Control." IEEE Trans. Automat. Contr., April 1976.

8.   V. M. Popov, "Delay-Feedback, Time-Optimal, Linear Time-Invariant Control Systems." Ordinary Differential Equations, L. Weiss, Ed. New York; Academic 1972.

AN ADAPTIVE LEAD PREDICTION ALGORITHM FOR MANEUVERING TARGET ENGAGEMENT

Pak T. Yip & Norman P. Coleman
USA ARRADCOM
Dover, NJ 07801

ABSTRACT. An algorithm concept which processes target bearing and range input data and provides "optimal" estimates of target position, velocity and acceleration a time-of-flight in the future is discussed. Since the algorithm concept involves certain important statistical assumptions about target acceleration dynamic models, these assumptions will be discussed in detail along with several important methods used in the model identification process. Secondly, the filter algorithm itself will be discussed. This algorithm involves the parallel processing of target range and bearing data by several extended Kalman Filters corresponding to distinct maneuver characteristics of anticipated target vehicles. At time of fire the filter with the largest computed likelihood function is selected for lead prediction. Finally, results of simulation studies in which actual target path data is used to generate filter input data for hit probability evaluation is discussed. Comparisons are made between the adaptive algorithm and non-adaptive first order algorithms.

I. INTRODUCTION. This paper describes a multiple model adaptive Kalman Filter approach to the problem of estimating and predicting the position, velocity and acceleration states of tank targets of varying maneuverability. The estimation and prediction problem presupposes that the range and angle DATA (measurements corrupted by Gaussian white noise) is available. The target dynamics is described by a system equation. Our solution to this problem is an adaptive algorithm implementable in real time with a microprocessor to compute target position a projectile time of flight in the future. This study begins with the selection of the Antitank Missile Test (ATMT) Phase II data[1] to identify the filter acceleration models. It consists of three dimensional (x,y,z) position data recorded at approximately 10 samples per second. Maximum likelihood identification method is applied to this data to identify a finite set of Markov Acceleration Models which are representative of a broad spectrum of vehicle maneuvers considered likely to occur in actual engagements. These models provide the required state variable description of the target dynamics used in the formulation of the multiple model extended Kalman Filter Algorithm for lead prediction. The extended Kalman Filter is required in this application as a result of nonlinearities induced by target coordinate transformations and non-linear measurement equation.

The adaptive lead prediction concept is based on the simultaneous (parallel) processing of the discrete extended Kalman Filters corresponding to the distinct target modes identified from the ATMT data. The likelihood function associated with each filter is computed up to the time of fire of the weapon, and the filter having the greatest likelihood is automatically selected for lead prediction.

141

In the present study, only the azimuth and range information of the target is processed in the filter with the target elevation considered constant. The performance of this design is examined with a Monte Carlo simulation and the sensitivity of the lead estimates to measurement noise, level of target maneuver, range sampling rate, and time of flight of projectile are analyzed to determine the feasibility of using this algorithm for fire control lead prediction against various maneuvering targets.

II. DATA ANALYSIS. The ATMT data consists of six tracks produced by a M60A1 tank, a Scout Vehicle and a Twister Vehicle undergoing evasive maneuvers. The M60A1 tank is capable of speeds of 10 to 16 miles per hour and with a maximum acceleration of approximately .3g. The Scout is an armored reconnassiance vehicle capable of moving at a speed of 15 to 25 miles per hour and a maximum acceleration of approximately .5g. Since our only interest is in modeling the acceleration, the position data is sampled at a frequency of 2 cps and twice differentiated to obtain the acceleration estimates which are then resolved into along-track and cross-track components. The power spectral density of this data is computed by the maximum entropy method₃ which assumes the data is generated by an autoregressive process. The power spectral desity $S(f)$ is given by

$$S(f) = \frac{2\sigma_q^2}{1 - \sum_{i=1}^{M} \alpha_i \exp(-j2\pi f i)}$$

where α_q is the standard deviation of a Gaussian noise process; α_i is the i-th coefficient of the autoregressive process; M is the number of coefficients, and the coefficients α_i' are estimated recursively .

The number of the autoregressive coefficients is usually larger than 3 which is not desirable for Kalman Filtering. However, the power density spectrum affords enough information for estimating essential poles and zeros of a simplier model structure. Later the maximum likelihood identification program is used to fine tune the pole and zero estimates.

The simplified model determined from the spectral analysis has the following form:

$$A(s) = \frac{s + \gamma}{s^2 + \beta_1 s + \beta_2} q(s)$$

where $q(s)$ is the Gaussian noise process; $A(c)$ is the system acceleration; γ, β_1 and β_2 are parameters to be identified for the chosen tracks and each of the along-track and cross-track formulations.

III. DISCRETE EXTENDED KALMAN FILTER. The system and the measurement equations are readily defined as follows:

$$\underline{x}_k = \underline{f}(\underline{x}_{k-1}, dt) + \underline{q}(s)$$

$$\underline{z}_k = \underline{h}(\underline{x}_k) + \underline{Y}_k$$

142

where $\underline{x}_k$ is the system state vector at the discrete time kdt in the Cartesian coordinate system, $\underline{\phi}$ the system function containing all information about the system dynamics, $\underline{q}_k$ the plant noise vector, $\underline{z}_k$ the measurement vector, $\underline{h}(\underline{x}_k)$ a vector containing the true range and azimuth angle of the target position at the time kdt, $\underline{r}_k$ the measurement noise vector, and dt the time between two samples.

The necessary statistics and conditions are :

$$\text{cov}(\underline{q}_i , \underline{q}_j) = Q_i \delta_{ij}$$

$$\text{cov}(\underline{r}_i , \underline{r}_j) = R_i \delta_{ij}$$

$$\text{cov}(\underline{q}_i , \underline{r}_j) = 0, \qquad \forall i,j$$

$$E(\underline{x}_o) = \hat{\underline{x}}_o$$

$$\text{cov}(\hat{\underline{x}}_o) = P_o$$

where δ_{ij} is the Kronecker Delta.

Given the above, the discrete Extended Kalman Filter equations can be written as follows: The predicted state estimate vector is given by

$$\hat{\underline{x}}_{k+1|k} = \underline{\phi}(\hat{\underline{x}}_k, dt)$$

and the state error a priori covariance matrix by

$$P_{k+1|k} = \phi P_k \phi^T + Q_k$$

where

$$\phi = \frac{\partial \underline{\phi}(\underline{x}, dt)}{\partial \underline{\hat{x}}} \Big|_{\underline{x}=\underline{x}_k}$$

$$\underline{\phi} = \hat{\underline{x}}_k + \dot{\underline{x}}_k dt + \ddot{\underline{x}}_k (dt)^2/2$$

$$\dot{\underline{x}}_k = \frac{d\underline{x}_k}{dt}$$

$$\ddot{\underline{x}}_k = \dot{\underline{x}}_k \frac{\partial \dot{\underline{x}}_k}{\partial \underline{x}}$$

The updated state estimate vector can be written as follows:

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_{k+1|k} + K\tilde{\underline{z}}_{k+1}$$

where

$$\tilde{\underline{z}}_{k+1} = \underline{z}_{k+1} - \underline{h}(\hat{\underline{x}}_{k+1|k})$$

$$K = P_{k+1|k} H^T (H P_{k+1|k} H^T + R_{k+1})^{-1}$$

$$H = \frac{\partial \underline{h}(\underline{x})}{\partial \underline{x}} \Big|_{\underline{x}=\hat{\underline{x}}_{k+1|k}}$$

$$\underline{h}(\hat{x}_{k+1|k}) = \{[(x_1)^2 + (x_2)^2]^{1/2}, \tan^{-1}(x_1/x_2)\}$$

143

x_1, x_2 represent x, y position state estimates respectively in fixed Cartesian coordinates. The state error a posteriori covariance matrix is given by

$$P_{k+1} = P_{k+1|k} - KHP_{k+1|k}$$

and

$$Q_k = \int_{t_{k-1}}^{t_k} \Phi(t_k - \tau) Q_s \Phi^T(t_k - \tau) d\tau$$

where the continuous case plant noise covariance matrix, Q_s, is known.

The continuous time system dynamic equations used in deriving the discrete time equations are given by

$$\dot{x}_1 = x_3 , \qquad \dot{x}_2 = x_4$$

$$\dot{x}_3 = (x_3 A_a + x_4 A_c)/V$$

$$\dot{x}_4 = (x_4 A_a - x_3 A_c)/V$$

$$\dot{x}_5 = -\beta_{a1} x_5 - \beta_{a2} x_6 , \qquad \dot{x}_6 = x_5$$

$$\dot{x}_7 = -\beta_{c1} x_7 - \beta_{c2} x_8 , \qquad \dot{x}_8 = x_7$$

$$A_a = \frac{s + \gamma_a}{s^2 + \beta_{a1} s + \beta_{a2}} q_a$$

$$A_c = \frac{s + \gamma_c}{s^2 + \beta_{c1} s + \beta_{c2}} q_c$$

$$V = (x_3^2 + x_4^2)^{1/2}$$

where x_3 and x_4 are the corresponding x and y components of the velocity vector; A_a is the target acceleration along the velocity vector; A_c is the target acceleration perpendicular to the velocity vector.

With this filter, target range and angle measurements may be processed to generate target state estimate recursively. Before defining an adaptive filter procedure, the parameters of the Markov model need to be identified.

IV. LIKELIHOOD FUNCTION & MAXIMUM LIKELIHOOD IDENTIFICATION OF PARAMETERS.

Given a parameter vector $\underline{a}$, the probability of occurence of the measurement vector sequence $\underline{z}_k$ can be represented by a multivariate Gaussian distribution.

$$p(\underline{z}^k; \underline{a}) = p(\underline{z}_k | \underline{z}^{k-1}; \underline{a}) \cdots p(\underline{z}_2 | \underline{z}^1; \underline{a}) p(\underline{z}_1; \underline{a})$$

$$p(\underline{z}_k | \underline{z}^{k-1}; \underline{a}) = \frac{\exp(-[1/2]\tilde{z}_k S_k^{-1} \tilde{z}_k)}{(2\pi)^{n/2} (\det S_k)^{1/2}}$$

$$S_k = HP_{k|k-1} H^T + R_k$$

where

$P(\underline{z}^k; \underline{\alpha})$ = the likelihood function

n = number of elements in the measurement vector $\underline{z}_k$.

In order to identify the best parameter vector $\underline{\alpha}$ to give a maximum $P(\underline{z}^k; \underline{\alpha})$, we can equivalently minimize the negative log likelihood function:

$$M(\underline{z}^k; \underline{\alpha}) = \sum_{i=1}^{k} \{(1/2)\tilde{z}_i^T S_i^{-1} \tilde{z}_i + (1/2)\ell n(\det S_i)\} .$$

Since the term $(2\pi)^{n/2}$ in the likelihood function dose not contribute any interesting information it has benn eliminated in forming $M(\underline{z}^k; \underline{\alpha})$. the Gauss-Newton method is used in the minimization procedure.

$$\underline{\alpha}_{j+1} = \underline{\alpha}_j - \rho D^{-1} \frac{\partial M(\underline{z}^k; \underline{\alpha}_j)}{\partial \underline{\alpha}_j}$$

where ρ = 1 for this method, and D, the expected Hessian

$$D = E\{\frac{\partial^2 M(\underline{z}^k; \underline{\alpha}_j)}{\partial \underline{\alpha}_j^2}\} .$$

The test for convergence is given by

$$(\underline{\alpha}_{j+1} - \underline{\alpha}_j)^T D (\underline{\alpha}_{j+1} - \underline{\alpha}_j) < 10^{-3} .$$

V. PARALLEL FILTERS & ADAPTIVE ESTIMATION. Target state prediction for maneuvering ground targets have never been a simple task to undertake. The major uncertainty comes from the target driver's (stochastic) decision to maneuver. However, it appears there exists a maximum level of maneuver that the ground vehicles studied can attain. This maximum level provides a non-trivial range of dynamic motion that can be quantized to a finite number of maneuver levels. In this study, five filters are incorporated into the multiple model filter. Model M1 (Filter 1) is a simple 4 states constant velocity filter. The remaining 4 filters are identified with various maneuver levels.

The adaptive estimation is a straight forward decision making process. Measurement in range and azimuth angle are processed through the parallel filters. The filter having the largest likelihood function is automatically chosen to provide the best estimate for lead prediction and gun orders. Two concepts of adaptive prediction are examined. In concept A the likelihood functions account for the entire measurement history up to the time of fire. Thus this adaptive prediction concept is good against targets with constant maneuver level. In concept B, only the last ten samples prior to the firing time are used to compute the likelihood functions. This adaptive filter concept tends to be more sensitive to changes in target maneuver levels.

VI. SIMULATION. A Monte Carlo simulation of 100 runs was set up to process a number of 10 second segments from the ATMT data representing various maneuver levels for the M60, M113ter and Scout Vehicles. These segments of data are different in those used for the parameter identification tasks discussed earlier.

For evaluating the system performance, the perpendicular miss distance of the predicted line of sight from the real target position is defined as the prediction error E_p in meters. The firing time points are fixed for each segment under process. The performance indicator ph at each firing time point is defined as the ratio of the number of times that the prediction error E_p is less than 1.15 meters to the total number of runs. Actually, they are hit probabilities considering the prediction errors alone.

Assuming engagement range of approximately 2000 meters, $45°$ cross range (across the range vector), 1σ range measurement error of 2 meters, 1σ azimuth tracking error of 0.3 mils, a projectile speed of 1500 meters per second and using the adaptive prediction concept A, the hit probability results are illustrated in Figure 1 and summarized in the following table:

Target Type	Number of Cases, 7 Firing Points per Case	Mean ph Const. Velocity Prediction	Adaptive Prediction
M60A1	13	.41	.49
Scout	10	.27	.38
Twister	8	.20	.26

For an engagement range of approximately 1158, $60°$ cross range, 1σ range measurement error of 3 meters, 1σ azimuth tracking error of 0.3 mils, a projectile speed of 1158 meters per second and using the adaptive prediction concept B, the hit probability results are summarized in the following table:

Target Type	Number of Cases, 7 Firing Points per Case	Mean ph Const. Velocity Prediction	Adaptive Prediction
M60A1	6	.51	.56
Twister	6	.31	.37

With the latter conditions, the sensitivities of the system are observed for a particular maneuvering segment as shown in Figure 2. Figure 3 illustrates the system range (hence the time of flight of projectile) sensitivity. Figure 4 illustrates the system sensitivity to angular measurement noise. Figure 5 illustrates the system sensitivity to range measurement noise. Figure 6 illustrates the system sensitivity to range sampling rate.

VII. <u>DISCUSSION & FUTURE PLAN</u>. This study has demonstrated that maneuvering target acceleration may be adequately modeled as a discrete set of stationary Markov processes whose parameters can be identified off line. Parallel discrete extended Kalman filters have been used to successfully process range and angle measurements. The adaptive selection of the most appropriate filter at each time step, based on its largest likelihood function, has been accomplished on line. Representatives maneuver patterns and levels used in this study were taken from the ATMT data base. The results from the Monte Carlo simulations indicate that the performance of the multiple model adaptive filter design is generally comparable to a filter which is tuned to the target dynamics of that particular tracking interval. In particular, the results show that the adaptive prediction consistently performed better than the constant velocity prediction with an improvement in prediction range from 10 to 40 percent.

Since the range data is currently not a uniformly accessable measurement, the range sampling rate has been examined as an area of uncertainty.

146

together with range, angular measurement noise, and range measurement noise. The results indicate that the system performance for the azimuth channel is heavily dependent of angular measurement noise and projectile time of flight in terms of range, and is not very sensitive to range measurement noise and range sampling rate. The results also indicate that higher probability of hit can be obtained in the cross range geometry than in the down range (coming down along the range vector) geometry.

Implementation of this filter algorithm in real time with a state of the art microprocessor is in the planning stage. We have noticed that Bierman's UD factorization for the state error covariance propagation is a desirable feature considering computation accuracy and stability. Several variations of the existing filter algorithm are also under consideration. Finally, a complete real time simulation of the fire control system with the auto-tracker or human operator in the loop and filter modifications to improve maneuver detection will be subjects of our future work.

References
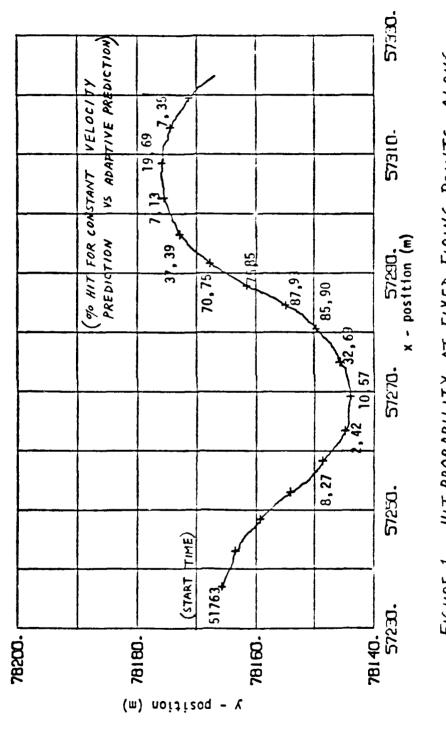
1. "Antitank Missile Test, Experiment FC019," Phase II, Project Analysis, ACN22273, USADEC, Fort Ord, CA, 30 June 1975.

2. N. K. Gupta, R. K. Mehra, "Computational Aspects of Maximum Likelihood Estimation & Reduction in Sensitivity Function Calculations," IEEE Trans. on AC, December 1974, pp 774-783.
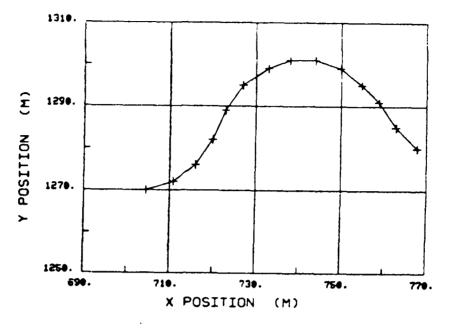
3. Tad J. Ulrych, Thomas N. Bishop, "Maximum Entropy Spectral Analysis & Autoregressive Decomposition," Reviews of Geophysics & Space Physics, Volume 13, No. 1, February 1975, pp 183-200.

4. G. J. Bierman, "Factorization Methods for Discrete Sequential Estimation," Academic Press, New York, 1977.

FIGURE 1. HIT PROBABILITY AT FIXED FIRING POINTS ALONG
A SEGMENT OF PATH 431, M60A1 TANK DATA

149

FIGURE 2. ATMT DATA SEGMENT 36
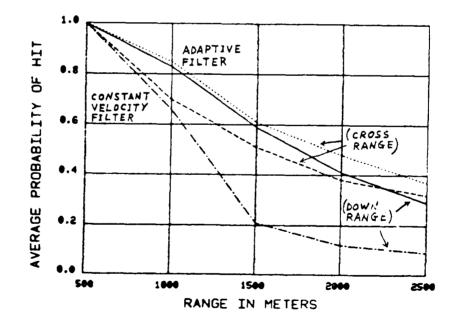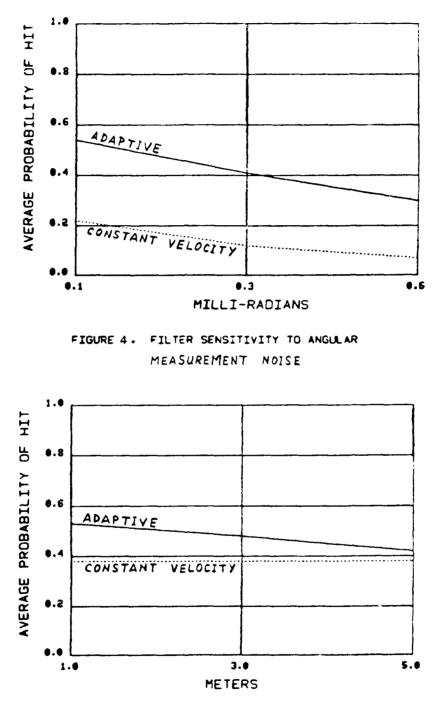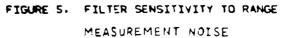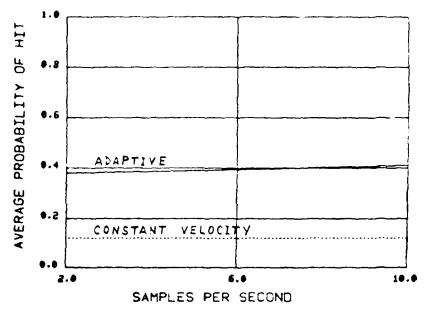
FIGURE 3. SYSTEM RANGE SENSITIVITY

150

FIGURE 4. FILTER SENSITIVITY TO ANGULAR
MEASUREMENT NOISE

FIGURE 5. FILTER SENSITIVITY TO RANGE
MEASUREMENT NOISE

151

FIGURE 6.   FILTER SENSITIVITY TO RANGE

SAMPLING RATE

ON VOLTERRA INTEGRAL EQUATIONS OF PULSE-CONVOLUTION TYPE

Edward W. Ross, Jr.
Staff Matnematician
US Army Natick R&D Command
Natick, MA 01760

ABSTRACT.   This paper presents a discussion of methods for solving
Volterra Integral Equations of first kind and pulse-convolution type.
The present context of the problem is the response of dyes to laser
excitation in the picosecond pulse range. Data on the excitation and
response pulses are given, and it is desired to find estimates of the
system function. The characteristic difficulties with this procedure
are discussed, and a method is described and illustrated that appears to
be optimal in the worst-case limit.

I.   INTRODUCTION. The paper is about methods for solving certain
Volterra Integral Equations of the first kind. The general form of
such equations is

$$\int_a^t K(t,s)\, x(s)\, ds = f(t), \qquad a \le t \le b. \tag{1}$$

It is assumed that the kernal $K(t,s)$ is known in the triangular region

$$a \le s \le t$$

$$a \le t \le b,$$

and the function $f(t)$ is known in $a \le t \le b$. We want to find the function
$x(t), a \le t \le b$.

The physical problem that concerns us is that of estimating the
response of various dyes to irradiation by pulses of laser light in the
pico-second range of pulse widths. A model which is commonly used in the
study of such systems can be written

$$\int_o^t h(t-s)E(s)ds = f(t), \qquad 0 \le t < \infty \tag{2}$$

where $E(s)$ is the excitation pulse of laser intensity, $f(t)$ is the
fluorescence pulse of the light from the dye in solution, and $h(t-s)$ is
the system function, which describes the effect of the dye on the
excitation pulse. It is assumed that the excitation begins at or after
t=0 and that the system function is causal, i.e.

$$E(U) = 0, \quad h(U) = 0, \quad U \le 0 \tag{3}$$

We want to find the system function h (U).

Under these conditions the integral equation can be written in the alternate forms

$$\int_{s=0}^{t} E(t-s)h(s)ds = f(t), \quad 0 \leq t < \infty \tag{4}$$

$$\int_{s=-\infty}^{\tau} E(t-s)h(s)ds = f(t), \quad 0 \geq t < \infty \tag{5}$$

and it is these forms that we shall study. The functions $E(t)$ and $f(t)$ are given at discrete, unevenly spaced points, and not, in general, at the same points, i.e. we know

$$f_i = f(t_i) \quad i=1,\ldots, M$$

$$E_j = E(U_j) \quad j=1,\ldots, N$$

and $M \geq N$. The values f_i and E_j are read from photographs of oscilloscope traces. Because of instrumental difficulties associated with these extremely short pulses, there is some fuzziness in the photographs and some uncertainty as to the baseline values.

We want to choose a practical method that enables us to find as much as we can about the function $h(t)$ from data of this type. Equation (4) is clearly a special case of (1), in which

$$a = 0, \, b = \infty, \, K(t,s) = E(t-s), \, \mathbf{X}(s) = h(s). \tag{6}$$

The main features of this special case, which set it apart from (1), are

(i) The kernel, E, is of convolution type.
(ii) The functions $E(U)$ and $f(t)$ are both pulse-like; in particular

$$K(o,o) = K(t,t) = E(t-t) = E(0) = 0 \tag{7}$$

(iii) Except for scale factors, the functions $E(U)$ and $f(U)$ are quite similar in shape, though perhaps uniformly shifted in time.

II. BACKGROUND. The books of Delves and Walsh, Reference [1], and Baker, [2], contain recent accounts of more-or-less practical methods for numerical solution of integral equations. Generally, it is much easier to solve equations of second kind, typically

$$\int_{a}^{b} K(t,s) \, x(s)ds + x(t) = f(t),$$

than those of the first kind,

$$\int_{a}^{b} K(t,s) \, x(s)ds = f(t),$$

154

whether the equations are of Volterra type (b=t) or Fredholm type (b fixed). In our case we see from (4) and (5) that there is no real distinction between equations of Fredholm and Volterra type, but both are affected by the ailments that are endemic to integral equations of the first kind.

These difficulties are well-described in [1]. If we regard (1) as an operator equation,

$$K \ x = f,$$

the difficulties boil down to the fact that the range of the operator K is too small and hence its nullspace is too large. That is, solutions exist only for certain functions. f, and more than one solution exists when f does have the required form. When numerical methods are used, this behavior usually manifests itself as near-rank-deficiency or lack of uniqueness in some system of algebraic equations.

The methods commonly given for solving these problems are as follows:

(i) Quadrature methods, i.e. replacing the integral by a finite sum and solving the resulting linear algebraic system either exactly or by least squares.

(ii) Integral Transform Methods, i.e. finding transforms of the functions, solving for the transform of $x(t)$ and inverting.

(iii) Parametric or Basis Function procedures, i.e. assuming a general form of $x(t)$, containing unspecified parameters, then solving for these parameters, e.g. by least squares.

Frequently when these methods are used on equations of the first kind, they do not work well. Various procedures, loosely described as regularization, have been advanced for avoiding these difficulties. The methods of Tykhonov and Phillips, Singular-Value Analysis and Cross-Validation are of this general type when applied to the Quadrature or Parametric schemes, and smoothing of the integral transform accomplishes something similar for procedures of class (ii).

For many problems, any of these methods may be employed. However, it is easy to see that all come to some kind of grief in our case. Principally, this is because we are forced to deal with the situation where $E(t)$ and $f(t)$ are nearly identical, apart from a shift and a multiplicative constant. It is obvious from (5) that in the limit when

$$f(t)=cE(t-b)$$

the meaningful solution of the integral equation is

$$h(t) = c\delta(t-b)$$

Thus when $f(t)$ and $E(t)$ are only slightly different, we must expect to find the h is a rapidly changing function of t. Any method (e.g. quadrature) that relies on smoothness will encounter substantial difficulties in this case.

If we attempt to take Fourier Transforms, we get from (5) and the convolution property

$$\Phi_E(w) \; \Phi_h(w) = \Phi_f(w),$$

where $\Phi_f(w)$ is the Fourier Transform of $f(t)$, etc. We may solve for Φ_h,

$$\Phi_h(w) = \Phi_f(w)/\Phi_E(w), \tag{6}$$

but this gets into trouble because $\Phi_E(w)$ will vanish at some points. Hence we can expect problems with this method as well.

Parametric Methods can probably be made to work, if we are skillful at guessing the basis functions, but are objectionable because we may be inadvertently constraining the form of the solution to be incorrect if our guesses are poor. Moreover, this method usually will involve a (possibly) non-linear, iterative least-squares solution to the problem of minimizing the errors at the data points t_i. The computational costs of this are unpleasant.

III. THE CUMULANT METHOD. To this writer, it appears that the essence of the problem with these methods is that the data $f(t_j)$ and $E(t_j)$ contain less and less information about $h(t)$ as the functions $E(t)$ and $f(t)$ become more nearly similar. In the limit where

$$f(t) = cE(t-b)$$
$$h(t) = c\delta(t-b),$$

the only information about $h(t)$ that the integral equations can possibly furnish is the two numbers c and b. On the other hand when $f(t)$ and $E(t)$ are sufficiently dissimilar, at least one of the standard methods will usually find stable estimates of the entire shape of $h(t)$.

This suggests that the most suitable method for the case where E and f are generally similar is one which concentrates on finding only a small number of resultants. The obvious choice is the low-order moments (or something equivalent). By focussing all the information in the data on the estimation of a few, low-order moments, we shall find these quantities with all the precision that the information can provide. If, on the other hand, we attempt to find the entire shape of the function, $h(t)$, we are diffusing the comparatively small amount of information across a large number of ordinates, none of which can then be found with much accuracy.

It is convenient to carry out these notions in terms of the cumulants, rather than the moments, of the functions E, f and h. These are derivable as follows: let

$$\psi_E(w) = \ln \phi_E(w) - \ln \phi_E(o)$$

$$= \text{cumulant generating function of E.}$$

and similarly define $\psi_f(w)$ and $\psi_h(w)$. Then from (8) and the fact that

$$\phi_f(o) = \phi_E(o) \phi_h(o)$$

we obtain

$$\psi_h(w) = \psi_f(w) - \psi_E(w) \tag{9}$$

The cumulants, $\psi_n(E)$, are defined in terms of the coefficients in the Taylor expansion of ψ_E about w=o, i.e.

$$\psi_E(w) = \sum_{n=1}^{\infty} \psi_n(E) \frac{(iw)^n}{n!}$$

or

$$\psi_n(E) = i^{-n} \left. \frac{d^n \psi_E(w)}{dw^n} \right|_{w=o}$$

The cumulant generating function is thus the logarithm of the Fourier Transform of a function having unit area, i.e. a probability density function. For such a function it is well-known that

$$\psi_1(E) = M_E = \text{mean of E}$$

$$\psi_2(E) = \sigma_E^2 = \text{variance of E}$$

$$\psi_3(E) = \gamma_1(E)\sigma_E^3 \tag{10}$$

$$\psi_4(E) = \gamma_2(E)\sigma_E^4$$

where $\gamma_1(E)$ and $\gamma_2(E)$ are the skewness and kurtosis of E, and similarly for f and h. Combining (9) and (10), we have

$$M_h = M_f - M_E$$

$$\sigma_h = (\sigma_f^2 - \sigma_E^2)^{1/2}$$

$$\gamma_1(h) = \{\gamma_1(f)\sigma_f^3 - \gamma_1(E)\sigma_E^3\}/\sigma_h^3 \tag{11}$$

$$\gamma_2(h) = \{\gamma_2(f)\sigma_f^4 - \gamma_2(E)\sigma_E^4\}/\sigma_h^4$$

Thus, we have only to find the resultants M, γ_1, σ and γ_2 for f and E and use the above formulas to find the analogous resultants for h.

157

The most satisfactory method for finding the resultants M, σ, γ_1, and γ_2 for f and E is to fit cubic splines to the data and integrate the splines. We define

$$\mu_k(E) = \int_{-\infty}^{\infty} s^k E(s)\,ds \qquad\qquad k=0,1,\ldots 4$$

and similarly for $f(t)$. For each k, a cublic spline is fitted to the integrand and the spline is integrated exactly to obtain μ_k. Then, e.g.

$$M_E = \mu_1(E)/\mu_0(E)$$

$$\sigma_E^2 = \frac{\mu_2(E)}{\mu_0(E)} - M_E^2$$

$$\gamma_1(E) = \left\{\frac{\mu_3(E)}{\mu_0(E)} - M_E(M_E^2 + 3\sigma_E^2)\right\}/\sigma_E^3$$

$$\gamma_2(E) = \left\{\frac{\mu_4(E)}{\mu_0(E)} - 4M_E\frac{\mu_3(E)}{\mu_0(E)} + 3M_E^2(M_E^2 + 2\sigma_E^2)\right\}/\sigma_E^4$$

(12)

and similarly for f.

This spline procedure has several desirable features.

(i) It can handle unevenly spaced data.
(ii) It makes only very modest assumptions about the shape of the fitted curve.
(iii) It is conveniently executed by available software.

The second of these features is very important. It means that the method does not impose any constraint on the solutions except the rather mild one of 2nd derivative continuity. This imparts a substantial advantage to this scheme, as compared with the parametric, or basis-function procedure, in which implicit, prior assumptions about the shape of the solution are unavoidable.

It is scarcely necessary to remark that considerable caution is necessary in using this spline method. Each of the quantities M_h, σ_h, $\gamma_1(h)$ and $\gamma_2(h)$ is found by subtracting other quantities that may differ only slightly, which means that accuracy will be a problem. As E and f become more similar, the quantities $\gamma_2(h)$ and $\gamma_1(h)$ will behave more erratically. Eventually, even σ_h^2 will become so small that we cannot conclude anything about it except that it is nearly zero. At that point, the function h is so narrow and sharply peaked that our method and data cannot distinguish it from a δ-function.

IV. SOFTWARE.

A small set of FORTRAN programs was written in order to carry out and test this method. The main program, MAIN, reads the data, contaminates it with Gaussian, uniform noise, calls the subroutine MOMTS and calculates

M_h, σ_h, $\gamma_1(h)$, $\gamma_2(h)$ from the moments of E and f. MOMTS calculates the moments of E and f from the data..

In addition to these, a program was written for test purposes which generates Gaussian pulses tinged with noise and makes two calculations.

(i) It finds M_h, σ_h, $\gamma_1(h)$ and $\gamma_2(h)$ exactly as MAIN does.

(ii) It calculates the ordinates of the function h(t) by the quadrature method, using the singular-value decomposition and discarding singular vectors of small singular values.

This program permits us to compare these two methods of finding h(t).

The spline method for estimating cumulants invokes the IMSL subroutines ICSSCU and DCSQDU for fitting cubic splines and integrating them. The first of these allows smoothing of the data, but it was found that smoothing had very little effect on the accuracy of moment estimates (as one would expect).

The quadrature method (see [1]) replaces the integral equation

$$f_i = \int_0^{t_i} E(t_i - s)h(s)\,ds \qquad\qquad i = 1, \ldots, N$$

by the trapezoidal formula

$$f_i = \sum_{k=0}^{i} \Delta W_k E_{i-k} h_k = \sum_{k=0}^{i} A_{ik} h_k$$

where Δ is the mesh spacing in the integration

$$E_{i-k} = E(i\Delta - k\Delta)$$

$$W_k = 1, \quad \text{for} \quad 1 \leq k \leq i-1$$

$$ = 1/2, \quad \text{for} \quad k = 0, \quad k = i$$

$$h_k = h(k\Delta).$$

$$A_{ik} = \Delta W_k E_{i-k}.$$

This leads us to a linear algebraic system with NxN matrix $A = [A_{ik}]$.

A is a lower triangular matrix that is nearly of Toeplitz form. The matrix equation

$$Ah = f$$

is solved using rank-selection based on the singular-value decomposition. The programs MINFIT and MINSOL from the IMSL ROSEPACK collection are used for this purpose.

V. EXAMPLES.

We present two examples in which the cumulant and quadrature methods are compared on two different problems, loosely speaking an easy problem and a hard one. In all cases the exact pulses E, f and h are Gaussian with

$$E = e^{-1/2(\frac{x-M_e}{\sigma_E})^2} \quad , \qquad f = e^{-1/2 (\frac{x-M_f}{\sigma_f})^2}$$

$$h = C_h\, e^{-1/2(\frac{x-M_h}{\sigma_h})^2}$$

with $M_h = M_f - M_e$, $\sigma_h = (\sigma_f^2 - \sigma_E^2)^{1/2}$, $C_h = \dfrac{\sigma_f}{\sqrt{2\pi}\,\sigma_E \sigma_h}$.

The data points are at $t_i = i, i = 0, \cdots, 24$ and $\sigma_{noise} = .0005$.

$$\text{Easy example:} \quad M_E = 8.0, \quad \sigma_E = 1.6$$
$$M_f = 14.0, \quad \sigma_f = 2.2$$

and the exact solution has

$$M_h = 6.0, \quad \tau_h = 1.5100, \quad \gamma_1(h) = \gamma_2(h) = 0$$

Figure 1 shows the exact forms of the pulses E, f and h, together with the points obtained from the quadrature method with rank-selection based on the singular values. Five different trials, using different random number seeds for noise generation, gave the results shown in Table 1. The results for the five trials are not distinguishable on the scale of Figure 1.

These results show that both methods were satisfactory. The quadrature method with rank selection based on singular values gave accurate and stable estimates of the function values. The cumulants up to $\gamma_1(h)$ are also found with reasonable precision, but $\gamma_2(h)$ is unstable. Probably we would regard the quadrature method as the better one because it provides somewhat more complete information about h. The matrix A has N=25 and rank that ranges from 11 to 14, so A is very rank-deficient even in this easy case. The rank was found so that the solution for h agreed better with the exact solution (in the L_2 sense) than for any other rank. Naturally, this method cannot be used when the exact solution is unknown, as will usually be so in practice. The results show that, while the rank may vary substantially, the solution values are pretty stable.

Figure 2 and Table 2 show like results for the hard example:

$$M_E = 8.0 \qquad \sigma_E = 1.6$$

$$M_f = 13.0 \qquad \sigma_f = 1.7$$

which has as exact solution the narrower, sharper pulse with

$$M_h = 5.0, \qquad \tau_h = .57446, \qquad \gamma_1(h) = \gamma_2(h) = 0$$

Both methods had a difficult time with this example. However, the quadrature method gave almost no accurate information beyond the fact that the h-pulse was located near t=5. The pulse is depicted consistently as somewhat lower and less sharp than it really is, undoubtedly because of the severe smoothing that has been done in the rank-selection process. This consistency is unfortunate because it implies that, in a situation where we did not know the true solution, inconsistency in the estimates might not occur to warn us of impending trouble.

Although the cumulant method also did poorly, its results were better than the quadrature method on two counts. First, it provides good estimates of the area and location of the h-pulse, and acceptable estimates of σ_h as well. Second, although the estimates of skewness and kurtosis are bad, their inconsistency is a clear warning not to trust them. It appears therefore that the cumulant method gives us more useful information than the quadrature method in this case.

VI. DISCUSSION AND CONCLUSIONS. The examples support the intuitive notion that, when $E(t)$ and $f(t)$ are similar, the best procedure is to estimate a few, low-order cumulants of $h(t)$. The calculations using splines are simpler than those involved in the other procedures and focus on the only quantities that can be predicted with any accuracy and stability in this worst-case limit. The method does not require evenly-spaced data and makes the mildest possible hypotheses about the solution.

Even if the functions $E(t)$ and $f(t)$ are not much alike, it may be worthwhile to use this cumulant procedure as a preliminary or adjunct to a more complete analysis. In particular, if the basis function method is employed, it may be very helpful to have at hand the information that the cumulants provide about the general shape of the function.

Obviously further work is needed to clarify both theoretical and practical aspects of these methods.

REFERENCES

[1]. L.M. Delves and J. Walsh et al, Numerical Solution of Integral Equations, Oxford, Clarendon Press, 1974.

[2]. C. T. H. Baker, The Numerical Treatment of Integral Equations, Oxford University Press, 1977.

[3]. A. N. Tihonov, Solution of Incorrectly Formulated Problems and the Method of Regularization, Soviet Math. Doklady, vol. 4, 1963, pp. 1035-1038.

[4]. D. L. Phillips, A Technique for the Numerical Solution of Certain Integral Equations of the First Kind, J. Assoc. Comp. Macn., vol. 9, 1962, pp. 54-97.

[5]. Grace Wahba, Practical Approximate Solutions to Linear Operator Equations when the Data are Noisy, Tech. Rept #430, U. of Wisconsin, Dept. of Statistics, Sept. 1975.

TABLE 1: Easy Example Results
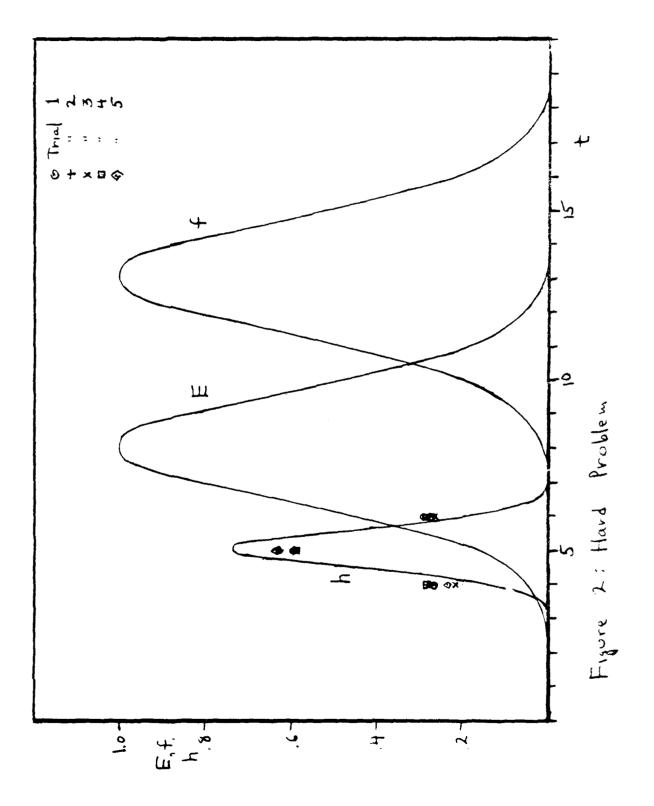
Trials

	Exact	1	2	3	4	5	
$\mu_o(h)$	1.375	1.375	1.374	1.374	1.375	1.374	Cumulant-Spline Method
M_h	6.0	6.006	6.005	6.005	5.998	6.004	
σ_h	1.5100	1.519	1.520	1.513	1.477	1.505	
$\gamma_1(h)$	0	.1133	.0824	.2560	-.3178	.1726	
$\gamma_2(h)$	0	-.2529	1.029	1.751	-4.319	.1269	

rank	-	14	12	13	12	11	
$\|err\|_2$	-	.00151	.00196	.00177	.00229	.00185	
$\|h\|_2$	-	.59451	.59438	.5935	.5941	.5941	
$h(3)$	.0505	.0487	.0537	.0473	.0485	.0500	
$h(4)$	.1511	.1450	.1492	.1527	.1547	.1516	
$h(5)$	.2917	.2973	.2882	.2946	.2927	.2922	Quadrature Method
$h(6)$	.3633	.3628	.3642	.3605	.3585	.3624	
$h(7)$	.2917	.2884	.2957	.2887	.2933	.2923	
$h(8)$	.1511	.1540	.1495	.1533	.1551	.1502	
$h(9)$	.0505	.0489	.0473	.0533	.0465	.0492	

		Trials				
	Exact	1	2	3	4	5
$\mu_0(h)$	1.0625	1.063	1.064	1.062	1.063	1.062
M_h	5.000	5.003	5.004	4.992	5.005	5.003
σ_h	.57446	.5529	.6222	.5804	.5354	.6037
$\gamma_1(h)$	0	.4039	4.543	-3.468	-.2004	2.784
$\gamma_2(h)$	0	-62.77	71.43	-6.227	-82.12	44.69

Cumulant Spline Method

rank	-	15	14	17	13	15
$\|err\|_2$	-	.00093	.00163	.00140	.00140	.00233
$\|h\|_2$	-	.7206	.7202	.7246	.7153	.7305
$h(3)$	.0017	-.0558	-.0726	.0081	-.0617	-.0300
$h(4)$	.1622	.2642	.2790	.2166	.2805	.2349
$h(5)$	.7379	.5947	.5939	.6293	.5844	.6275
$h(6)$	.1622	.2872	.2762	.2610	.2825	.2667
$h(7)$	.0017	-.0646	-.0518	-.0469	-.0537	-.0787

Quadrature Method

164

Figure 1: Easy Problem

165

Figure 2: Hard Problem

166

Cubic Splines and Approximate Solution
of Singular Integral Equations

Erica Jen and R.P. Srivastav

Department of Applied Mathematics & Statistics

State University of New York at Stony Brook

Stony Brook, New York 11794

ABSTRACT: Of concern here is the numerical solution of singular
integral equations of Cauchy type, i.e., equations involving
principal value integrals. The unknown function is expressed
as the product of an appropriate weight function and a cubic
spline. The problem is reduced to a system of linear algebraic
equations, which determine the approximate values of the function
at each of the knots. It turns out that the maximum error can be
estimated. The procedure has been tested on a variety of problems
arising in mixed boundary value problems of elasticity. The
prospects for refinement and extension are also discussed.

Cubic Splines and Approximate Solution
of Singular Integral Equations

Erica Jen and R. P. Srivastav

1. <u>Introduction</u>: Many boundary value problems in physics and engineering can be reduced to the problem of solving a singular integral equation of the form:

$$(1.1) \quad a(s)g(s) + \frac{b(s)}{\pi} \int_{-1}^{1} \frac{g(t)dt}{t-s} + \int_{-1}^{1} K(t,s)g(t)dt = f(s), \quad -1 < s < 1.$$

The singular integral is to be taken in the sense of the Cauchy principal value. The functions a, b, K, and f are known; and g is the unknown function. Equations of this type have been studied extensively in the classical theories of elasticity and hydrodynamics. These equations also arise in the mathematical treatment of such diverse fields as radiative transfer, neutron transport, and particle physics. In such contexts, the function g usually represents either a potential (e.g., temperature, displacement, velocity potential) or a flux-type quantity (e.g. heat flux, stress, charge density).

The theoretical foundations for the study of singular integral equations were laid by Hilbert's work in analytic function theory and Poincaré's investigation of the general theory of tides. Important contributions were subsequently made by Noether in her work on the so-called "index theorems," and by Carleman in his derivation of an explicit solution for the basic equation. The classic theorems of

Fredholm served as a model for the development of the analytical

approach now known as the Carleman-Vekua regularization of systems

of singular integral equations. The theory of singular integral

equations of one variable is fairly well-understood, due to these

early results and to the contributions of Muskhelishvili [1], Gakhov [2],

and Privalov [3].

The attention to direct methods of solution is of comparatively

recent origin. In 1969, F. Erdogan [5] proposed a technique which

explicitly builds the "correct" singular behavior of the solution in

the approximating sequence of functions. The index theory provides the

weight function, and the Jacobi polynomials orthogonal with respect to

this weight are used to represent the approximate solution by the

relation

(1.2) $g(t) = \phi(t)w(t)$,

where $w(t)$ is the weight and $\phi(t)$ is a linear combination of the Jacobi

polynomials of degree $\leq N$. Subsequently, Erdogan and Gupta [6] developed

a Gauss-Chebyshev type formula for numerical evaluation of singular

integrals. An excellent exposition of the Erdogan-Gupta procedure is

given in [7]. Theocaris and Ioakimidis [8] have proposed a variant of

this method, based on the Lobatto-Chebyshev formula, which eliminates

the need for the extrapolation to determine the value of the function

at the endpoints of the interval. Both methods rely on a discretization

of the equation which involves the values of the function at the zeros

of certain polynomials. Dow and Elliot [9] also suggest the use of the

orthogonal polynomials to solve the singular integral equation. A proof

of the convergence of their algorithm is given, under the assumption that the approximating sequence of polynomials $\{f_n\}$ converges to f in the Hölder norm. It is to be noted that $\{f_n\}$ is a sequence of interpolatory polynomials, and its convergence in the Hölder norm is not entirely obvious.

In the evaluation of non-singular integrals whose integrands are of only low-order differentiability, the Gaussian integration formulae are nearly as accurate as the trapezoidal or Simpson's rule ([16], p. 187). In the evaluation using a product integration rule of singular integrals, however, the accuracy of these formulae has not been established. Moreover, the methods based on Gaussian integration formulae suffer from the previously mentioned requirement that the collocation points coincide with the zeros of certain polynomials. Hence, if either the right hand side of (1.1) or the kernel K fluctuates over a small interval, the methods are effective only if a large number of points is used. Thus there appears to be a need for further development of low-cost, low-accuracy methods.

One possible approach due to Gerasoulis and Srivastav [11] uses piecewise linear functions to determine the function $x(t)$ in (1.2). This procedure permits the analytical evaluation of the integral expressions, and yields satisfactory results for certain test problems with known solutions. Gerasoulis [12] obtained an improvement in accuracy by using quadratic interpolation. These earlier results motivated the development of a cubic spline approximation method. In addition to providing higher accuracy, the spline method can also be expected to be applicable to the numerical solution of singular

integro-differential equations.

Spline methods have been used successfully for the solution of non-singular integral equations (see Ahlberg, Nilson, and Walsh [13], Netravali and Figueiredo [14]). In fact, Gabdulhaev [15] found that under certain conditions, the spline method with equally spaced knots is optimal among all collocation methods for the numerical solution of Fredholm integral equations of the second kind.

The organization of the rest of the paper is as follows: Section 2 describes a procedure for the solution of (1.1) with $a(s) = 0$ and $b(s) = 1$; Section 3 is devoted to error analysis; and section 4 contains comparisons of the numerical results obtained from spline approximation and other methods.

2. Reduction to a Linear Algebraic System: The general strategy of direct methods for the solution of integral equations is to discretize the original equation by considering it only at a finite set of points in the domain, and to use some numerical integration formula to obtain a system of algebraic equations for the values assumed by the unknown function at these points. The accuracy of the solution is affected by both the choice of collocation points and the quadrature formula used.

Consider the case where $a(s) = 0$ and $b(s) = 1$ in (1.1), and the solution is known to possess square root singularities at ± 1. (The method described below is applicable in general, although in some cases it may be necessary to evaluate certain integral expressions numerically.) Set

$$(2.1) \qquad g(t) = c(t)(1-t^2)^{-\frac{1}{2}}.$$

171

In most applications, it is possible to exploit the symmetry properties of the problem, and to work with either odd or even functions. Therefore, assume the number of node points to be $(2n+1)$, and let $-1 = t_0 < t_1 < \ldots < t_{2n} = 1$. Replace the unknown function $\phi(t)$ by splines $S(t) = S_j(t)$, $(j = 1, 2, \ldots, 2n)$ on the interval $[t_{j-1}, t_j]$. It is computationally convenient to use the form [13]

$$(2.2) \quad S_j(t) = \frac{M_{j-1}}{6h_j}(t_j - t)^3 + \frac{M_j}{6h_j}(t - t_{j-1})^3 + \left(\frac{\phi_j}{h_j} - \frac{M_j h_j}{6}\right)(t - t_{j-1})$$

$$\left(\frac{\phi_{j-1}}{h_j} - \frac{M_{j-1} h_j}{6}\right)(t_j - t), \qquad j = 1, 2, \ldots, 2n$$

where $h_j = t_j - t_{j-1}$, $\phi_j = \phi(t_j)$, and $M_j = S_j''(t_j) = S_{j-1}''(t_j)$. (Although the above expressions for $S_j(t)$ involve the moments, or second derivatives, of splines, it is possible to use instead their first derivatives.)
The function $K(t, s_k)$ is approximated by $K_j(t, s_k)$ using a cubic interpolation formula in each of the intervals $[t_{j-1}, t_j]$. In this way, the original equation is replaced by a discrete analogue

$$(2.3) \quad \sum_{j=1}^{2n} \int_{t_{j-1}}^{t_j} \frac{S_j(t)dt}{\sqrt{1-t^2}(t-s_k)} + \sum_{j=1}^{2n} \int_{t_{j-1}}^{t_j} \frac{S_j(t)K_j(t,s_k)dt}{\sqrt{1-t^2}} = f(s_k),$$

$$k = 1, 2, \ldots, 2n$$

where the collocation points are chosen so that $t_{k-1} < s_k < t_k$. All the quantities in (2.4) can be evaluated analytically to yield $2n$ linear equations for the $(4n+2)$ unknowns $M_0, M_1, \ldots, M_{2n}, \phi_0, \phi_1, \ldots, \phi_{2n}$. An additional $(2n-1)$ equations are furnished by the continuity of the derivatives of splines; namely,

$$(2.4) \quad M_{j-1} + 2M_j\left(\frac{h_j + h_{j+1}}{h_j}\right) + \frac{h_{j+1}}{h_j}M_{j+1} = \frac{6}{h_j}\left(\frac{\phi_{j+1} - \phi_j}{h_{j+1}} - \frac{\phi_j - \phi_{j-1}}{h_j}\right),$$

$$j = 1, 2, \ldots, 2n-1.$$

172

Two equations relating the values of the moments at the endpoints are needed. These equations are usually chosen to be of the form

$$(2.5) \qquad \alpha_0 M_0 + \beta_0 M_1 = C_0, \qquad \beta_{2n} M_{2n-1} + \alpha_{2n} M_{2n} = C_{2n}.$$

Finally, a single equation is obtained from the compatibility condition

$$(2.6) \qquad \sum_{j=1}^{2n} \int_{t_{j-1}}^{t_j} \frac{S_j(t)\,dt}{\sqrt{1-t^2}} = k, \qquad k \text{ constant.}$$

Thus a total of $(4n+2)$ equations in as many variables is obtained. The coefficient matrix for the system of equations is of the form

$$\begin{pmatrix} A^1 \\ A^2 \\ A^3 \end{pmatrix}$$

where A^1 is the $2n \times (4n+2)$ submatrix of coefficients obtained from the integral equation evaluated at the $2n$ collocation points;

A^2 is the $1 \times (4n+2)$ submatrix of coefficients obtained from the compatibility condition;

and A^3 is the $(2n+1) \times (4n+2)$ submatrix of coefficients obtained from the moments conditions and the continuity relations for splines.

In order to display the elements of the coefficient matrix in convenient form, some operator notation is needed. Define operators I_k, J_k by

$$(I_k f)(s) = \int_{t_{k-1}}^{t_k} \frac{f(t)\,dt}{\sqrt{1-t^2}\,(t-s)}, \qquad J_k f = \int_{t_{k-1}}^{t_k} \frac{f(t)\,dt}{\sqrt{1-t^2}}.$$

Note that for polynomial functions f, the expressions for $(I_k f)(s)$ and $J_k f$ can be expressed analytically. In particular,

173

$$(I_k t^p)(s) = \int_{t_{k-1}}^{t_k} \frac{t^{p-1}}{\sqrt{1-t^2}}\,dt + s\int_{t_{k-1}}^{t_k} \frac{t^{p-2}}{\sqrt{1-t^2}}\,dt + \ldots + s^{p-1}\int_{t_{k-1}}^{t_k} \frac{dt}{\sqrt{1-t^2}}$$

$$+ s^p(I_k 1)(s), \qquad\qquad p = 1,2,\ldots$$

and ([16], p. 147)

$$(I_k 1)(s) = \frac{1}{\sqrt{1-s^2}} \ell n \left| \frac{\left(-1+\sqrt{1-s^2}+s\tan\frac{\theta_k}{2}\right)\tan\frac{\theta_{k-1}}{2}+s-(\sqrt{1-s^2}+1)\tan\frac{\theta_k}{2}}{\left(-1-\sqrt{1-s^2}+s\tan\frac{\theta_k}{2}\right)\tan\frac{\theta_{k-1}}{2}+s-(-1-\sqrt{1-s^2})\tan\frac{\theta_k}{2}} \right|$$

where $\theta_k = \arcsin t_k$.

Then the elements of the submatrix A^1 are given for $j = 1,2,\ldots,2n$ by

$$A^1_{j,1} = \left\{ I_i\left[\frac{1}{6h_i}(t_i-t)^3 - \frac{h_i}{6}(t_i-t)\right]\right\}_j (s_j)$$

$$+ J_i\left[\frac{1}{6h_i}(t_i-t)^3 K_{i,j}(t) - \frac{h_i}{6}(t_i-t)K_{i,j}(t)\right]$$

$$+ \left\{ I_{i-1}\left[\frac{1}{6h_{i-1}}(t-t_{i-2})^3 - \frac{h_{i-1}}{6}(t-t_{i-2})\right]\right\}_j (s_j)$$

$$+ J_{i-1}\left[\frac{1}{6h_{i-1}}(t-t_{i-2})^3 K_{i-1,j}(t) - \frac{h_{i-1}}{6}(t-t_{i-2})K_{i-1,j}(t)\right]$$

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

$$= \left\{ I_i\left[\frac{1}{h_i}(t_i-t)\right]\right\}_j (s_j) + J_i\left[(t_i-t)K_{i,j}(t)\right]$$

$$+ \left\{ I_{i-1}\left[\frac{1}{h_{i-1}}(t-t_{i-2})\right]\right\}_j (s_j) + J_{i-1}\left[(t-t_{i-2})K_{i-1,j}(t)\right]$$

$$i = 2n+2,\ 2n+3,\ldots,4n+2$$

where $(I_k f)(s)$ and $J_k f$ are taken to be zero for $k = 0,\ 2n+1$.

The elements of A^2 are given by

174

$$A^2_{1,i} = J_i\left[\frac{1}{6h_i}(t_i-t)^3 - \frac{h_i}{6}(t_i-t)\right] + J_{i-1}\left[\frac{1}{6h_{i-1}}(t-t_{i-2})^3 - \frac{h_{i-1}}{6}(t-t_{i-2})\right],$$

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

$$= J_i\left[\frac{1}{h_i}(t_i-t)\right] + J_{i-1}\left[\frac{1}{h_{i-1}}(t-t_{i-2})\right], \qquad i = 2n+2,\ 2n+3,\ldots,4n+2$$

where again $J_k f = 0$ for $k = 0, 2n+1$.

The elements of A^3 are given by

$$
\begin{aligned}
A_{1,i} &= \alpha_0 & i &= 1 \\
&= \beta_0 & i &= 2 \\
&= 0 & &\text{otherwise}
\end{aligned}
$$

$$
\begin{aligned}
A_{2n+1,i} &= \beta_{2n} & i &= 2n \\
&= \alpha_{2n} & i &= 2n+1 \\
&= 0 & &\text{otherwise}
\end{aligned}
$$

and for $j = 2,3,\ldots,2n$ by

$$
\begin{aligned}
A_{j,1} &= 1 & i &= j-1 \\
&= 2\left(\frac{h_j + h_{j-1}}{h_{j-1}}\right) & i &= j \\
&= \frac{h_j}{h_{j-1}} & i &= j+1 \\
&= \frac{6}{h_j h_{j-1}} & i &= 2n+j \\
&= -\frac{6}{h_j h_{j-1}} - \frac{6}{h_j^2} & i &= 2n+j+1 \\
&= \frac{6}{h_j^2} & i &= 2n+j+2 \\
&= 0 & &\text{otherwise}
\end{aligned}
$$
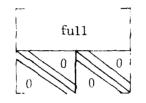
From the above, it can be seen that the coefficient matrix A has the structure

Note that when solving the system by Gaussian elimination using only partial pivoting, it may be advisable to rearrange the matrix so that the τ_j's are computed first, thus reducing the effect of round-off error propagation.

3. <u>Error Analysis</u>: Define the functions ϕ^*, ϕ^ϵ as follows:

(i) ϕ^* is the Type II cubic spline on the true values

   $\phi(t_i)$, $i = 0,1,\ldots,2n$, with $\phi^{*''}(t_0) = \phi''(t_0)$, $\phi^{*''}(t_{2n}) = \phi''(t_{2n})$;

(ii) ϕ^ϵ is the spline on the computed values for $\phi(t_i)$, $i = 0,1,\ldots,2n$.

Let

$$\underline{x} = [\phi''(t_0),\phi''(t_1),\ldots,\phi''(t_{2n}),\phi(t_0),\phi(t_1),\ldots,\phi(t_{2n})]$$
$$\underline{x}^* = [\phi^{*''}(t_0),\phi^{*''}(t_1),\ldots,\phi^{*''}(t_{2n}),\phi^*(t_0),\phi^*(t_1),\ldots,\phi^*(t_{2n})]$$
$$\underline{x}^\epsilon = [\phi^{\epsilon''}(t_0),\phi^{\epsilon''}(t_1),\ldots,\phi^{\epsilon''}(t_{2n}),\phi^\epsilon(t_0),\phi^\epsilon(t_1),\ldots,\phi^\epsilon(t_{2n})]$$

and

$$\underline{f} = [f(s_1),f(s_2),\ldots,f(s_{2n}),k,C_0,0,\ldots,0,C_{2n}].$$

It will be assumed below that the splines used are the natural splines, so $C_0 = C_{2n} = 0$. The system of equations which is being solved can therefore be represented as

176

(3.1) $A\underline{x}^\epsilon = f.$

The vector $\underline{x}$ of true values satisfies

(3.2) $A\underline{x} = \tilde{f}$

where $\tilde{f} = f+\theta$, and θ is the vector of errors in the numerical integration due to the use of splines. Hence

(3.3) $\|x^\epsilon - x\| \le \|A^{-1}\| \cdot \|\theta\|$.

The above inequality can be used to obtain an error estimate for the spline method. The j-th component of θ is given by

(3.4) $\theta_j = \dfrac{1}{\pi}\displaystyle\int_{-1}^{1} \dfrac{(\phi^* - \phi)\,dt}{(t-s_j)\sqrt{1-t^2}} + \displaystyle\int_{-1}^{1} \dfrac{(K_\epsilon^* - K_j)\,dt}{\sqrt{1-t^2}}$, $j = 1,2,\ldots,2n$

where K_ϵ is the piecewise cubic approximant to K,

(3.5) $\theta_{2n+1} = \displaystyle\int_{-1}^{1} \dfrac{(\phi^* - \phi)}{\sqrt{1-t^2}}\,dt.$

$\theta_{2n+2} = \phi''(t_0)$

$\theta_{4n+2} = \phi''(t_{2n})$

and $\theta_j = 0$ for other values of j. Let

(3.6) $\epsilon(t) = \phi^* - \phi$, $\epsilon'(t) = \phi^{*'} - \phi'.$

Then it has been shown (see [17], p. 107) that

(3.7) $|\epsilon| \le \dfrac{5}{384}\|\phi^{(iv)}\|h^4,$ $|\epsilon'| \le \dfrac{1}{24}\|\phi^{(iv)}\|h^3$

where $h = \max_j h_j$.

The second term in (3.4) is easily shown to be less than or equal to

$\pi\{\max_{-1\le t \le 1} |\phi^*| \cdot \max|K_\epsilon - K| + \max|\epsilon| \cdot \max|K|\}$

which is $O(h^4)$.

177

Consider now the first term in (3.4). Suppose the mesh is uniform; i.e. $h_j = h$ for all j. The results below are not significantly affected if this assumption does not hold. Furthermore, assume that the collocation points s_j are chosen to be the midpoints of the intervals (t_{j-1}, t_j). For $s_1 \in (t_0, t_1)$,

$$\left| \int_{-1}^{1} \frac{(\phi^* - \phi)dt}{(t-s_1)\sqrt{1-t}} \right| \leq \left| \int_{t_0}^{t_1} \frac{[\varepsilon(t) - \varepsilon(s_1)]dt}{(t-s_1)\sqrt{1-t^2}} \right| + |\varepsilon(s_1)| \cdot \left| \int_{t_0}^{t_1} \frac{dt}{(t-s_1)\sqrt{1-t^2}} \right|$$

$$+ \sum_{k=2}^{2n} \left| \int_{t_{k-1}}^{t_k} \frac{\varepsilon(t)dt}{(t-s_1)\sqrt{1-t^2}} \right|$$

(3.8)
$$\leq \|\varepsilon'\| \cdot \frac{\sqrt{h}}{\sqrt{2-h}} + \|\varepsilon\| \cdot M_1 + |\varepsilon_1| \cdot M_2 \cdot \frac{2}{h} \sum_{k=1}^{2n-1} \frac{1}{2k-1}$$

where

$$M_1 = \left| \int_{t_0}^{t_1} \frac{dt}{(t-s_1)\sqrt{1-t^2}} \right| = \frac{2}{\sqrt{h}\sqrt{4-h}} \ell n \left| \frac{1 + \sqrt{\frac{2-h}{4-h}}}{1 - \sqrt{\frac{2-h}{4-h}}} \right|$$

$$= O(h^{-\frac{1}{2}})$$

$$M_2 = \max_{\substack{k \\ k \neq 1}} \left| \int_{t_{k-1}}^{t_k} \frac{dt}{\sqrt{1-t^2}} \right| = \left| \int_{t_1}^{t_2} \frac{dt}{\sqrt{1-t^2}} \right| = \cos^{-1}(-1+h)$$

$$= O(h^{\frac{1}{2}})$$

and

$$\sum_{k=1}^{2n-1} \frac{1}{2k-1} = O(h^{\delta}), \quad \delta > 0.$$

Therefore

$$\left| \int_{-1}^{1} \frac{(\phi^* - \phi)dt}{(t-s_j)\sqrt{1-t}} \right| = O(h^{\frac{1}{2}-\delta}), \quad \delta > 0.$$

This estimate is not sharp, but appears adequate. For $s_{2n} \in (t_{2n-1}, t_{2n})$ the same estimate holds. Next consider $s_j \in (t_{j-1}, t_j)$ with $j \neq 1, 2n$. In this case,

$$\left| \int_{-1}^{1} \frac{(\phi^* - \phi)dt}{(t-s_j)\sqrt{1-t}} \right|$$

$$\leq \left| \int_{t_{j-1}}^{t_k} \frac{[\varepsilon(t) - \varepsilon(s_j)]dt}{(t-s_j)\sqrt{1-t^2}} \right| + \varepsilon(s_j) \cdot \left| \int_{t_{j-1}}^{t_j} \frac{dt}{(t-s_j)\sqrt{1-t^2}} \right| + \sum_{\substack{k=1 \\ k \neq j}}^{2n} \left| \int_{t_{k-1}}^{t_k} \frac{\varepsilon(t)dt}{(t-s_j)\sqrt{1-t^2}} \right|$$

178

$$(3.9) \quad \leq \|\varepsilon'\| \frac{\sqrt{h}}{\sqrt{2-h}} + \|\ldots\| \cdot M_3 + \|\varepsilon\| \cdot M_4 \cdot \frac{4}{h} \sum_{k=1}^{n} \frac{1}{2k-1}$$

where

$$M_3 = \left| \int_{t_{j-1}}^{t_j} \frac{dt}{(t-s_j)\sqrt{1-t^2}} \right| = \left| \int_{t_{j-1}}^{t_j} \frac{[(1-t^2)^{-\frac{1}{2}}-(1-s_j^2)^{-\frac{1}{2}}]dt}{t-s_j} + \int_{t_{j-1}}^{t_j} \frac{(1-s_j^2)^{-\frac{1}{2}}dt}{t-s_j} \right|$$

$$= \int_{t_{j-1}}^{t_j} \frac{\xi(t)dt}{[1-\xi(t)^2]^{3/2}}, \qquad t_{j-1} < \xi(t) < t$$

$$= \frac{\xi}{(1-\xi^2)^{3/2}} \cdot h \leq \frac{h(1-h)}{[1-(1-h)^2]^{3/2}}$$

$$= O(h^{-1/2})$$

and

$$M_4 = \max_{\substack{k \\ k \neq j}} \int_{t_{k-1}}^{t_k} \frac{dt}{\sqrt{1-t^2}} = O(h^{1/2})$$

Therefore

$$\int_{-1}^{1} \frac{(\phi^*-\phi)dt}{(t-s_j)\sqrt{1-t^2}} = O(h^{\frac{1}{2}-\delta}), \qquad \delta > 0.$$

Note that the estimate (3.3) should be used with caution since the quantity $\|A^{-1}\|$ depends on h. The appendix lists values of $\|A^{-1}\|$ and $\|A^{-1}\cdot\ldots\|$ to be used in estimating the error involved in solving a particular equation. In practice, it has been found that for problems with known solutions, the spline method produces results of accuracy considerably better than would be expected from (3.3), thus indicating that the error bounds above could probably be significantly improved.

4. <u>Numerical Results</u>: The spline method has been used to solve a number of singular integral equations, including the following:

Example 1.

$$(4.1) \qquad \frac{1}{\pi}\int_{-1}^{1} \frac{g(t)}{t-s}dt + \frac{1}{\pi}\int_{-1}^{1} \sin(t-s)g(t)dt = J_1(1)\cos s + 1, \qquad -1 < s < 1$$

where J_1 is the Bessel function of the first kind of order 1. The solution $g(t)$ is required to satisfy the compatibility condition

$$\int_{-1}^{1} g(t)dt = 0.$$

Moreover, $g(t)$ is assumed to possess square-root singularities at ± 1, and hence to be expressible in the form

$$(4.2) \qquad g(t) = \phi(t)(1-t^2)^{-1/2}.$$

Then it can be seen that the true solution is given by (4.2) with $\phi(t) = t$. The table below displays the maximum error ϵ in the values for $\phi(t)$ computed using the spline method.

n^*	ϵ
2	0.000041
3	0.000003
4	0.000001

*n is the number of nodes taken in the interval $[0,1]$. For $n = 7$, the computed solution was accurate to the limits of single-precision computation (8 digits).

Example 2.

$$\frac{1}{\pi}\int_{-1}^{1} \frac{g(t)}{t-s}dt - \lambda \int_{-1}^{s} g(t)dt = p_0, \qquad\qquad -1 < s < 1$$

subject to the compatibility condition

$$\int_{-1}^{1} g(t)dt = 0.$$

The above equation arises in the plane elasticity problem for a plate bonded to an elastic half-plane. The solution $g(t)$ is again assumed to be of the form

$$g(t) = \phi(t)(1-t^2)^{-1/2}.$$

The table below displays some values of $\phi(t)$ obtained by the spline method for $\lambda = 1/5$, $p_0 = 1$ compared with those given by Erdogan-Gupta [6]. In this case, natural splines were used; i.e. it was assumed that

$$\phi''(t_0) = \phi''(t_{2n}) = 0.$$

180

t	Erdogan-Gupta (n=40)	Spline (n=10)
0.11753	0.08016	0.0017
0.27144	0.18664	0.18665
0.41866	0.29205	0.29207
0.55557	0.39543	0.39544
0.67880	0.49522	0.49522
0.78531	0.58935	0.58954
0.87249	0.67482	0.67417
0.93819	0.74762	0.74052
0.98078	0.80246	0.80504
0.99923	0.83172	0.82753
1.00000	0.83531¹	0.82862²

¹obtained through extrapolation

²obtained directly

(It should be noted that the spline method used equally spaced nodes. For purposes of comparison, the values of $g(t)$ at the Erdogan-Gupta node points were obtained using spline interpolation. It is to be expected that the spline method would be more accurate at its own nodes than at intermediate points.)

As can be seen from the table above, the results for the two methods are in close agreement for $t < 0.78$. The discrepancy near the endpoint $+1$ would seem to indicate that natural splines were an inappropriate choice. When "not-a-knot" conditions [18] were used instead, the value obtained for $g(1)$ was 0.8335.

Example 3.

$$\frac{1}{\pi}\int_{-1}^{1}\frac{g(t)}{t-s}dt + \frac{1}{\pi}\int_{-1}^{1}\frac{t(t^2-s^2)}{(t^2+s^2)^2}g(t)dt = s, \qquad -1 < s < 1$$

subject to the compatibility condition

$$\int_{-1}^{1}g(t)dt = 0.$$

181
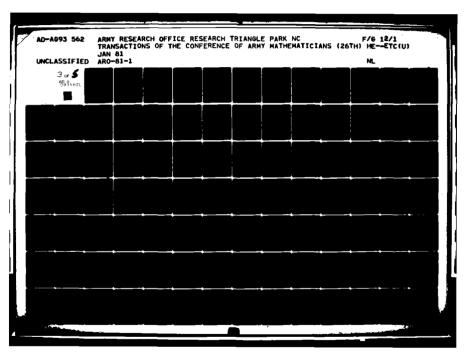
The above equation arises in the problem of a cruciform crack in an infinite isotropic elastic medium under constant load along its four branches. As before, the function g(t) is assumed to be of the form

$$g(t) = \phi(t)(1-t^2)^{-1/2}.$$

The table below provides a comparison of results obtained from the Erdogan-Gupta method [8] (Col. I); its Lobatto-Chebyshev variant [8] (Col. II); and the spline method (Col. III). Note the correct value for $\phi(1)$ has been calculated by Rooke and Sneddon [19] to be 0.8656.

n	Erdogan-Gupta	Lobatto-Chebyshev	Spline
3	0.8564	0.8597	0.8846
4	0.8788	0.8639	0.8641
5	0.8629	0.8645	0.8658
6	0.8658	0.8644	0.8657
7	0.8655	0.8642	0.8656
8	0.8628	0.8641	0.8656
9	0.8650	0.8640	0.8656
10	0.8628	0.8638	0.8656

APPENDIX

The error estimates (3.8) and (3.9) can be written using (3.7) as

$$(A.1) \quad h^{7/2} \cdot \|\phi^{(iv)}\| \left\{ \frac{1}{24\sqrt{2-h}} + \frac{5}{192} \left(\frac{1}{\sqrt{4-h}} \ell n \left| \frac{1+\sqrt{\frac{2-h}{4-h}}}{\sqrt{\frac{2-h}{4-h}}} \right| + \frac{1}{\sqrt{2-h}} \sum_{k=1}^{2n-1} \frac{1}{2k-1} \right) \right\}$$

$$= C_1 \cdot \|\phi^{(iv)}\|$$

and

$$(A.2) \quad h^{7/2} \cdot \|\phi^{(iv)}\| \left\{ \frac{1}{24\sqrt{2-h}} + \frac{5}{384} \left(\frac{1-h}{(2-h)^{3/2}} + \frac{4}{\sqrt{2-h}} \sum_{k=1}^{n} \frac{1}{2k-1} \right) \right\}$$

$$= C_2 \cdot \|\phi^{(iv)}\|.$$

From (3.3), the error in solving the equation

$$\frac{1}{\pi} \int_{-1}^{1} \frac{g(t)dt}{t-s} = f(s), \qquad\qquad -1 < s < 1$$

is therefore given by

$$\|A^{-1}\| \cdot \max(C_1, C_2) \cdot \|\phi^{(iv)}\|.$$

The table below lists values of $\|A^{-1}\|_\infty$ and $\max(C_1, C_2)$ for selected values of n.

n	$\|A^{-1}\|_\infty$	$\max(C_1, C_2)$	$\|A^{-1}\|_\infty \cdot \max(C_1, C_2)$
1	6.3×10^0	9.3×10^{-2}	5.9×10^{-1}
2	3.8×10^1	8.3×10^{-3}	3.2×10^{-1}
4	1.9×10^2	7.9×10^{-4}	1.5×10^{-1}
8	7.8×10^2	7.7×10^{-5}	6.0×10^{-2}
10	2.9×10^3	7.5×10^{-6}	2.2×10^{-2}

References

[1] Muskhelishvili, N. I. Singular Integral Equations, P. Noordhoff Ltd., Gronigen, Holland, 1953.

[2] Gakhov, F. D. Boundary Value Problems, Pergamon Press, New York, 1966.

[3] Privalov, I. Boundary Properties of Analytic Functions (in Russian), Moscow-Leningrad, 2nd edition, 1950.

[4] Ivanov, V. V. The Theory of Approximate Methods and Their Application to the Numerical Solution of Singular Integral Equations, Noordhoff International Publishing, Leyden, 1976.

[5] Erdogan, F. "Approximate solutions of systems of singular integral equations," SIAM J. Appl. Math. 17, 1041-1060 (1969).

[6] Erdogan, F. and Gupta, G. D. "On the numerical solution of singular integral equations," Q. Appl. Math. 30, 525-534 (1972).

[7] Erdogan, F., Gupta, G. D., and Cook, T. S. "Numerical solution of singular integral equations," Mechanics of Fracture, Volume 1: Methods of analysis and solutions of crack problems, Noordhoff International Publishing, Leyden, 1973.

[8] Theocaris, P. S. and Ioakimidis, N. I. "Numerical integral methods for the solution of singular integral equations," Q. Appl. Math. 35, 173-183 (1977).

[9] Dow, M. L. and Elliot, D. "The numerical solution of singular integral equations over (-1,1)," SIAM J. Num. Anal. 16, 115-134 (1979).

[10] Stroud, A. H. Numerical Quadrature and Solution of Ordinary Differential Equations, Springer-Verlag, New York, 1974.

[11] Gerasoulis, A. and Srivastav, R. P. "A method for the numerical solution of singular integral equations with a principal value integra," (to appear).

[12] Gerasoulis, A. "Product integration methods for the solution of singular integral equations of Cauchy type," Rutgers University Dept. of Computer Science Report No. DCS-TR-86, New Brunswick, 1979.

[13] Ahlberg, J. H., Nilson, E. H., and Walsh, J. L. The Theory of Splines and Their Applications, Academic Press, New York, 1967.

[14] Netravali, A. N. and Figueiredo, R. J. P. "Spline approximation to the solution of the linear Fredholm integral equation of the second kind," SIAM J. Num. Anal. II, 538-549 (1974).

[15] Gabdulhaev, B. G., "Optimization of collocation methods," Soviet Math. Dokl. 20, 823-827 (1979).

[16] Gradshteyn, I. S. and Ryzhik, Table of Integrals, Series, and Products, Academic Press, New York, 1965.

[17] Hall, C. A. and Meyer, W. W. "Optimal error bounds for cubic spline interpolation," J. Approximation Theory 16, 105-122 (1976).

[18] De Boor, C. A Practical Guide to Splines, Springer-Verlag, New York, 1978.

[19] Rooke, D. P. and Sneddon, I. N. "The crack energy and stress intensity factor for a cruciform crack deformed by internal pressure," Int. J. Engng. Science 7, 1079-1089 (1969).

CAN DISSIPATION PREVENT THE BREAKING OF WAVES?

Constantine M. Dafermos
Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, R. I. 02912

ABSTRACT. We show that dissipative mechanisms induced by friction, viscosity or thermal diffusion prevent the breaking of relatively weak waves but are ineffective against waves of large amplitude.

I. INTRODUCTION. Compressible weak waves that can be modeled as solutions to quasilinear hyperbolic systems of conservation laws are generally amplified, as they propagate, and eventually break, due to the formation of shock waves. It is interesting to consider the situation where this destabilizing mechanism coexists and thus competes with dissipation. Damping induced by viscosity of the rate type is so powerful that it dominates and prevents the breaking of any wave. Far more interesting is the situation where damping is induced by "friction", thermal diffusion or viscosity of the Boltzmann type. In these cases the dissipation mechanism is subtler so it may prevent the breaking of relatively weak waves but is ineffective against waves of large amplitude.

In Section II we exhibit results of this type in the context of a simple model in which complete proofs can be displayed in short space. In Section III we survey related results for more complicated systems of physical interest in which the analysis, too elaborate to be presented here in detail, is based upon the same principles as the analysis of the model case of Section II so that the reader will have already gotten a taste of its flavor!

II. A MODEL CASE. We consider wave phenomena governed by the Hopf equation

$$u_t + uu_x = 0. \tag{II.1}$$

Expansion waves are spreading out and get weaker while compression waves are amplified and eventually break. Specifically,

PROPOSITION II.1. The Cauchy problem for (II.1) with initial conditions $u(0,x) = \bar{u}(x) \in C^1(-\infty,\infty)$, with bounded derivative, has a global C^1-smooth solution if and only if $\bar{u}_x(x) \geq 0$, $-\infty < x < \infty$. When $\bar{u}_x(x)$ takes negative values, a local C^1-smooth solution exists which

breaks down at $t = [-\inf \overline{u}_x]^{-1}$.

PROOF. The characteristic equations read

$$\begin{cases} \dfrac{dx}{dt} = u \\[2em] \dfrac{du}{dt} = 0 \end{cases} \qquad\qquad (II.2)$$

so that characteristics are straight lines along which u remains constant. Thus, when $\overline{u}(x)$ is nondecreasing on $(-\infty,\infty)$, the fan of characteristics diverges and a global C^1-smooth solution exists. On the other hand, when $\overline{u}_x(x)$ takes negative values, characteristics collide and shock waves develop.

In order to determine the exact time the first wave breaks, we follow the evolution of u_x along characteristics. We set $v(t,x) = u_x(t,x)$ and take the derivative of (II.1) with respect to x, thus

obtaining

$$v_t + uv_x + v^2 = 0 \qquad\qquad (II.3)$$

or

$$\frac{dv}{dt} + v^2 = 0. \qquad\qquad (II.4)$$

Thus $\inf\limits_x v(t,x)$ will be bounded for $t < [-\inf \overline{u}_x(x)]^{-1}$ but will tend to $-\infty$ as $t \to [-\inf \overline{u}_x(x)]^{-1}$. As a matter of fact, when $\overline{u}_x(x)$ attains a minimum on $(-\infty,\infty)$ at a point $\overline{x}, v$ along the characteristic issuing at $\overline{x}$ will tend to $-\infty$, as $t \to [-\overline{u}_x(\overline{x})]^{-1}$ so that the wave emanating from $\overline{x}$ will break at $t = [-\overline{u}_x(\overline{x})]^{-1}$. ▤

Let us now investigate the effect of the presence of frictional damping,

$$u_t + uu_x + \mu u = 0, \quad \mu > 0. \qquad\qquad (II.5)$$

We will show that even compression waves do not break so long as their amplitude is not large.

PROPOSITION II.2. The Cauchy problem for (II.5) with initial
conditions $u(0,x) = \bar{u}(x) \in C^1(-\infty,\infty)$, with bounded derivative, has a
global C^1-smooth solution if and only if $\bar{u}_x(x) \geq -\mu$, $-\infty < x < \infty$, and
the amplitude of waves emanating from points $\bar{x}$ with $\bar{u}_x(\bar{x}) > -\mu$
decays to zero exponentially, as $t \to \infty$. When $\bar{u}_x(x)$ takes values less
than $-\mu$, a local C^1-smooth solution exists which breaks down as
$t \to \mu^{-1}\log[m/(m+\mu)]$, where $m = \inf \bar{u}_x(x)$.

PROOF. In the place of (II.2) we now have characteristic
equations

$$\begin{cases} \dfrac{dx}{dt} = u \\[2em] \dfrac{du}{dt} = -\mu u. \end{cases} \tag{II.6}$$

Setting, as before, $v(t,x) = u_x(t,x)$ and taking the derivative of
(II.5) with respect to x we obtain, in the place of (II.4),

$$\frac{dv}{dt} + v^2 + \mu v = 0. \tag{II.7}$$

It is now clear that if $\bar{u}_x(x) \geq -\mu$, $-\infty < x < \infty$, $u_x(t,x)$ will be bounded
and a global C^1-smooth solution will exist. Furthermore, v along any
characteristic issuing from a point $\bar{x}$ with $\bar{u}_x(\bar{x}) > -\mu$ decays to zero
exponentially, as $t \to \infty$. On the other hand, when $\bar{u}_x(x)$ takes values
less than $-\mu$, $\inf_x u_x(t,x)$ will be bounded for $t < \mu^{-1}\log[m/(m+\mu)]$
but will tend to $-\infty$ as $t \to \mu^{-1}\log[m/(m+\mu)]$, where $m = \inf \bar{u}_x(x)$.
In fact, if $\bar{u}_x(x)$ attains its minimum $m < -\mu$ at a point $\bar{x}$, v along
the characteristic issuing at $\bar{x}$ will tend to $-\infty$ as $t \to \mu^{-1}\log[m/(m+\mu)]$
so that the wave emanating from $\bar{x}$ will break at $t = \mu^{-1}\log[m/(m+\mu)]$. ▦

As seen from the above proof, the advantage of the method of
characteristics lies in that it yields explicitly the threshold amplitude
beyond which waves break as well as the time the first wave breaks. On
the other hand, the method is very special and it may be expected to work
only when the equations are very simple. We now state another result in
the same spirit which is less precise but whose proof is more versatile
and thus amenable to far reaching generalizations:

PROPOSITION II.3. Consider the Cauchy problem for (II.5) under initial conditions $u(0,x) = \bar{u}(x)$ with $\bar{u}_x(x), \bar{u}_{xx}(x)$ in $L^2(-\infty,\infty)$. Then, if

$$\|\bar{u}_x\|_{L^2} \|\bar{u}_{xx}\|_{L^2} < \frac{2\mu^2}{25}, \qquad (II.8)$$

there exists a global C^1-smooth solution $u(t,x)$ such that $u_x(t,\cdot)$, $u_{xx}(t,\cdot)$ are in $L^2(-\infty,\infty)$ and their L^2 norms decay to zero exponentially, as $t \to \infty$.

PROOF. We first give the idea of the proof. Assuming that a sufficiently smooth solution $u(t,x)$ exists on $[0,T) \times (-\infty,\infty)$, we differentiate (II.5) with respect to x, we multiply by $2u_x$, we integrate over $[0,s) \times (-\infty,\infty)$, $0 < s < T$, and integrate by parts, using the identity $2(uu_x)_x u_x = (uu_x^2)_x + u_x^3$, thus arriving at

$$\int_{-\infty}^{\infty} u_x^2(s,x)dx + \int_0^s\int_{-\infty}^{\infty} (2\mu+u_x)u_x^2 dxdt = \int_{-\infty}^{\infty} \bar{u}_x^2(x)dx, \qquad (II.9)$$

from which we could get an L^2 bound on u_x, uniform in time, if we had $|u_x(t,x)| < 2\mu$. This appears, of course, useless since pointwise bounds are locally stronger than L^2 bounds so one would have to assume more to get less. One may attempt to obtain pointwise bounds on u_x by establishing first L^2 bounds on u_{xx}. To this end, we differentiate (II.5) twice with respect to x, we multiply by $2u_{xx}$, we integrate over $[0,s) \times (-\infty,\infty)$ and integrate by parts to derive the analog of (II.9) for second derivatives. The anticipated difficulty is that we now may have to assume pointwise bounds on u_{xx} in order to obtain L^2 bounds on u_{xx}. This danger, however, does not materialize! The derived estimate, upon using in the integration by parts the identity

$$2(uu_x)_{xx}u_{xx} = (uu_{xx}^2)_x + 5u_x u_{xx}^2, \quad \text{reads}$$

$$\int_{-\infty}^{\infty} u_{xx}^2(s,x)dx + \int_0^s\int_{-\infty}^{\infty} (2\mu+5u_x)u_{xx}^2 dxdt = \int_{-\infty}^{\infty} \bar{u}_{xx}^2(x)dx. \qquad (II.10)$$

The miracle is that only a pointwise bound $5|u_x(t,x)| < 2\mu$ on u_x is needed in order to get a uniform L^2 bound on u_{xx}! This is not a coincidence but rather a consequence of the algebraic structure of the

190

operator of differentiation. Consequently, as it will become evident
in what follows, this methodology has wide applicability.

It is now easy to synthesize our proof. Let $[0,T)$, $T \leq \infty$, be
the maximal interval with the property that there is a C^1-smooth solution
$u(t,x)$ on $[0,T) \times (-\infty,\infty)$ such that $u_x(t,\cdot), u_{xx}(t,\cdot)$ are in
$L^\infty_{loc}([0,T); L^2(-\infty,\infty))$ and

$$\| u_x(t,\cdot) \|_{L^2} \| u_{xx}(t,\cdot) \|_{L^2} < \frac{2\mu^2}{25} , \quad 0 \leq t < T. \quad (II.11)$$

The existence of such a T follows from a straightforward local existence
theorem and assumption (II.8). For $s \in [0,T)$ we have estimates (II.9)
and (II.10); (the given derivation of (II.10) is only formal, within the
present function class, but the estimate can be easily established
rigorously either through difference approximations or via a density
argument).

By account of (II.11) and

$$u_x^2(t,x) = \int_{-\infty}^{x} (u_x^2)_x dx \leq 2 \| u_x(t,\cdot) \|_{L^2} \| u_{xx}(t,\cdot) \|_{L^2}, \quad (II.12)$$

we get

$$| u_x(t,x) | < \frac{2\mu}{5} , \quad -\infty < x < \infty, \quad 0 \leq t < T, \quad (II.13)$$

which, in conjunction with (II.9) and (II.10), implies that $\| u_x(s,\cdot) \|_{L^2}$
and $\| u_{xx}(s,\cdot) \|_{L^2}$ are nonincreasing on $[0,T)$. Thus, if $T < \infty$, we
may extend $u(t,x)$ up to $t = T$ and (II.11) will now hold for $t = T$.
But then, by the local existence theorem, $u(t,x)$ can be extended onto
a small interval beyond T, still satisfying (II.11), and this is a
contradiction since $[0,T)$ was assumed maximal. Therefore, $T = \infty$ and
the solution is global.

Exponential decay of $\| u_x(t,\cdot) \|_{L^2}$ and $\| u_{xx}(t,\cdot) \|_{L^2}$ follows
easily from (II.9) and (II.10) upon observing that $2\mu + u_x(t,x)$ and
$2\mu + 5u_x(t,x)$ are uniformly positive on $[0,\infty) \times (-\infty,\infty)$. ▨

Let us now equip (II.1) with a dissipative mechanism induced by
viscosity of the Boltzmann type:

$$u_t + uu_x + \int_0^t a'(t-\tau)uu_x d\tau = 0, \quad (II.14)$$

where $a(t)$ is a smooth relaxation function with properties to be
specified below.

191

The casual observer does not discern any similarity between (II.5) and (II.14): Damping is instantaneous in (II.5) but distributed over the entire history of the solution in (II.14).Nevertheless,the following argument (compare with MacCamy [7]) reveals a close similarity between these two equations.

Let $k(t)$ be the resolvent kernel associated with $a'(t)$; that is $k(t)$ is the solution to the linear Volterra integral equation:

$$k(t) + \int_0^t a'(t-\tau)k(\tau)d\tau = -a'(t).$$ (II.15)

Taking the convolution of (II.14) with $k(t)$ and after a simple calculation we arrive at

$$u_t + uu_x + k(0)u + \int_0^t k'(t-\tau)ud\tau = k(t)\bar{u}$$ (II.16)

where $\bar{u}(x) = u(0,x)$. We thus observe that when $-a'(0) = k(0) > 0$, (II.16) contains the frictional damping term $k(0)u$. In fact, when $a(t) = e^{-t}$, then $k(t) \supset 1$ so that (II.16) essentially reduces to (II.5) (the forcing term $\bar{u}$ on the right-hand side of (II.16) can be handled easily.)

It is not easy to establish the existence of global smooth solutions to (II.16) by the method of characteristics since the integration in (II.16) is along lines $x = $ const. rather than along characteristics. In contrast, it is straightforward to adapt the energy method employed in the proof of Proposition II.3, provided that $k(t) \in L^2(0,\infty)$ (in order to handle the forcing term $k(t)\bar{u}$) and that there is $\mu > 0$ with the property

$$\int_0^s v(t)\frac{d}{dt} \int_0^t k(t-\tau)v(\tau)d\tau dt \geq \mu \int_0^s v^2(t)dt,$$ (II.17)

for any $s \in (0,\infty)$ and every $v(t) \in L^2_{loc}(0,\infty)$. For assumptions on $a(t)$ that would guarantee the above properties of $k(t)$, we refer the reader to [7]. Under these conditions, we obtain easily, in the place of (II.9) and (II.10),

$$\int_{-\infty}^\infty u_x^2(s,x)dx + \int_0^s\int_{-\infty}^\infty (2\mu+u_x)u_x^2 dxdt$$ (II·18)

$$\leq \int_{-\infty}^\infty \bar{u}_x^2(x)dx + 2\int_0^s k(t)\int_{-\infty}^\infty \bar{u}_x u_x dxdt,$$

$$\int_{-\infty}^\infty u_{xx}^2(s,x)dx + \int_0^s\int_{-\infty}^\infty (2\mu+5u_x)u_{xx}^2 dxdt$$ (II.19)

$$\leq \int_{-\infty}^\infty \bar{u}_{xx}^2(x)dx + 2\int_0^s k(t)\int_{-\infty}^\infty \bar{u}_{xx}u_{xx}dxdt,$$

respectively. On the strength of the above estimates, following closely
the pattern of the proof of Proposition II.3, we establish

PROPOSITION II.4. Assume that the relaxation function $a(t)$
induces, through (II.15), a resolvent kernel $k(t) \in L^2(0,\infty)$ which
satisfies (II.17). Consider the Cauchy problem for (II.14) (or,
equivalently, for (II.16)) with initial conditions $u(0,x) = \bar{u}(x)$ where
$\bar{u}_x$ and $\bar{u}_{xx}$ are in $L^2(-\infty,\infty)$. If $\|\bar{u}_x\|_{L^2}$, $\|\bar{u}_{xx}\|_{L^2}$ are sufficiently
small, then there exists a global C^1-smooth solution $u(t,x)$ and
$u_x(t,\cdot)$, $u_{xx}(t,\cdot)$ are in $L^\infty([0,\infty);L^2(-\infty,\infty)) \cap L^2([0,\infty);L^2(-\infty,\infty))$.

III. SURVEY OF KNOWN RESULTS. Consider the second order quasi-
linear wave equation in one space variable,

$$w_{tt} - \sigma(w_x)_x = 0, \quad \sigma' > 0, \sigma'' > 0, \tag{III.1}$$

which is equivalent (upon setting $u = w_x$, $v = w_t$) to the genuinely
nonlinear system of hyperbolic conservation laws

$$\begin{cases} u_t - v_x = 0 \\ v_t - \sigma(u)_x = 0 \ . \end{cases} \tag{III.2}$$

The characteristic speeds of (III.2) are $\pm\sqrt{\sigma'(u)}$ generating a family
of forward and a family of backward characteristics.

The principal Riemann invariants are defined by

$$r = v + \int_0^u \sqrt{\sigma'(\xi)}\,d\xi, \quad s = v - \int_0^u \sqrt{\sigma'(\xi)}\,d\xi. \tag{III.3}$$

For C^1-smooth solutions, r remains constant along forward characteris-
tics and s remains constant along backward characteristics.

By monitoring the evolution of $r_x(s_x)$ along forward (backward)
characteristics, Lax [6] has established the following analog to
Proposition II.1:

PROPOSITION III.1. The Cauchy problem for (III.1) (or, equivalent-
ly, for (III.2)) under initial conditions $w_x(0,x) = u(0,x) = \bar{u}(x) \in C^1(-\infty,\infty)$,
$w_t(0,x) = v(0,x) = \bar{v}(x) \in C^1(-\infty,\infty)$ has a global C^1-smooth solution if and

only if $r(\bar{u}(x),\bar{v}(x))_x \leq 0$, $s(\bar{u}(x),\bar{v}(x))_x \leq 0$, $-\infty < x < \infty$. When $r(\bar{u}(x),\bar{v}(x))_x$ and/or $s(\bar{u}(x),\bar{v}(x))_x$ take positive values, there is a local C^1-smooth solution which breaks down at $t = T$ given asymptotically (for small initial data) by

$$T \sim \min\{[\frac{\sigma''(0)}{\sigma'(0)} \sup r(\bar{u}(x),\bar{v}(x))_x]^{-1}, [\frac{\sigma''(0)}{\sigma'(0)} \sup s(\bar{u}(x),\bar{v}(x))_x]^{-1}\}.$$

(III.4)

We now equip (III.1) with a frictional damping mechanism, viz.,

$$w_{tt} - \sigma(w_x)_x + \mu w_t = 0, \quad \mu > 0,$$

(III.5)

or, in system form,

$$\begin{cases} u_t - v_x = 0 \\ v_t - \sigma(u)_x + \mu v = 0 . \end{cases}$$

(III.6)

Nishida [10] estimates the growth of $r_x(s_x)$ along forward (backward) characteristics and deduces the following theorem, analogous to Proposition II.2:

PROPOSITION III.2. The Cauchy problem for (III.5) (or, equivalently, for (III.6)) under initial conditions $w_x(0,x) = u(0,x) = \bar{u}(x) \in C^1(-\infty,\infty)$, $w_t(0,x) = v(0,x) = \bar{v}(x) \in C^1(-\infty,\infty)$ has a global C^1-smooth solution provided that $\bar{u}_x(x)$, $\bar{v}_x(x)$ are bounded and $|r(\bar{u}(x),\bar{v}(x))_x|$, $|s(\bar{u}(x),\bar{v}(x))_x|$ are sufficiently small.

On the other hand, Kosinski [5] and Slemrod [11] have shown that when $r(\bar{u}(x),\bar{v}(x))_x$ and/or $s(\bar{u}(x),\bar{v}(x))_x$ take large values, waves generally break and no global smooth solution exists.

The problem of existence of global solutions to (III.5) (as well as to the multidimensional analog of (III.5)) was also studied by Matsumura [9] via energy estimates akin to those used in the proof of Proposition II.3. This approach yields the following:

PROPOSITION III.3. Assume that σ is C^3-smooth and $\sigma'(0) > 0$. Consider the Cauchy problem for (III.5) under initial conditions $w_x(0,x) = \bar{u}(x)$, $w_t(0,x) = \bar{v}(x)$, with $\bar{u}, \bar{u}_x, \bar{u}_{xx}, \bar{v}, \bar{v}_x, \bar{v}_{xx}$ in $L^2(-\infty,\infty)$. When $\|\bar{u}\|_{L^2}$, $\|\bar{u}_x\|_{L^2}$, $\|\bar{u}_{xx}\|_{L^2}$, $\|\bar{v}\|_{L^2}$, $\|\bar{v}_x\|_{L^2}$ and $\|\bar{v}_{xx}\|_{L^2}$ are sufficiently small, there exists a global C^2-smooth solution with

derivatives of first, second and third order in $L^\infty([0,\infty); L^2(-\infty,\infty))$. Furthermore, as $t \to \infty$, second order derivatives decay to zero, uniformly as well as in $L^2(-\infty,\infty)$.

We now turn to damping mechanisms of the memory type. Our first example is

$$w_t = \int_0^t a(t-\tau)\sigma(w_x)_x \, d\tau \tag{III.7}$$

which is a model of the heat flow equation in a material with memory [3] (w is temperature). Here σ is a smooth function with $\sigma'(0) > 0$ and $a(t)$ is a relaxation function normalized so that $a(0) = 1$ and having properties to be specified below.

Upon differentiating (III.7) with respect to t we obtain

$$w_{tt} - \sigma(w_x)_x - \int_0^t a'(t-\tau)\sigma(w_x)_x \, d\tau = 0 \tag{III.8}$$

which bears to (III.5) the same relationship that (II.14) bears to (II.5). In particular, as in Section II, we may employ the resolvent kernel $k(t)$ of $a'(t)$ (cf. (II.15)) to rewrite (III.8) into the equivalent form

$$w_{tt} - \sigma(w_x)_x + k(0)w_t + \int_0^t k'(t-\tau)w_\tau \, d\tau = 0, \tag{III.9}$$

analogous to (II.16). Exploiting the similarity between (III.9) and (III.5), MacCamy [7] establishes the existence of global C^2-smooth solutions to (III.9) (and thereby to (III.7)) by adapting the afore-mentioned methodology of Nishida for (III.5), namely, by estimating the growth of r_x and s_x along characteristics; he imposes assumptions on $a(t)$ guaranteeing that $k(t)$ satisfies (II.17) and that $k'(t)$, $k''(t)$ decay sufficiently fast, as $t \to \infty$. A necessary condition for (II.17) is that $a(\infty) = 0$, an assumption compatible with the physical interpretation of (III.7). Subsequently, Dafermos and Nohel [1] established existence of global C^2-smooth solutions to (III.9) by means of energy estimates thus arriving at a theorem analogous to Proposition II.4. Finally, Staffans [13] proved existence by employing energy estimates derived directly for (III.7). Here is a representative result from [13]:

PROPOSITION III.4. Assume that σ is C^3-smooth with $\sigma'(0) > 0$, and that $a(t)$ is a strongly positive definite kernel with $a'(t)$, $a''(t) \in L^1(0,\infty)$. Consider the Cauchy problem for (III.7) under the initial condition $w(0,x) = w(x)$ where w_x, w_{xx}, w_{xxx} are in $L^2(-\infty,\infty)$. When $\|w_x\|_{L^2}$, $\|w_{xx}\|_{L^2}$, $\|w_{xxx}\|_{L^2}$ are sufficiently small, there exists a global C^2-smooth solution with derivatives of first, second and

third order in $L^\infty([0,\infty); L^2(-\infty,\infty))$. Furthermore, as $t \to \infty$, second order derivatives decay to zero uniformly as well as in $L^2(-\infty,\infty)$.

As another, related, example, consider

$$w_{tt} - \sigma(w_x)_x - \int_0^t a'(t-\tau)\phi(w_x)_x d\tau = 0 \qquad (III.10)$$

which is a model for the equation of motion of a nonlinear viscoelastic material. We normalize ϕ and a so that $\phi'(0) = \sigma'(0)$ and $a(0) = 1$. Here w is displacement, σ is the instantaneous elastic stress and $\sigma_e = \sigma - [1-a(\infty)]\phi$ is the equilibrium stress. The physically natural assumptions are $\sigma'(0) = \phi'(0) > 0$ and $\sigma_e'(0) = a(\infty)\phi'(0) > 0$.

When $\phi = \sigma$, (III.10) reduces to (III.8) and may therefore be rewritten in the form (III.9). However, here $a(\infty) > 0$ so that the kernel $k(t)$ cannot satisfy (II.17). Nevertheless, MacCamy [8] devised an alternative line of estimates, compatible with the physically reasonable assumptions, and established a global existence theorem which hinges upon pointwise bounds on r_x and s_x along characteristics. Subsequently, Dafermos and Nohel [1] and Staffans [13] considered the same problem by means of energy estimates. In [1] the estimates are derived for Equation (III.9) while in [13] the estimates are established directly for Equation (III.8). In fact, Proposition III.4 also covers the present situation in the special case $w_t(0,x) = 0$.

The general case (III.10), with ϕ different from σ, is studied by Dafermos and Nohel [2] through energy estimates. They assume that $a(t)$ is a strongly positive definite kernel and $a(t) = a(\infty) + A(t)$, where $a(\infty) > 0$, $A(t),A'(t),A''(t)$ in $L^1(0,\infty)$, and establish an existence theorem analogous to Proposition III.4.

We should emphasize that the aforementioned methods of Dafermos and Nohel [1,2] and Staffans [13] apply also to the mixed initial-boundary value problem for Equations (III.8) and (III.10) as well as to the corresponding problems for the two-and three-space dimensional versions of these equations.

As our last example we consider the conservation equations of one-dimensional nonlinear thermoelasticity:

$$\begin{cases} w_{tt} - \sigma(w_x,\theta)_x = 0 \\ \theta\eta(w_x,\theta)_t + q(\theta_x)_x = 0 \end{cases} \qquad (III.11)$$

where w is the motion, θ is the temperature, σ is the stress, η is the entropy and q is the heat flux. Equation (III.11)$_1$ expresses

conservation of linear momentum and $(III.11)_2$ expresses conservation of energy. It is important that

$$\sigma_\theta(u,\theta) + \eta_u(u,\theta) = 0. \qquad (III.12)$$

Physically natural assumptions are $\sigma_u > 0$, $\eta_\theta > 0$ and $q' < 0$. Dissipation is here induced by thermal diffusion which manifests itself through the presence of the term $q(\theta_x)_x$. The question is whether the coupling between the two equations in (III.11) is sufficiently effective so that the "parabolic" $(III.11)_2$ may prevent the breaking of waves by the "hyperbolic" $(III.11)_1$. It turns out that the effectiveness of the coupling is indeed ensured by (III.12).

Slemrod [12] considers the mixed initial-boundary value problem for (III.11) on $(0,1) \times (0,\infty)$ with initial conditions $w_x(0,x) = \bar{u}(x)$, $w_t(0,x) = \bar{v}(x)$, $\theta(0,x) = \bar{\theta}(x)$ and boundary conditions $w_x(t,0) = w_x(t,1) = 0$ and $\theta(t,0) = \theta(t,1) = 0$. By means of energy estimates he establishes the existence of a global smooth solution under the assumption that $\bar{u}$, $\bar{u}_x$, $\bar{u}_{xx}$, $\bar{v}$, $\bar{v}_x$, $\bar{v}_{xx}$, $\bar{v}_{xxx}$, $\bar{\theta}, \bar{\theta}_x, \bar{\theta}_{xx}$, $\bar{\theta}_{xxx}$ and $\bar{\theta}_{xxxx}$ are in $L^2(0,1)$ and their L^2-norms are sufficiently small.

In contrast to the examples discussed before, the proof of the above result depends crucially upon the one-dimensionality of the body and it is not known whether thermal dissipation may prevent the breaking of waves in two- and three-dimensional thermoelasticity.

Another dissipation mechanism that may prevent the breaking of waves is induced by attenuation due to spreading of a wave of fixed energy into a large portion of space. It is clear that the effectiveness of this mechanism will increase with the dimension of space and the results obtained so far require dimensionality higher than the dimension of physical space. For relevant information the reader may consult the interesting article by Klainerman [4].

June 26, 1980

IV. REFERENCES

[1]   DAFERMOS, C.M., and J.A. NOHEL, Energy methods for nonlinear
      hyperbolic Volterra integrodifferential equations. Comm. P.D.E.
      4(1979), 219-278.

[2]   DAFERMOS, C.M., and J.A. NOHEL, A nonlinear hyperbolic Volterra
      equation in viscoelasticity. Am. J. Math. (to appear).

[3]   CURTIN, M.E., and A.C. PIPKIN, A general theory of heat conduction
      with finite wave speeds. Arch. Rat. Mech. Anal. 31 (1968), 113-126.

[4]   KLAINERMAN, S., Global existence for nonlinear wave equations.
      Comm. Pure Appl. Math. 33(1980), 43-101.

[5]   KOSINSKI, W., Gradient catastrophe in the solution of nonconserva-
      tive hyperbolic systems. J. Math. Anal. Appl. 61(1977), 672-688.

[6]   LAX, P.D., Development of singularities of solutions of nonlinear
      hyperbolic partial differential equations. J. Math. Phys. 5(1964),
      611-613.

[7]   MacCAMY, R.C., An integrodifferential equation with applications
      in heat flow. Q. Appl. Math. 35(1977), 1-19.

[8]   MacCAMY, R.C., A model for one-dimensional, nonlinear viscoelasticity.
      Q. Appl. Math. 35(1977), 21-33.

[9]   MATSUMURA, A., Global existence and asymptotics of the solutions of
      the second order quasilinear hyperbolic equations with first order
      dissipation. Publ. Res. Inst. Math. Sci. Kyoto U., Ser. A, 13(1977),
      349-379.

[10] NISHIDA, T., Global smooth solutions for the second order quasi-
      linear wave equations with the first order dissipation (Unpublished).

[11] SLEMROD, M. Instability of steady shearing flows in a nonlinear
      viscoelastic fluid. Arch. Rat. Mech. Anal. 68(1978), 211-225.

[12] SLEMROD, M., Global existence, uniqueness and asymptotic stability
      of classical smooth solutions in one-dimensional non-linear thermo-
      elasticity. Arch. Rat. Mech. Anal. (to appear).

[13] STAFFANS, O., On a nonlinear hyperbolic Volterra equation. SIAM J.
      Math. Anal. (to appear).

INFLUENCE-FUNCTION APPROACH TO THE COMPUTING
OF THE DEFLECTION OF A FLOATING ELASTIC PLATE AND
NONUNIQUENESS OF THE SOLUTIONS

Chunsuke Takagi
Physical Sciences Branch
U.S. Army Cold Regions Research and Engineering Laboratory
Hanover, New Hampshire 03755

ABSTRACT. We have developed an analytical machinery that enables us to
compute the deflection of a floating elastic plate expectedly valid to any
shape under any boundary conditions. The boundary-value problems of the semi-
infinite plate are studied. We have found that the solutions are not unique.
The reason for the nonuniqueness is that uncontrolled deflection is produced
in the outside of a part that the plate covers in the infinite plane. The
condition that must be imposed to produce a unique solution for a plate cover-
ing a region other than the entire infinite plane is expectedly that the de-
flection is zero everywhere outside the region that the plate occupies.

INTRODUCTION

The topic of this paper is some boundary-value problems of the differen-
tial equation

$$\nabla^4 w + w = P\Delta(x-x_o)\delta(y-y_o) , \tag{1}$$

which governs the deflection w of a floating elastic plate sustaining a concen-
trated load P at a point (x_o, y_o), where ∇^2 is the Laplacian operator and $\delta(\ )$
the delta function. We use the nondimensional coordinates in (1).

The solution of (1) for a plate covering the entire infinite plane is

$$w(x,y) = -\frac{P}{2\pi} \text{kei}\sqrt{(x-x_o)^2 + (y-y_o)^2} . \tag{2}$$

This solution was derived by Wyman (14), but his derivation is not complete be-
cause he did not show that the substitution of (2) into the left-hand side of
(1) yields the singularity on the right-hand side of (1).

Livesley (9) used the double Fourier-transform to solve the differential
equation obtained by replacing the right-hand side of (1) with a distributed
load $q(x,y)$. He found the solution, however, only for the simple-edge boundary
condition with regard to the semi-infinite plate by use of the reflection
principle. Kerr (6) extended the application of the reflection principle to
many shapes of floating plates of simple-edge boundary condition by using Wyman's
solution (2) as an influence function.

With the unstated but implicit objective of finding a break through the dif-
ficulty of solving (1) for the deflection of floating elastic plates of various
shapes and various boundary conditions, Kerr (7) solved some simple boundary-
value problems of (1) with two different methods and showed that several

relationships involving Bessel functions must hold true. Takagi (11,12) analytically proved all the relationships shown by Kerr (7). Thus, a double-Fourier-transform formula,

$$\text{kei}\sqrt{(x-x_o)^2 + (y-y_o)^2} = -\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(\xi^2+\eta^2)^2+1}e^{i\xi(x-x_o)+i\eta(y-y_o)}d\xi d\eta \quad , \quad (3)$$

key to the development in this paper, where $i = \sqrt{-1}$, was found.

We can now show that (2) satisfies (1) exactly. On substituting the right-hand side of (3), the left-hand side of (1) becomes

$$\frac{P}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{i\xi(x-x_o) + i\eta(y-y_o)}d\xi d\eta \quad ,$$

which transforms to the right-hand side of (1) when the expression

$$\int_{-\infty}^{\infty}e^{i\xi x}d\xi = 2\pi\delta(x) \tag{4}$$

of the delta function as the Fourier transform of unity (ref. 3,4) is used.

Use of the key formula (3), analysis in Takagi (11,12), recognition that the solution w of eq (1) is a generalized function (Appendix I), and use of the Fourier transform of the generalized functions (3,4,5,8) have enabled us to institute an analytical machinery, called virtual reaction method, which we have expected to be effective for solving (1) for the deflection of floating elastic plates of various shapes and various boundary conditions, problems which have hitherto been impossible to solve.

In the following, first we show the derivation of (3) and some theorems basic to the operation of the virtual reaction method. Second, for solving the deflection of a semi-infinite plate, we introduce a set of virtual reactions that may include the effects of the reflection principle. Third, we solve the boundary-value problems of the semi-infinite plate by use of the virtual reaction method. It is found that the reflection principle yields a solution only for the simple-edge boundary condition, but not for other boundary conditions. With or without the effect of reflection principle, the virtual reaction method consistently yields a solution for every boundary condition. All the solutions, however, are not unique. The reason for the nonuniqueness is discovered and the condition that we expect will produce the unique solution is presented.

BASIC THEORY

We prove (3) by way of the following four propositions.

Proposition 1. The Fourier reciprocal relationship holds between the Hankel function of zeroth order and the exponential function in the following:

$$\int_{-\infty}^{\infty} e^{ix\zeta}\, H_0^{(1)}(\kappa e^{\pi i/2}\sqrt{a^2+\zeta^2})\,d\zeta \;=\; \frac{-2i}{\sqrt{x^2+\beta^2}}\, e^{-a\sqrt{x^2+\beta^2}}\;,$$

$$\int_{-\infty}^{\infty} e^{-a\sqrt{x^2+\beta^2}\,-\,i\zeta x}\,\frac{dx}{\sqrt{x^2+\beta^2}} \;=\; \pi i H_0^{(1)}(\kappa e^{\pi i/2}\sqrt{a^2+\zeta^2})\;,$$

(5)

where $H_0^{(1)}(\ )$ is a Hankel function of zeroth order; x and ζ are real and non-negative such that $x^2+\zeta^2 \neq 0$; and β is complex such that $\beta \neq 0$, $\arg\beta < \pi/2$, $x^2+\beta^2 \neq 0$. By $\sqrt{H}$ we mean such a branch of H_0, $\sqrt{H} \geq 0$. This most convention is observed throughout the paper.

Proof: We show the proof of (5)₂ in the following with the corrections of the misprints in the proof of Takagi [11,12] and in a more readable form. Formula (5)₂ is the Fourier inverse of (5)₁. The proof of (5)₁ without use of the Fourier transform is shown in Takagi [11,12].

Use of Barnes' integral representation of $H_0^{(1)}$ as

$$H_0^{(1)} \;=\; -\frac{1}{2\pi^2}\int_{-c-\infty i}^{-c+\infty i}\Gamma^2(-s)\left(-\frac{iz}{2}\right)^{2s}ds\;,$$

(a)

where c is any positive number and $|\arg(-iz)| < \pi/2$ [Watson [4], p. 210], transforms the single integral on the left-hand side of (5)₂

$$I \;=\; \int_{-\infty}^{\infty} e^{ix\zeta} H_0^{(1)}(\kappa e^{\pi i/2}\sqrt{a^2+\zeta^2})\,d\zeta$$

(b)

to the repeated integrals

$$I \;=\; -\frac{1}{2\pi^2}\int_{-\infty}^{\infty} e^{ix\zeta}\,d\zeta\int_{-c-\infty i}^{-c+\infty i}\Gamma^2(-s)\left(\frac{\beta}{2}\sqrt{a^2+\zeta^2}\right)^{2s}ds\;.$$

(c)

The absolute convergence of the integral (c) carries over to (b), because the condition $|\arg(-iz)| < \pi/2$ in (a) transforms to $\arg\beta < \pi/2$ in (c), which is one of the preassigned conditions of the formula (5)₁. Therefore, the order of integration in (c) may be exchanged. Moreover, restricting the original range of x, which is $x \leq 0$, to $x > 0$, we let $\zeta = in$ in (c) so that it becomes

$$ I = -\frac{a}{2\pi^2} \int_{-c-\infty i}^{-c+\infty i} \Gamma^2(-s)\left(\frac{\mu a}{2}\right)^{2s} ds \int_{-\alpha}^{\alpha} e^{iax\eta}(1+\eta^2)^s d\eta \quad . \tag{d}$$

To evaluate the internal single integral in (d)

$$ M = \int_{-\infty}^{\infty} e^{iaxz}(1+z^2)^s dz $$

by the contour integral method, we first note that the original range of x (i.e. $-\infty < x < \infty$) may be restricted to $0 < x$, $x \neq 0$, because I is an even function of x, as the right-hand side of (b) shows. The condition $x \neq 0$ is added to make the following contour integration feasible. Consider the contour in Figure 1 that starts at origin 0, goes along the positive real axis to A (i.e. $z = \infty$), takes a 90° turn along the infinitely large circle to reach B (i.e. $z = i\infty$), comes down along the imaginary axis to C (i.e. $z = i$), makes a 360° turn along an infinitely small circle clockwise around C, goes upward along the imaginary axis to reach D (i.e. $z = i\infty$), takes a 90° turn along the infinitely large circle to reach E (i.e. $z = -\infty$), and finally reaches origin 0, thus completing a circuit. No singularity of the integrand $\exp(iaxz)(1 + z^2)^s$ exists inside this closed contour. Among the integrals along the path mentioned above, the integrals along AB and DE vanish, provided $ax > 0$. The integral around C also vanishes. Therefore, on the condition that $x > 0$, and $x > 0$, we have
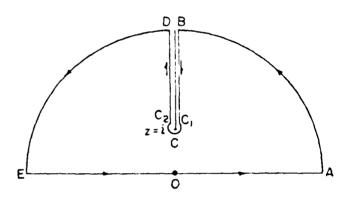
Figure 1. Transformation of integral M on the z-plane.

$$ M = -(\int_{B}^{C_1} + \int_{C_2}^{D})e^{iaxz}(1+z^2)^s dz \ , $$

where C_1 and C_2 are the initial and terminal points of the infinitely small circle around C. Letting $z = it$, where t is real, M reduces to

$$M = i(1 - e^{-2\pi i s}) \int_1^\infty e^{-axt}(1-t^2)^s dt \quad .$$

We now let s be

$$s = -\frac{1}{2} + ip$$

i.e., let c in (a) be 1/2, where p is a real number, in order to integrate M by use of the formula

$$K_\nu(z) = \frac{\Gamma(1/2)(z/2)^\nu}{\Gamma(\nu+1/2)} \int_1^\infty e^{-zt}(t^2-1)^{\nu-1/2} dt \quad , \tag{e}$$

which is valid when $\mathrm{Re}(\nu + 1/2) > 0$ and $|\arg(z)| < \pi/2$ (Watson 13, p. 172). When we let $\nu - 1/2 = s$ and $z = ax$ to integrate M, these two conditions are satisfied. Thus letting

$$(1-t^2)^s = e^{\pi i s}(t^2-1)^s \quad ,$$

M integrates to

$$M = -2\sin(\pi s) \frac{\Gamma(s+1)}{\sqrt{\pi}(ax/2)^{s+1/2}} K_{s+1/2}(ax) \quad .$$

In this way, (d) transforms to a single integral:

$$I = \frac{a}{\pi^2} \int_{-\frac{1}{2} - \infty i}^{-\frac{1}{2} + \infty i} \Gamma^2(-s)\left(\frac{\beta a}{2}\right)^{2s}\sin(\pi s)\frac{\Gamma(s+1)}{\sqrt{\pi}(ax/2)^{s+1/2}} K_{s+1/2}(ax)ds \quad .$$

Changing the Gamma function of the negative argument to the positive argument by the reflection formula

$$\Gamma(-s) = \frac{-\pi}{\Gamma(1+s)\sin(\pi s)} \quad ,$$

I becomes

$$I = \sqrt{2a/\pi^3 x} \int_{-\frac{1}{2} - \infty i}^{-\frac{1}{2} + \infty i} \frac{\pi}{\sin(\pi s)} \frac{K_{s+1/2}(ax)}{\Gamma(s+1)}\left(\frac{\beta^2 a}{2x}\right)^s ds \quad .$$

Taking the residues, I integrates to

$$I = -i\sqrt{8a/\pi x} \sum_{n=0}^\infty \frac{(-1)^n}{n!}\left(\frac{\beta^2 a}{2x}\right)^n K_{n+1/2}(ax) \quad .$$

Replacing $K_{n+1/2}(ax)$ with

$$K_\nu(z) = \frac{1}{2}\left(\frac{z}{2}\right)^\nu \int_0^\infty \exp\left(-t-\frac{z^2}{4t}\right)t^{-\nu-1}dt$$

(Watson 13, p. 183, I becomes

$$I = -\frac{ia}{\sqrt{\pi}}\sum_{n=0}^\infty \frac{(-1)^n}{n!}\int_0^\infty \left(\frac{a^2\beta^2}{4t}\right)^n \exp\left(-t-\frac{a^2x^2}{4t}\right)\frac{dt}{\sqrt{t^3}} \quad .$$

The order of the summation and integration may be exchanged, and we get

$$I = -\frac{ia}{\sqrt{\pi}}\int_0^\infty \exp\left(-t-\frac{a^2(x^2+\beta^2)}{4t}\right)\frac{dt}{\sqrt{t^3}} \quad .$$

Leting $t = \xi^{-2}$, this becomes

$$I = -\frac{2ai}{\sqrt{\pi}}\int_0^\infty \exp\left(-\xi^{-2}-\frac{a^2(x^2+\beta^2)}{4}\xi^2\right)d\xi \quad . \tag{f}$$

To integrate (f), we introduce a lemma:

$$\int_0^\infty e^{-\mu^2\xi^2 - \xi^{-2}}d\xi = e^{-2\mu}\frac{\sqrt{\pi}}{2\mu} \quad ,$$

if $\arg(\mu)$ is in the range

$$n\pi - \frac{\pi}{4} < \arg(\mu) < n\pi + \frac{\pi}{4} \quad ,$$

where n is an integer. When μ is in the above range, the integral is convergent. To prove the lemma, we first note that the integral

$$L = \int_0^\infty e^{-\mu^2\xi^2 - \xi^{-2}}d\xi$$

transforms to

$$L = e^{-2\mu}N \quad , \tag{g}$$

where

$$N = \int_0^\infty e^{-(\mu\xi-\xi^{-1})^2}d\xi \quad . \tag{h}$$

Letting $\xi = 1/(\mu\eta)$ and changing the resulting contour $0 \to \mu^{-1}\infty$ to $0 \to \infty$, we get

$$N = \int_0^\infty e^{-(\mu\eta - \eta^{-1})} \frac{1}{\mu\eta^2} d\eta .$$

(i)

Addition of (h) and (i) yields

$$2N = \frac{1}{\mu} \int_0^\infty e^{-(\mu\xi - \xi^{-1})} (\mu + \frac{1}{\xi^2}) d\xi .$$

Letting

$$\mu\xi - \xi^{-1} = t ,$$

we get

$$2N = \frac{1}{\mu} \int_{-\infty}^{\mu\infty} e^{-t} dt .$$

Changing the range of integration to the range from $-\infty$ to $+\infty$, N integrates to $N = \pi/(2\mu)$. Substituting this value into (g), the lemma is proved.

Letting μ be

$$\mu = \frac{a}{2} \sqrt{a^2 + \beta^2}$$

in the lemma, (f) is integrated, because μ above is obviously in the range prescribed before. Thus, under the conditions $x \neq 0$ and $a \neq 0$, Formula (5) is proved. Applying the analytical continuation, the condition $a \neq 0$ is extended to the condition $a^2 + \beta^2 \neq 0$. Because the integral is convergent at $x = 0$, the condition $x \neq 0$ may be removed. The proof is thus completed.

Proposition 2. The formula

$$\int_{-\infty}^{\infty} \frac{1}{z^2 + \gamma^2} e^{iyz} dz = \frac{\pi}{\gamma} e^{-\gamma|y|}$$

(6)

is true for any real number y, where γ is a complex number such that

$$Re(\gamma) > 0 .$$

(7)

Proof: First let us assume that $y > 0$. Then the integration of $e^{iyz}/(z^2 + \gamma^2)$ with regard to z along the upper semicircle BCA of infinitely large radius (see Figure 3) is equal to zero. Therefore, the integral on the left-hand side of (6) may transform to the contour integral passing through the real axis A B

and the infinitely large semicircle BCA. Because of (7) there is only one pole $z = \gamma i$ inside the contour. Applying the residue theorem, the lemma is proved for the case of $y > 0$.

If $y < 0$, we take the lower semicircle BDA. Inside the contour that passes through the real axis AOB and the infinitely large semicircle BDA, there is only one pole $z = -\gamma i$, where use is made of (7). Applying the residue theorem the lemma is proved for the case of $y < 0$.

If $y = 0$, (6) is still true as may be proved by the straightforward integration. The proof is thus completed.

Proposition 3. The single Fourier-transforms in (5) are equivalent to the following double Fourier-transforms,
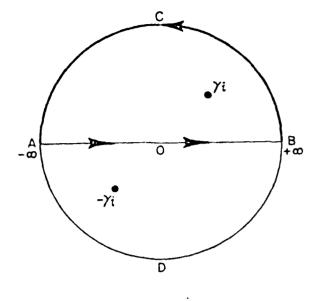
Figure 2 Contour integration for proving (6)

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_o^{(1)}(\beta e^{(\pi i/2)} \sqrt{x^2 + y^2}) e^{i\xi x + i\eta y} dx dy \;=\; \frac{-4i}{\xi^2 + \eta^2 + \beta^2} \;,$$

$$\tag{8}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\xi^2 + \eta^2 + \beta^2} e^{-ix\xi - iy\eta} d\xi d\eta \;=\; \pi^2 i H_o^{(1)}(\beta e^{(\pi i/2)} \sqrt{x^2 + y^2}) \;.$$

Proof: Application of the double Fourier-transform changes $(8)_1$ or $(8)_2$ to $(8)_2$ or $(8)_1$ respectively.

We show that application of a single Fourier-transform on $(8)_2$ yields $(5)_1$; i.e., we calculate

$$1 \;=\; \int_{-\infty}^{\infty} H_o^{(1)}(\beta e^{(\pi i/2)} \sqrt{x^2 + y^2}) e^{iax} dx \;,$$

which, by $(8)_2$, becomes

206

$$I = \frac{1}{\pi^2 i} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\xi^2 + \eta^2 + \beta^2} e^{-iy\eta} d\xi d\eta \int_{-\infty}^{\infty} e^{-ix\xi + iax} dx \quad .$$

Applying the Fourier transform (4) of unity on the internal single integral, I becomes

$$I = \frac{2}{\pi i} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\xi^2 + \eta^2 + \beta^2} e^{-iy\eta} \delta(a-\xi) d\xi d\eta \quad .$$

Using the sampling property, eq (51), of the delta function, I becomes

$$I = \frac{2}{\pi i} \int_{-\infty}^{\infty} \frac{1}{\xi^2 + (a^2 + \beta^2)} e^{-iy\eta} d\eta \quad .$$

Use of the integration formula (6) yields

$$I = \frac{-2i}{\sqrt{a^2 + \beta^2}} e^{-|y|\sqrt{a^2 + \beta^2}} \quad ,$$

which is the right-hand of $(5)_1$ expressed with the current notation.

We show that application of a single Fourier-transform on $(5)_1$, yields $(8)_1$. Letting x, ξ, and a in $(5)_1$ be ξ, x, and $|y|$, one may use $(5)_2$ to integrate

$$J = \int_{-\infty}^{\infty} e^{-i\eta y} dy \int_{-\infty}^{\infty} H_0^{(1)} (\beta e^{(\pi i/2)} \sqrt{x^2 + y^2}) e^{i\xi x} dx \quad .$$

Thus we have

$$J = \frac{-2i}{\sqrt{\xi^2 + \beta^2}} \int_{-\infty}^{\infty} e^{-i\eta y - |y|\sqrt{\xi^2 + \beta^2}} dy \quad .$$

Dividing the range of integration into positive and negative semi-infinite parts, J integrates to

$$J = \frac{-2i}{\sqrt{\xi^2 + \beta^2}} \left\{ \frac{1}{\sqrt{\xi^2 + \beta^2} + i\eta} + \frac{1}{\sqrt{\xi^2 + \beta^2} - i\eta} \right\} \quad ,$$

which reduces to the righ-hand side of $(8)_1$.

The equivalence of the two equations in (5) to the two equations in (8) is thus proved.

We prove in the next proposition a formula that is equivalent to (3).

Proposition 4.

$$\text{kei}\sqrt{x^2+y^2} \;=\; -\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(\xi^2+\eta^2)^2 + 1}\, e^{-ix\xi-iy\eta}\,d\xi d\eta \quad . \tag{9}$$

Proof: We first note that $\text{ker}x$ and $\text{kei}x$ are the real and imaginary parts of the right-hand side of the equation

$$\text{ker}(x) \;+\; i\text{kei}(x) \;=\; \frac{\pi i}{2}\, H_0^{(1)}(xe^{(3\pi i/4)}) \quad . \tag{a}$$

The conjugate complex of (a) is

$$\text{ker}(x) \;-\; i\text{kei}(x) \;=\; \frac{\pi i}{2}\, H_0^{(1)}(xe^{(\pi i/4)}) \quad , \tag{b}$$

as is proved below.

A form of the conjugate complex of the right-hand side of (a) is found to be

$$-\frac{\pi i}{2}\, H_0^{(2)}(xe^{(-3\pi i/4)}) \quad , \tag{c}$$

when use is made of the formula

$$\overline{H_0^{(1)}(z)} \;=\; H_0^{(2)}(\overline{z}) \;,$$

which is a special case of the formula 9·1·40 in Olver (4), where a bar indicates the conjugate complex of the underlying symbol and z a complex number. The expression (c) transforms to the right-hand side of (b) by use of the formula

$$H_0^{(2)}(ze^{-\pi i}) \;=\; -\,H_0^{(1)}(z) \quad ,$$

which is a special case of the formula 9·1·39 in Olver (4).

Solving (a) and (b) simultaneously for $\text{kei}x$, we find

$$\text{kei}(x) \;=\; \frac{\pi}{4}\{H_0^{(1)}(xe^{(3\pi i/4)}) \;-\; H_0^{(1)}(xe^{(\pi i/4)})\} \;. \tag{d}$$

Changing the argument x in (d) to $\sqrt{x^2+y^2}$, and applying $(8)_2$ to the first and second terms with $\beta = \exp(\pi i/4)$ and $\beta = \exp(-\pi i/4)$, respectively, we find

208

$$\text{kei}\sqrt{x^2+y^2} \;=\; \frac{1}{4\pi i} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-ix\xi - iy\eta}\{\frac{1}{\xi^2+\eta^2+i} \;-\; \frac{1}{\xi^2+\eta^2-i}\}d\xi d\eta \quad , \qquad \text{(e)}$$

which transforms to (9).

None of the formulas developed in the above is considered to be entirely new, because the formula

$$\int_{0}^{\infty} K_o(\beta\sqrt{x^2+\xi^2})\cos(\alpha\xi)d\xi \;=\; \frac{\pi}{2\sqrt{\alpha^2+\beta^2}}\, e^{-x\sqrt{\alpha^2+\beta^2}} \quad ,$$

which may be derived from $(5)_1$, is listed in Erdélyi (2). Some more propositions are derived in the following for the operation of the virtual reaction method that we develop below.

Proposition 5.

$$\text{kei}\sqrt{x^2+y^2} \;=\; \frac{1}{4i}\int_{-\infty}^{\infty} e^{-ix\xi}\{\frac{1}{\sqrt{\xi^2+i}}\, e^{-|y|\sqrt{\xi^2+i}} \;-\; \frac{1}{\sqrt{\xi^2-i}}\, e^{-|y|\sqrt{\xi^2-i}}\}d\xi, \quad (10)$$

where y is either positive, negative, or zero.

Proof: The double integral on the right-hand side of (e) above may be rewritten to the following repeated integrals.

$$\text{kei}\sqrt{x^2+y^2} \;=\; \frac{1}{4\pi i}\int_{-\infty}^{\infty} e^{-ix\xi}d\xi\int_{-\infty}^{\infty} e^{-iy\eta}\{\frac{1}{\eta^2+(\xi^2+i)} \;-\; \frac{1}{\eta^2+(\xi^2-i)}\}d\eta \quad .$$

Applying (6) to the integration with regard to η, we find (10).

Proposition 6. For $y \neq 0$,

$$\frac{\partial^n}{\partial y^n}\, \text{kei}\sqrt{x^2+y^2}$$

$$= \frac{1}{4i}(-\text{sgn}y)^n\int_{-\infty}^{\infty}\{(\xi^2+i)^{(n-1)/2}e^{-|y|\sqrt{\xi^2+i}} \;-\; (\xi^2-i)^{(n-1)/2}e^{-|y|\sqrt{\xi^2-i}}\}e^{-ix\xi}d\xi ,$$

$$\qquad\qquad (11)$$

where n is a nonnegative integer and

$$sgn(y) = 1 \quad \text{for} \quad y > 0$$
$$= -1 \quad \text{for} \quad y < 0 \quad .$$

For $y = 0$,

$$\left(\frac{\partial^n}{\partial y^n} kei\sqrt{x^2+y^2}\right)_{y=o} = -\frac{(-1)^n}{2\pi} \int_{-\infty}^{\infty} e^{-ix\xi} d\xi \int_{-\infty}^{\infty} \frac{\eta^n}{(\xi^2+\eta^2)^2+1} d\eta \quad . \tag{12}$$

Proof: Because, when $y \neq 0$, either of the following functions $(1/\sqrt{\xi \pm i})\exp(-|y|\sqrt{\xi^2 \pm i})$ is a good function of ξ in the sense of Lighthill (1958), the differentiation of (10) with regard to y may be carried out inside the integral sign on the right-hand side (see theorem 7.23 at p. 206 of Jones (1966)). Thus, for $y \neq 0$, the n-time differentiation of (10) with regard to y yields (11).

For $y = 0$, using (9) for $kei\sqrt{x^2+y^2}$, one finds (12), where part iii of theorem 7.14 at p. 198 of Jones (1966) is used to justify the exchange of the order of operating integrations and operating differentiation.

As (11) shows, the first derivative and all the even-order derivatives are continuous at $y = 0$. Especially

$$\left(\frac{\partial^3}{\partial y^3} kei\sqrt{x^2+y^2}\right)_{y=+o} = -\pi\delta(x) \quad ,$$

$$\left(\frac{\partial^3}{\partial y^3} kei\sqrt{x^2+y^2}\right)_{y=-o} = \pi\delta(x) \quad .$$

At $y = 0$, as (12) shows, for any positive odd integers $n = 2k+1$,

$$\left(\frac{\partial^{2k+1}}{\partial y^{2k+1}} kei\sqrt{x^2+y^2}\right)_{y=o} = 0 \quad ,$$

where $k \geq 0$. In the theory of generalized functions, the right-hand side of (12) can be estimated even for $n \geq 4$, because one may use the formula

$$\int_{-\infty}^{\infty} \xi^n e^{-ix\xi} d\xi = 2\pi i^n \frac{d^n \delta(x)}{dx^n} \tag{13}$$

to evaluate the divergent part arising from the internal single integral in (12). Eq. (13) is proved by successive differentiation of the Fourier transform (4) of unity (Gel'fand and Shilov (3), p. 38).

<u>Proposition 7.</u>

$$\int_{-\infty}^{\infty} e^{isx} \, kei\sqrt{(x-a)^2 + (y-b)^2}\,dx$$

$$= \frac{\pi}{2i} e^{ias}\{\frac{1}{\sqrt{s^2+1}} \, e^{-|y-b|\sqrt{s^2+1}} - \frac{1}{\sqrt{s^2-1}} \, e^{-|y-b|\sqrt{s^2-1}}\} \tag{14}$$

<u>Proof.</u> Using (10), we get

$$\int_{-\infty}^{\infty} e^{isx} \, kei\sqrt{(x-a)^2 + (y-b)^2}\,dx$$

$$= \frac{1}{4i} \int_{-\infty}^{\infty} e^{ia\xi}\{\frac{1}{\sqrt{\xi^2+1}} \, e^{-|y-b|\sqrt{\xi^2+1}} - \frac{1}{\sqrt{\xi^2-1}} \, e^{-|y-b|\sqrt{\xi^2-1}}\}d\xi \int_{-\infty}^{\infty} e^{i(s-\xi)x}\,dx.$$

Use of the Fourier transform (4) of unity and the sampling property, (51), of
the delta function reduce the right-hand side of the above equation to that of (14).

<u>Proposition 8.</u>

$$\int_{-\infty}^{\infty} e^{isx}(\frac{\partial^n}{\partial y^n} \, kei\sqrt{(x-a)^2 + (y-b)^2})\,dx$$

$$\tag{15}$$

$$= \frac{\pi}{2i} e^{ias}\{(s^2+1)^{(n-1)/2} e^{(y-b)\sqrt{s^2+1}} - (s^2-1)^{(n-1)/2} e^{(y-b)\sqrt{s^2-i}} \quad \text{for } b>y,$$

$$= \frac{\pi}{2i} (-1)^n e^{ias}\{(s^2+1)^{(n-1)/2} e^{(b-y)\sqrt{s^2+1}} - (s^2-1)^{(n-1)/2} e^{(b-y)\sqrt{s^2-i}}$$

$$\text{for } y>b,$$

where n is a nonnegative integer, b and y are unequal real numbers, and a is
a real number.

<u>Proof.</u> The case n = 0 is proved by (14). Because the theory of generalized functions
allows the exchange of the order of integration and operating differentiation in (14)
for y ≠ b (see Theorem, 7.23 at p, 206 Of Jones (5)), n-times differentiation of (14)
with regard to y yields (15).

<div align="center">THE VIRTUAL REACTION METHOD</div>

Let us consider a semi-infinite plate whose sole boundary is the x-axis.
In this analysis, the values of w(x,y) and its derivatives on the x-axis are not
the values at y = 0, but are chosen to be the limits as the positive y tends to
y = 0. We impose three kinds of boundary conditions:

the simple-edge condition,

$$w(x,+0) = 0 \ ,$$

$$\frac{\partial^2 w(x,+0)}{\partial y^2} = 0 \ ;$$

(16)

the fixed-edge condition,

$$w(x,+0) = 0 \ .$$

$$\frac{\partial w(x,+0)}{\partial y} = 0 \ ;$$

(17)

and the free-edge condition,

$$\frac{\partial^2 w(x,+0)}{\partial y^2} + \nu \frac{\partial^2 w(x,+0)}{\partial x^2} = 0 \ ,$$

(18)

$$\frac{\partial^3 w(x,+0)}{\partial y^3} + (2-\nu) \frac{\partial^3 w(x,+0)}{\partial x^2 \partial y} = 0 \ ;$$

where $\partial^{p+q} w(x,+0)/\partial x^p \partial y^q$ stands for $\lim_{y \to +0} \partial^{p+q} w(x,y)/\partial x^p \partial y^q$, in which p and q are nonnegative integers, and ν the Poisson ratio.

We assume that a concentrated load P is sustained at a point A($x = 0$, $y = y_0$), where $y_0 \geqq 0$. The singularity at A causes a singularity at B($x = 0$, $y = -y_0$), where, in consideration of the reflection principle, we place another concentrated load αP(i.e., a virtual reaction) in which α is an arbitrary real number. We place on the x-axis two unknown virtual reactions — an unknown vertical line load $p(x)$ and an unknown line couple $m(x)$. These loads cause the deflection

$$w(x,y) = -\frac{P}{2\pi} \text{kei}\sqrt{x^2 + (y-y_0)^2} - \frac{\alpha P}{2\pi} \text{kei}\sqrt{x^2 + (y+y_0)^2} +$$

(19)

$$+ \frac{1}{2\pi}\int_{-\infty}^{\infty} p(t)\text{kei}\sqrt{(x-t)^2 + y^2}dt + \frac{1}{2\pi}\int_{-\infty}^{\infty} m(t)(\frac{\partial}{\partial y}\text{kei}\sqrt{(x-y)^2 + y^2})dt \ .$$

The limits as positive y tends to $y = 0$ must be employed, because the third and higher odd-order derivatives of $w(x,y)$ in (19) with regard to y are discontinuous at $y = 0$ [in more detail, the values of the third and higher odd-order derivatives at $y = +0$, 0, and -0 are definite in the sense of generalized functions but, due to the existence of the integrand $\mathrm{kei}\sqrt{(x-t)^2 + y^2}$, are not equal with each other, as explained following Proposition 6]. Our task is to determine $p(x)$ and $m(x)$ to satisfy the boundary conditions stated above.

The Fourier transforms of (19) and its derivatives are facilitated by use of the differentiation formula (15) and

$$\int_{-\infty}^{\infty} e^{isx} \frac{\partial^n f(x,y)}{\partial x^n} \, dx = (-is)^n \int_{-\infty}^{\infty} e^{isx} f(x,y) dx \quad , \tag{20}$$

where $f(x,y)$ is a generalized function that vanishes at $|x| = \infty$. Eq. (20) is derived by repeating partial integrations on the left-hand side.

We now set the restriction

$$y_o > y > 0 \quad , \tag{21}$$

which is the range of y we must operate on to find the limits involved in the boundary conditions. Once the limits are found, we relax the condition to $y_o > y \geq 0$ by redefining the value $y = 0$ to be the limit at $y = +0$.

Use of (15) and (20) under the condition (21) yields the Fourier transform of (19)

$$
\begin{aligned}
\int_{-\infty}^{\infty} e^{isx} w(x,y) dx \ = \ & - \ \frac{P}{2i} \{ \frac{1}{\sqrt{s^2+i}} e^{(y-y_o)\sqrt{s^2+i}} - \frac{1}{\sqrt{s^2-i}} e^{(y-y_o)\sqrt{s^2-i}} \} - \\
& - \ \frac{\alpha P}{4i} \{ \frac{1}{\sqrt{s^2+i}} e^{-(y+y_o)\sqrt{s^2+i}} - \frac{1}{\sqrt{s^2-i}} e^{-(y+y_o)\sqrt{s^2-i}} \} + \\
& + \ \frac{1}{4i} \tilde{p}(s) \{ \frac{1}{\sqrt{s^2+i}} e^{-y\sqrt{s^2+i}} - \frac{1}{\sqrt{s^2-i}} e^{-y\sqrt{s^2-i}} \} - \\
& - \ \frac{1}{4i} \tilde{m}(s) \{ e^{-y\sqrt{s^2+i}} - e^{-y\sqrt{s^2-i}} \} \quad ,
\end{aligned} \tag{22}
$$

where we have introduced the notation $\widetilde{f}(s)$ to denote the Fourier transform of a function $f(x)$,

$$\bar{f}(s) = \int_{-\infty}^{\infty} e^{isx} f(x)dx \quad . \tag{23}$$

Using (20) for the m-time differentiation by x and performing the n-time differentiation with regard to y directly on (22), we find

$$\int_{-\infty}^{\infty} e^{isx} \frac{\partial^{m+n} w(x,y)}{\partial x^m \partial y^n} dx$$

$$= -P \frac{(-is)^m}{4i} \{(s^2+i)^{(n-1)/2} e^{(y-y_o)\sqrt{s^2+i}} - (s^2-i)^{(n-1)/2} e^{(y-y_o)\sqrt{s^2-i}}\} -$$

$$-aP \frac{(-is)^m(-1)^n}{4i} \{(s^2+i)^{(n-1)/2} e^{-(y+y_o)\sqrt{s^2+i}} - (s^2-i)^{(n-1)/2} e^{-(y+y_o)\sqrt{s^2-i}}\} +$$

$$+ \frac{(-is)^m(-1)^n}{4i} \bar{p}(s)\{(s^2+i)^{(n-1)/2} e^{-y\sqrt{s^2+i}} - (s^2-i)^{(n-1)/2} e^{-y\sqrt{s^2-i}}\} -$$

$$- \frac{(-is)^m(-1)^n}{4i} \bar{m}(s)\{(s^2+i)^{n/2} e^{-y\sqrt{s^2+i}} - (s^2-i)^{n/2} e^{-y\sqrt{s^2-i}}\} \quad . \tag{24}$$

In the region (21), the operation of integration with regard to x and the operation of differentiations are interchangeable. Letting $y = 0$ in (24), we find the formula

$$\int_{-\infty}^{\infty} e^{isx} \frac{\partial^{m+n} w(x,+0)}{\partial x^m \partial y^n} dx$$

$$= -[1+(-1)^n a]P \frac{(-is)^m}{4i} \{(s^2+i)^{(n-1)/2} e^{-y_o\sqrt{s^2+i}} - (s^2-i)^{(n-1)/2} e^{-y_o\sqrt{s^2-i}}\} +$$

$$+ \frac{(-is)^m(-1)^n}{4i} \bar{p}(s)\{(s^2+i)^{(n-1)/2} - (s^2-i)^{(n-1)/2}\} -$$

$$- \frac{(-is)^m(-1)^n}{4i} \bar{m}(s)\{(s^2+i)^{n/2} - (s^2-i)^{n/2}\} \quad , \tag{25}$$

which facilitates the transformation of the boundary conditions (16), (17), and (18). The Fourier transforms of the expressions contained in the boundary conditions (16), (17), (18) are written out by use of (25), and shown in Appendix II. In the final forms of the transformation of the boundary conditions, the

214

expressions shown on the left-hand sides of the following formulas are replaced with the respective right-hand sides,

$$\sqrt{s^2+i} - \sqrt{s^2-i} = 2i[\sqrt{s^2+i} + \sqrt{s^2-i}]^{-1} ,$$

$$(s^2+i)^{3/2} - (s^2-i)^{3/2} = 2i(2s^2 + \sqrt{s^4+1})[\sqrt{s^2+i} + \sqrt{s^2-i}]^{-1} , \qquad (26)$$

because the right-hand sides show clearly the asymptotic forms when $|s|$ increases indefinitely.

To replace $p(t)$ and $n(t)$ in (19) with $\tilde{p}(s)$ and $\tilde{m}(s)$, respectively, we substitute the formulas

$$p(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{p}(s)e^{-ist}ds ,$$

$$\qquad (27)$$

$$m(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{m}(s)e^{-ist}ds ,$$

into (19). Thus we find

$$w(x,y) = -\frac{P}{2\pi} \text{kei}\sqrt{x^2 + (y-y_o)^2} - \frac{\alpha P}{2\pi} \text{kei}\sqrt{x^2 + (y-y_o)^2} +$$

$$+ \frac{1}{8\pi i} \int_{-\infty}^{\infty} \tilde{p}(s)e^{-ixs}\{\frac{1}{\sqrt{s^2+i}} e^{-y\sqrt{s^2+i}} - \frac{1}{\sqrt{s^2-i}} e^{-y\sqrt{s^2-i}}\}ds -$$

$$\qquad (28)$$

$$- \frac{1}{8\pi i} \int_{-\infty}^{\infty} \tilde{m}(s)e^{-ixs}\{e^{-y\sqrt{s^2+i}} - e^{-y\sqrt{s^2-i}}\}ds ,$$

where use is made of $(15)_2$ for $n = 0$ and 1 with the substitution of $x = -t$, $a = -x$, and $b = 0$ to transform the integrands. Eq (28) is convenient for the substitution of $\tilde{p}(s)$ and $\tilde{m}(s)$ that will be found as a solution in the following.

THE BOUNDARY-VALUE PROBLEMS

Solution for the simple edge.

Simultaneously solving the Fourier transforms (56) and (58) (in Appendix II) of the boundary conditions in (16), we find

$$p(s) = -P \frac{1+\alpha}{2i} \sqrt{s^4+1}(\sqrt{s^2+1} + \sqrt{s^2-1})\{\frac{1}{\sqrt{s^2+1}} e^{-y_0\sqrt{s^2+1}} - \frac{1}{\sqrt{s^2-1}} e^{-y_0\sqrt{s^2-1}}\} \quad ,$$

<div align="right">(29)</div>

$$\tilde{m}(s) = -P \frac{1+\alpha}{2i} (\sqrt{s^2+1} + \sqrt{s^2-1})\{e^{-y_0\sqrt{s^2+1}} - e^{-y_0\sqrt{s^2-1}}\} \quad .$$

Substituting the equations in the above into (28), we find the solution for the simple-edge condition:

$$w(x,y) = -P \frac{1}{2\pi} \text{kei}\sqrt{x^2 + (y-y_0)^2} - P \frac{\alpha}{2\pi} \text{kei}\sqrt{x^2 + (y+y_0)^2} +$$

$$+ P \frac{1+\alpha}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\sqrt{s^4+1}(\sqrt{s^2+1} + \sqrt{s^2-1})\{\frac{1}{\sqrt{s^2+1}} e^{-y\sqrt{s^2+1}} - \frac{1}{\sqrt{s^2-1}} e^{-y\sqrt{s^2-1}}\} \times$$

$$\times \{\frac{1}{\sqrt{s^2+1}} e^{-y_0\sqrt{s^2+1}} - \frac{1}{\sqrt{s^2-1}} e^{-y_0\sqrt{s^2-1}}\}ds -$$

<div align="right">(30)</div>

$$- P \frac{1+\alpha}{2\pi}\int_{-\infty}^{\infty} e^{-isx}(\sqrt{s^2+1} + \sqrt{s^2-1})\{e^{-y\sqrt{s^2+1}} - e^{-y\sqrt{s^2-1}}\} \times$$

$$\times \{e^{-y_0\sqrt{s^2+1}} - e^{-y_0\sqrt{s^2-1}}\}ds \quad .$$

The formulas

$$\sqrt{x^2+i} + \sqrt{x^2-i} = \sqrt{2} \sqrt{\sqrt{x^4+1} + x^2} \quad ,$$

<div align="right">(31)</div>

$$\sqrt{x^2+i} - \sqrt{x^2-i} = \sqrt{2}i \sqrt{\sqrt{x^4+1} - x^2} \quad ,$$

where x is a real number, are convenient for the conversion, if necessary, of the complex form of the right-hand side of (30) (and the similar formulas that appear later) to a real form. Equations in (31) are proved by squaring them.

If we let $\alpha = -1$ in (30), the effects of the virtual reactions $p(x)$ and $m(x)$ disappear, and we find Kerr's (6) solution obtained by applying the reflection principle on the concentrated load P that is situated in the semi-infinite plane. This solution by Kerr's (6), however, is not sufficiently general because it does not give a solution for $y_0 = 0$. A solution for $y_0 = 0$ is given by (30) if $\alpha \neq -1$ but arbitrary. (We note that the condition $y_0 > 0$ assumed initially for finding a solution need not be observed at this stage.) The solution, therefore, is not unique. Moreover, pushing this line of inference further, we find that other kinds of virtual ractions, such as higher-order derivatives with regard to y, or more generally, any point-, line-, or area-singularities distributed in the lower half plane, $0 \geqq y \quad - \bullet$, may be adopted instead of $p(x)$ and $m(x)$ with proper change of the formulations to represent the effect of the assumed virtual reactions. This conclusion applies, as may be seen below, to all the solutions of any other boundary conditions. The nonuniqueness reveals the defect of the virtual reaction method. The remedy is presented later.

Solution for the fixed edge.

Simultaneously solving the Fourier transforms (56) and (57) (in Appendix II) of the boundary conditions in (17), we find $(29)_1$ and

$$\tilde{m}(s) = -P\,\frac{1-\alpha}{2i}(\sqrt{s^2+i} + \sqrt{s^2-i})\{e^{-y_0\sqrt{s^2+i}} - e^{-y_0\sqrt{s^2-i}}\} \quad , \tag{32}$$

Comparison of (32) with $(29)_2$ shows that the solution for the fixed edge is found by changing $-(1+\alpha)$ in the second intergral on the right-hand side of (30) to $1 - \alpha$. In this case it is impossible to eliminate all the effects of the virtual reactions $p(x)$ and $m(x)$ by giving α an appropriate number. In other words, the reflection principle fails to yield a solution for the fixed-edge boundary condition.

Solution for the free-edge.

Use of the appropriate Fourier transforms of the derivatives of $w(x,y)$ in Appendix II transforms the boundary conditions in (18) to the simultaneous equations for $\tilde{p}(s)$ and $\tilde{m}(s)$,

$$\tilde{p}(s)\{(\nu s^2 + \sqrt{s^4+1})/(\sqrt{s^2+i} + \sqrt{s^2-i})\} - \tilde{m}(s)\sqrt{s^4+1} = A\sqrt{s^4+1} \quad ,$$

$$\tag{33}$$

$$\tilde{p}(s) - \tilde{m}(s)\{(\nu s^2 + \sqrt{s^4+1})/(\sqrt{s^2+i} + \sqrt{s^2-i})\} = B \quad ,$$

where

$$A = \frac{P(1+\alpha)}{2i} \left[\frac{E}{\sqrt{s^2+i}} e^{-y_0\sqrt{s^2+i}} - \frac{\overline{E}}{\sqrt{s^2-i}} e^{-y_0\sqrt{s^2-i}} \right] \quad ,$$

$$B = \frac{P(1-\alpha)}{2i} \left[\overline{E} e^{-y_0\sqrt{s^2+i}} - E e^{-y_0\sqrt{s^2-i}} \right] \quad ,$$

$$E = (1-\nu)s^2 + i \quad ,$$

$$\overline{E} = (1-\nu)s^2 - i \quad .$$

$$(34)$$

The solution of the simultaneous equations on the above is expressed here in the following intermediate form,

$$p(s) = M / \{ \Delta (\sqrt{s^2+i} + \sqrt{s^2-i}) \} \quad ,$$

$$\tilde{m}(s) = N / \{ \Delta (\sqrt{s^2+i} + \sqrt{s^2-i}) \} \quad ,$$

$$(35)$$

where the numerators M and N are

$$M = -A\sqrt{s^4+1}(\nu s^2 + \sqrt{s^4+1}) + B\sqrt{s^4+1}(\sqrt{s^2+i} + \sqrt{s^2-i}) \quad ,$$

$$N = -A\sqrt{s^4+1}(\sqrt{s^2+i} + \sqrt{s^2-i}) + B(\nu s^2 + \sqrt{s^4+1}) \quad ,$$

$$(36)$$

and Δ is

$$\Delta = \sqrt{s^4+1} - \left(\frac{\nu s^2 + \sqrt{s^4+1}}{\sqrt{s^2+i} + \sqrt{s^2-i}} \right)^2 \quad ,$$

$$= \frac{(\sqrt{s^4+1} - \nu s^2)[\sqrt{s^4+1} + (2-\nu)s^2]}{(\sqrt{s^2+i} + \sqrt{s^2-i})^2} \quad .$$

$$(37)$$

Transforming the numerators M and N further, as shown below, we find the final solution in the following form:

$$\tilde{p}(s) = \frac{P}{2i} \frac{\sqrt{s^2+i} + \sqrt{s^2-i}}{[\sqrt{s^4+1} - \nu s^2][\sqrt{s^4+1} + (2-\nu)s^2]} [U e^{-y_0\sqrt{s^2+i}} - \bar{U} e^{-y_0\sqrt{s^2-i}}] \quad ,$$

(38)

$$\tilde{m}(s) = \frac{P}{2i} \frac{\sqrt{s^2+i} + \sqrt{s^2-i}}{[\sqrt{s^4+1} - \nu s^2][\sqrt{s^4+1} + (2-\nu)s^2]} [-V e^{-y_0\sqrt{s^2+i}} + \bar{V} e^{-y_0\sqrt{s^2-i}}] \quad ,$$

where

$$U = \sqrt{s^2-i} \left[\{(1-\nu)s^4[(1-\alpha) - \nu(1+\alpha)] - 2i\nu s^2 + (1-\alpha)\} - 2\sqrt{s^4+1}\{\alpha(1-\nu)s^2+i\} \right] \quad ,$$

(39)

$$V = \{(1-\nu)s^4[(1+\alpha) - \nu(1-\alpha)] + 2i\nu s^2 + (1+\alpha)\} + 2\sqrt{s^4+1}\{\alpha(1-\nu)s^2 + i\}$$

and $\bar{U}$ and $\bar{V}$ the conjugate complexes of U and V, respectively.

On substitution of A and B from (34), the numerator M becomes

$$M = \frac{P}{2i} [U e^{-y_0\sqrt{s^2+i}} - \bar{U} e^{-y_0\sqrt{s^2-i}}] \quad ,$$

(40)

where

$$U = -(1+\alpha)(\nu s^2 + \sqrt{s^4+1})E\sqrt{s^2-i} + (1-\alpha)\sqrt{s^4+1}(\sqrt{s^2+i} + \sqrt{s^2-i})\bar{E} \quad .$$

(41)

Substitution of M from (40) and Δ from (37), into the intermediate form $(35)_1$ yields the final form $(38)_1$. We transform U in (41) to

$$U = \sqrt{s^2-i}\{R + Q\sqrt{s^4+1}\} \quad ,$$

(42)

where

$$R = -(1+\alpha)\nu s^2 E + (1-\alpha)(s^2+i)\bar{E} \quad ,$$

(43)

$$Q = -(1+\alpha)E + (1-\alpha)\bar{E} \quad .$$

219

Substitution of E from (34)$_3$ and $\overline{E}$ from (34)$_4$ yields

$$R = (1-\nu)s^4[(1-\alpha) - \nu(1+\alpha)] - 2i\nu s^2 + (1-\alpha) ,$$

$$Q = -2[\alpha(1-\nu)s^2 + i] . \tag{44}$$

Substitution of R and Q from (44) into the expression of U in (42) yields the expression of U in (39)$_1$.

On substitution of A and B from (34), the numerator N becomes

$$N = \frac{P}{2i} [-Ve^{-y_0\sqrt{s^2+1}} + \overline{V}e^{-y_0\sqrt{s^2-i}}] , \tag{45}$$

where

$$V = (1+\alpha)(\sqrt{s^2+1} + \sqrt{s^2-i})E\sqrt{s^2-i} - (1-\alpha)(\nu s^2 + \sqrt{s^4+1})\overline{E} . \tag{46}$$

Substitution of N from (45) and Δ from (37)$_2$ into the intermediate form (35)$_2$ yields the final form (38)$_2$. We transform V in (46) to

$$V = T - Q\sqrt{s^4+1} , \tag{47}$$

where

$$T = (1+\alpha)E(s^2-i) - (1-\alpha)\overline{E}\nu s^2 . \tag{48}$$

Substitution of E from (34)$_3$ and $\overline{E}$ from (34)$_4$ into T in (48) yields

$$T = (1-\nu)s^4[(1+\alpha) - \nu(1-\alpha)] + 2i\nu s^2 + (1+\alpha) . \tag{49}$$

Substituting T from (49) and Q from (44)$_2$ into the expression of V in (47) yields the expression of V in (39)$_2$.

Substitution of $\tilde{p}(s)$ and $\tilde{m}(s)$ from (38) into (28) yields the solution for the free-edge condition. It is impossible also in this case to eliminate all the effects of the virtual reactions $p(x)$ and $m(x)$ by giving α an appropriate number. In other words, the reflection principle fails to yield a solution for the free-edge boundary condition.

Condition that must be assigned in the outside region.

Use of the Fourier transfor of generalized functions means dealing with the entire infinite plane. On the infinite plane that we are currently dealing with, however, a definite condition is given only in a part, say A, of our

interest and, except if any at infinity, no definite condition is given in the outside region, say B. The nonuniqueness emerges because use of the virtual reaction method, and the reflection principle as well, produces in fact uncontrolled deflection in the outside region B through the use of Wyman's solution (2), a solution with regard to the infinite plane. If the conditions are definite throughout the entire infinite plane, the solution must obviously be unique.

In physical terms, it is reasonable to assume that, even if a load is supported by a floating plate occupying a region A, the level of the water in the outside region B is invariable. On the other hand, even though a plate of our interest covers only a part of the entire infinite plane, use of Wyman's solution in the reflection principle and in the virtual reaction method reveals that there exists intrinsically an assumed plate covering the entire infinite plane. Taking this situation into consideration, we may interpret that the condition

$$w = 0 \tag{50}$$

in the outside region B of the assumed infinite plate replaces the condition that the level of water is constant in the region B of the physical infinite plane. The condition (50) implies that no virtual reactions must be imposed in the region B of the assumed infinite plate. Therefore, once we adopt the new condition, i.e., (50), we must disuse both the reflection principle and the virtual reaction method.

SUMMARY AND CONCLUSION

The deflection of a floating elastic plate has so far been analytically solved only for the entire infinite plane. With regard to the plates other than the infinite, solutions presumedly valid for the simple-edge boundary condition have been produced by the application of the reflection principle in which the above-mentioned analytical solution is used as an influence function.

We have introduced unknown virtual reactions that we can determine to satisfy given boundary conditions by using the solution with regard to the infinite plane as the influence function, and developed an analytical machinery that enables us to compute the deflection of a floating elastic plate expectedly valid to any shape under any boundary conditions. The mathematical method essentially consists in the use of the Fourier transform of generalized functions. The key to the new development is the discovery of the double Fourier transform of $kei\sqrt{x^2+y^2}$.

The boundary-value problems of the semi-infinite plate are studied with the virtual reaction method. It is found that the reflection principle yields a solution presumedly valid only for the simple-edge boundary condition. The virtual reaction method consistently yields a solution presumedly valid for every boundary condition. The alleged solutions, however, are not unique.

We have found the reason for the nonuniqueness: The influence function produces uncontrolled deflection in the outside of the semi-infinite plate, i.e., in the remainder of the entire infinite plane, which the Fourier transform of generalized functions necessarily includes in its range of application. The condition that must be imposed to produce a unique solution for a plate occupying

221

a region other than the entire infinite plane is that the deflection is zero everywhere outside the region that the plate occupies. Use of the influence function to obtain a solution in a region other than the entire infinite plane must be abandoned.

APPENDIX I

A BRIEF INTRODUCTION TO THE THEORY OF GENERALIZED FUNCTIONS

For the study of quantum mechanics, P.A.M. Dirac introduced in 1926 the delta function, $\delta(x)$. He defined the delta function to be equal to zero at $x \neq 0$, to be infinite at $x = 0$, and to satisfy

$$\int_{-\infty}^{\infty} \delta(x)dx = 1 \quad .$$

He also showed that, for any finite function $\phi(x)$, the relation

$$\int_{-\infty}^{\infty} \phi(x)\delta(x)dx = \phi(0) \tag{51}$$

holds true — the relation usually referred to (Hoskins (4), p. 35) as the sampling property of the delta function. The delta function has been proved to be a convenient mathematical tool but nevertheless had not been accepted by mathematicians as a valid mathematical concept until L. Schwartz clarified in 1945 the mathematical foundation underlying the delta function concept. Since Schwartz's work, the delta function is accepted by mathematicians as a straightforward mathematical concept.

To define the delta function in modern terms, it is the commonly accepted approach to start with the concept of functional. A functional is an operation through which a number is determined for a function. For instance, given a regular function $f(x)$, the integral

$$\int_{-\infty}^{\infty} \phi(x)f(x)dx$$

may determine a finite number for any function $\phi(x)$ if this integral is intergrable. Therefore the integral is a functional. It may be observed that a desire to give a functional an expression of integral implicitly prevails in the theory of generalized functions.

The modern concept of delta function is constructed not on Dirac's initial definition, which is a difficult mathematical concept, but on the sampling property (51), i.e., on the identification of the delta function with a functional that yields a number $\phi(0)$ for a function $\phi(x)$. This functional, however, cannot be expressed as an integral (in other words, (51) is deceptive) in the classical sense, because $\delta(x)$ in (51) is not an ordinary function. The modern approach overcomes this difficulty by taking the advantage of the following observation.

It is found that there are several sequences of regular functions whose limits exhibit the sampling property (see Gel'fand and Shilov (3), pp. 34-39). For instance, Fourier's single integral formula

$$\lim_{\nu \to \infty} \int_{-\infty}^{\infty} \frac{\sin(\nu x)}{x} f(x) dx = \pi f(0)$$

shows that an expression of the delta function is

$$\delta(x) = \lim_{\nu \to \infty} \frac{1}{\pi} \frac{\sin(\nu x)}{x} \quad . \tag{52}$$

The classical concept of function is, therefore, enlarged in the modern concept to the inclusion of the limits of the sequences of regular functions. The enlarged function concept is called the generalized functions. The delta function is a generalized function.

The algebraic process used in the above for defining generalized functions is the same as the one used for defining real numbers as the limits of the sequences of rational numbers. (Birkoff (1) is very readable on this subject.) In the same way as the real number is useful for computation, the generalized function is useful for analysis. For instance, consider a sequence f_1, f_2, ..., f_ν, ... of generalized functions of variables x_1, ..., x_j, ..., x_N, which converges to a generalized function f. Then, the theory of generalized functions (Gel'fand and Shilov (3), p. 29) asserts that the sequence of derivatives $\partial f_\nu / \partial x_j$ converges to $\partial f / \partial x_j$ — a property that does not necessarily hold true in the case of ordinary functions.

The expression (4) of the delta function as the Fourier transform of unity is derived when (52) is rewritten to

$$\delta(x) = \frac{1}{2\pi} \lim_{\nu \to \infty} \int_{-\nu}^{\nu} e^{i\xi x} d\xi \quad ,$$

which is equivalent to (4).

Use of the Fourier transform (4) of unity and the sampling property (51) of the delta function is the essence of the Fourier reciprocal relationship. To demonstrate this, let us assume the relationship

$$\int_{-\infty}^{\infty} f(t) e^{i\xi t} dt = g(\xi) \quad . \tag{53}$$

Multiplying this by $\exp(-i\eta\xi)$, integrating with regard to ξ, and exchanging the order of integrations, the above becomes

223

$$\int_{-\infty}^{\infty} f(t) dt \int_{-\infty}^{\infty} e^{i\xi(t-\eta)} d\xi = \int_{-\infty}^{\infty} g(\xi) e^{-i\eta\xi} d\xi \quad . \tag{54}$$

(The order of integrations may be exchanged, because we may assume that the double integral exists in the sense of generalized functions.) Using the formula (4) of the delta function, the left-hand side of (54) reduces to

$$2\pi \int_{-\infty}^{\infty} f(t)\delta(t-\eta) dt \quad ,$$

which becomes $2\pi f(\eta)$ by use of the formula (51) of the sampling property of the delta function. Eq (54) thus reduces to

$$\int_{-\infty}^{\infty} g(\xi) e^{-i\eta\xi} d\xi = 2\pi f(\eta) \quad , \tag{55}$$

which is the inverse of (53).

The brief introduction to generalized functions in this appendix is sufficient for a casual reading for comprehending the outline of this paper. For understanding the details of the theory of generalized functions, which is needed for a further development of the mathematics in this paper, it should be noted that currently three different approaches to the theory of generalized functions are available to applied mathematicians.

The most widely accepted approach is Gel'fand and Shilov's (3), which is based, however, on the modern concept of functional space and contains the materials that are actually more than we need. The easiest approach is presented by Hoskins (4), but it deals with only the case of a single independent variable, and ends up eventually in the functional space approach and in the use of theory of integrations. The brief introduction in this appendix is a digest of an essential part of the approach initiated by G. Temple in 1955, formalized by Lighthill (8), and extended by Jones (5). Although not widely recognized, this approach seems to be a natural one to the theory of generalized functions.

APPENDIX II

FOURIER TRANSFORMS ON THE BOUNDARIES OF THE SEMI-INFINITE PLATE

For the convenience of calculating the Fourier transforms of the boundary conditions (16), (17), (18), the Fourier transforms of the expressions contained in them are written out by use of (25), as shown below:

$$\int_{-\infty}^{\infty} e^{isx} w(x,+0)dx$$

$$= -\frac{P(1+\alpha)}{4i}\{\frac{1}{\sqrt{s^2+i}} e^{-y_o\sqrt{s^2+i}} - \frac{1}{\sqrt{s^2-i}} e^{-y_o\sqrt{s^2-i}}\} - \frac{\tilde{p}(s)}{2\sqrt{s^4+1}(\sqrt{s^2+i} + \sqrt{s^2-i})} ,$$

$$(56)$$

$$\int_{-\infty}^{\infty} e^{isx} \frac{\partial w(x,+0)}{\partial y}dx$$

$$= -\frac{P(1-\alpha)}{4i}\{e^{-y_o\sqrt{s^2+i}} - e^{-y_o\sqrt{s^2-i}}\} + \frac{\tilde{m}(s)}{(2\sqrt{s^2+i} + \sqrt{s^2-i})} ,$$

$$(57)$$

$$\int_{-\infty}^{\infty} e^{isx} \frac{\partial^2 w(x,+0)}{\partial y^2}dx$$

$$= -\frac{P(1+\alpha)}{4i}\{\sqrt{s^2+i} e^{-y_o\sqrt{s^2+i}} - \sqrt{s^2-i} e^{-y_o\sqrt{s^2-i}}\} +$$

$$(58)$$

$$+ \frac{\tilde{p}(s)}{2(\sqrt{s^2+i} + \sqrt{s^2-i})} - \frac{1}{2} \tilde{m}(s) ,$$

$$\int_{-\infty}^{\infty} e^{isx} \frac{\partial^2 w(x,+0)}{\partial x^2}dx$$

$$= \frac{P(1+\alpha)}{4i}\{\frac{s^2}{\sqrt{s^2+i}} e^{-y_o\sqrt{s^2+i}} - \frac{s^2}{\sqrt{s^2-i}} e^{-y_o\sqrt{s^2-i}}\} +$$

$$(59)$$

$$+ \frac{\tilde{p}(s)s^2}{2\sqrt{s^4+1}(\sqrt{s^2+i} + \sqrt{s^2-i})} ,$$

225

$$\int_{-\infty}^{\infty} e^{isx} \frac{\partial^3 w(x,+0)}{\partial y^3} dx$$

$$= -\frac{P(1-\alpha)}{4i}\{(s^2+i)e^{-y_0\sqrt{s^2+i}} - (s^2-i)e^{-y_0\sqrt{s^2-i}}\} -$$

$$-\frac{1}{2}\tilde{p}(s) + \frac{\tilde{m}(s)[2s^2 + \sqrt{s^4+1}]}{2(\sqrt{s^2+i} + \sqrt{s^2-i})},$$

(60)

$$\int_{-\infty}^{\infty} e^{isx} \frac{\partial^3 w(x,+0)}{\partial x^2 \partial y} dx$$

(61)

$$= +\frac{P(1-\alpha)}{4i}s^2\{e^{-y_0\sqrt{s^2+i}} - e^{-y_0\sqrt{s^2-i}}\} - \frac{\tilde{m}(s)s^2}{2(\sqrt{s^2+i} + \sqrt{s^2-i})}.$$

The fractions in these formulas are expressed in the forms given on the right-hand side of (26).

REFERENCES

1. Birkhoff, G., _A Survey of Modern Algebra_, Revised ed., The MacMillan Co., New York, 1953.

2. Erdélyi, A., et al., ed., _Tables of Integral Transforms_, Vol. 1, McGraw-Hill, New York, p. 56, No. (43).

3. Gel'fand, I.M., and Shilov, G.E., _Generalized Functions_, Vol. 1: Properties and Operations, translated by E. Saletan, Academic Press, New York, 1979.

4. Hoskins, R.F., _Generalized Functions_, John Wiley and Sons, New York, 1979.

5. Jones, D.S., *Generalized Functions*, McGraw-Hill Publishing Company Ltd., New York, 1966.

6. Kerr, A.D., "Elastic Plates on a Liquid Foundation," *Journal of the Engineering Mechanics Division*, Vol. 89, No. EM3, June 1963, pp. 59-71.

7. Kerr, A.D., "An Indirect Method for Evaluating Certain Infinite Integrals," *Journal of Applied Mathematics and Physics* (JAMP), Vol. 29, 1977, pp. 380-386.

8. Lighthill, M.J., *Introduction to Fourier Analysis and Generalized Functions*, Cambridge at the University Press, 1958.

9. Livesley, R.,.., "Some Notes on the Mathematical Theory of a Loaded Elastic Plate Resting on an Elastic Foundation," *Quaternary Journal of Mechanics and Applied Mathematics*, Vol. 6, 1953, Pt. 1, pp. 32-44.

10. Olver, F.W.J., "Bessel Functions of Integer Order," *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and I.A. Stegun, ed., National Bureau of Standards Applied Mathematics Series 55, Revised ed., U.S. Government Printing Office, Washington, 1970, p. 361.

11. Takagi, S. "Some Bessel Function Identities Arising in Ice Mechanics Problems," *CRREL Report 79-27*, Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire, 1979.

12. Takagi, S., "The Hankel-Fourier Reciprocal Relationship Arising in Ice Mechanics Problems," *Transactions of the Twenty-Fifth Conference of Army Mathematicians*, pp. 625-641, *ARO Report 80-1*, U.S. Army Research Office, P. . Box 12211, Research Triangle Park, North Carolina, 1980.

13. Watson, G.N., *A Treatise on the Theory of Bessel Functions*, Cambridge at the University Press, Cambridge, 1962.

14. Wyman, M., "Deflection of an Infinite Plate," *Canadian Journal of Research*, A28, pp. 293-302, 1950.

CYCLIC STRESS-STRAIN BEHAVIOR NEAR A NARROW ELLIPTICAL FLAW

Dennis M. Tracey and Colin E. Freese

Mechanics and Engineering Laboratory
Army Materials and Mechanics Research Center
Watertown, Massachusetts

ABSTRACT. The elastic-plastic, plane strain conditions at the ends of a narrow, elliptical flaw are examined for the case of a zero-tension, cyclic loading. The flaw considered has an aspect ratio of $\sqrt{1000}$ which corresponds to a length to tip radius ratio of 2000. Hence, the flaw serves as a model for a crack. A specialized numerical formulation was developed for solution of this problem. It involves aspects of both the finite element and mapping-boundary collocation methods. Attention is restricted to loadings which maintain the zone of plastic deformation close to the flaw ends. Numerical results demonstrate the forms taken by the stress and strain distributions within the plastic region, during a cyclic load pattern.

INTRODUCTION. We focus here on the analytical problem of determining the elastic-plastic, stress-strain states that develop ahead of a blunt-tipped flaw during cyclic loading. This is a basic problem which is encountered in studies aimed at developing improved predictive criteria for macrocrack growth in structural materials. For certain limited conditions, linear elastic fracture mechanics provides a satisfactory basis for crack growth prediction. Using results of crack growth in a material specimen, a crack growth "law" is established which predicts the growth rate da/dn in terms of the variation ΔK of the elastic, crack tip, stress singularity amplitude. When this linear elastic approach fails to give satisfactory predictions, the elastic-plastic analytical problems arise. Such is true, for instance, in

studies attempting to explain the phenomenon of crack growth retardation and the general crack behavior in variable amplitude load cycling. It is expected that plasticity effects, identified through analyses such as that presented here, will provide a framework for devising improved growth criteria. Of course, the nonlinear character of the elastic-plastic problem makes the task of identifying generally meaningful "plasticity effects" a very complex one, indeed. The work reported here hopefully represents a step toward resolution of this broad problem.

We limit attention here to load conditions which produce localized plasticity, i.e., a plastic zone which extends no further than a few tip radii from the flaw ends. Such is the case when the applied tension is restricted to small magnitudes relative to the material's yield stress. Our treatment of an elliptical flaw contrasts with the usual choice of a sharp crack model in fracture analysis. The large value 2000 for the length to tip radius ratio $2a/\rho$ appears to offer a suitable simulation of a natural macrocrack. Although the elliptical shape was chosen for reasons of analytical convenience, it has the blunt tip characteristic that we wished to include in this study. It is clear that sharp and blunt crack solutions will display major differences under conditions of localized yielding. We are interested in quantifying these differences and have conducted some work along these lines, although we reserve this particular subject for a subsequent report. We limit ourselves here to the discussion of the elliptical flaw analysis.

The flaw is considered to be isolated within an infinite domain. Remote uniaxial tension is imposed in a direction perpendicular to the flaw's length and plane strain constraint is specified. The continuum is

230

modeled as a non-hardening Prandtl-Reuss material. Hence, the Mises yield criterion applies and plastic flow occurs under a constant value of equivalent stress equal to the yield stress Y. The stress-strain relationships have a rate form, so an incremental analysis following the cyclic load path is necessary.

The numerical formulation was designed specifically for this problem, from aspects of the finite element and stress function boundary collocation methods. The problem is doubly symmetric, so that we consider one quadrant of the geometry. In a region surrounding the flaw end, where plastic deformation is anticipated, finite element approximations are made for the incremental displacement field. This region extends a distance of $4c$ ahead of the flaw, as illustrated in Figure (1). Only the mesh of the first quadrant is in fact used. The contour Γ defines the interface with the elastic region which is represented by a complex variable power series, stress function approximation. Conformal mapping was used to transform the flaw onto a unit circle, and this allowed utilization of analytic continuation principles for satisfaction of the traction free flaw boundary condition. In the transformation, Γ maps to a circular arc centered at the end of the flaw. A form of boundary collocation along Γ is employed, involving the unknown coefficients of the power series. Conditions of equilibrium and compatibility are enforced across Γ, and these provide equations relating nodal displacement changes to the change in applied tension.

This hybrid type of formulation was chosen both for reasons of mathematical accuracy and efficiency of analysis. It is well established that the mapping-stress function collocation approach is an effective way to solve elasticity problems, especially for infinite domains, Bowie [1]. Likewise, the suitability of the elastic-plastic finite element formulation has been amply demonstrated, Tracey and Freese [2].

231

The elastic solution to our problem is known, and it suggests that the flaw tip stress σ_{yy} exceeds the remote tension T by a factor of 64.2, i.e., $1 + 2\sqrt{a/\rho}$. In the tip vicinity, the stress gradient is severe: σ_{yy} falls by a factor of 5 from the flaw surface to the interface Γ. The mesh shown in the figure has element edges with lengths of the order of $\rho/10$. This level of spatial discretization was found to be suitable, based upon the numerical elastic solution. We found the σ_{yy} distribution to be well within one percent of exact, when element midpoint data was considered. Before discussing the numerical solution, we next outline the equations employed in the analysis.

FORMULATION. We refer the reader to Bowie [1] and Tracey and Freese [2] for complete descriptions of the mapping-collocation and finite element formulations which formed the basis of this work. Here, we begin our discussion with the equations which govern in the elastic region beyond the interface. The point of departure is the expression for the Airy stress function $W(x,y)$ in terms of the two analytic functions of the complex variable $z = x + iy$, $\phi(z)$ and $\psi(z)$:

$$W = \mathrm{Re}\ (\overline{z}\ \phi + \int \psi\ dz) \tag{1}$$

The field quantities are expressible in terms of ϕ and ψ. For instance, the stress increment $\Delta\sigma_{xx}$, displacement increment Δu_y and the x,y components of the force resultant acting on the generic arc AB are given by

$$\Delta\sigma_{xx} = \mathrm{Re}\ (-\overline{z}\ \phi'' + 2\phi' - \psi') \tag{2}$$

$$\Delta u_y = \mathrm{Im}\ ((1+\nu)\ \overline{z}\ \phi' - (3-\nu)\ \overline{\phi} + (1+\nu)\ \psi)/E \tag{3}$$

$$\Delta F_x + i\ \Delta F_y = -i\ [z\ \overline{\phi}' + \phi + \overline{\psi}]_A^B \tag{4}$$

In equation (3), E and ν are the elastic material constants. In equation (4), the bracket notation indicates the change in going from position A to B. The barred quantities represent complex conjugates, and the primes indicate differentiation.

The mapping function transforms the elliptical flaw in the z-plane onto the unit circle $|\zeta| \le 1$ in an auxiliary ζ-plane, according to

$$z = (a+b)\ \zeta/2 + (a-b)/(2\zeta) = \omega(\zeta) \qquad (5)$$

where a and b are the semi-axes of the flaw. The geometric detail of our problem in the ζ plane is illustrated in Figure (2). As can be seen, the interface Γ appears as a circular arc centered at $\zeta = 1$. The finite element grid was designed in this plane where it takes a uniform polar character, centered at the origin.

The boundary conditions that must be satisfied can be summarized as follows:

I
$$\left. \begin{array}{l} \Delta\sigma_{yy} \rightarrow \Delta T \\[4pt] \Delta\sigma_{xx} \rightarrow 0 \\[4pt] \Delta\sigma_{xy} \rightarrow 0 \end{array} \right\} \qquad \text{as } z \rightarrow \infty$$

II   flaw surface is traction free

III $\Delta u_x = \Delta\sigma_{xy} = 0$ along $x = 0$

            $\Delta u_y = \Delta\sigma_{xy} = 0$ along $y = 0$

IV $\Delta u_x, \Delta u_y, \Delta\sigma_{xx}, \Delta\sigma_{xy}, \Delta\sigma_{yy}$ continuous across Γ

The most significant advantage of mapping the flaw onto the unit circle is that the traction free flaw condition can be met by the method of analytic continuation. It follows from equation (4) that if ϕ is defined in the interior of the flaw, $|\zeta| < 1$, according to

$$\psi(\zeta) = -\omega(\zeta)\ \overline{\phi'(1/\bar\zeta)}/\overline{\omega'(1/\bar\zeta)} - \overline{\phi(1/\bar\zeta)} \tag{6}$$

or equivalently, in $|\zeta| \geq 1$, if ψ is given in terms of ϕ by

$$\psi(\zeta) = -\overline{\phi(1/\bar\zeta)} - \overline{\omega(1/\bar\zeta)}\ \phi'(\zeta)/\omega'(\zeta) \tag{7}$$

then the traction free conditions are implicitly satisfied. With the definition (7), the problem resorts to finding the single analytic function ϕ which satisfies conditions I, II, and IV.

The remote stress conditions I and symmetry conditions III require that

$$\phi \to \Delta T\ (a+b)\zeta/8 \qquad \text{as } \zeta \to \infty \tag{8}$$

$$
\left.
\begin{array}{ll}
\mathrm{Re}\ \phi = 0 & \text{along } x = 0 \\[6pt]
\mathrm{Im}\ \phi = 0 & \text{along } y = 0
\end{array}
\right\} \tag{9}
$$

These conditions are satisfied, of course, by the exact stress function for the elastic problem:

$$\phi^{\text{elastic}} = \Delta T\ [(a+b)\zeta - (3a+b)/\zeta]/8 \tag{10}$$

The representation chosen for the elastic-plastic problem consists of a finite term power series added to the elastic function:

$$\phi = \phi^{\text{elastic}} + \sum_{n=1}^{15} a_n\ \zeta/(\zeta^2-1)^n \tag{11}$$

The plastic deformation within Γ will locally perturb the solution (10). For this reason, the series was chosen to have negative powers of (ζ^2-1). The coefficients a_n are real to satisfy the conditions of symmetry. These unknowns are determined in conjunction with the finite element unknowns, as discussed next.

Bilinear, isoparametric finite elements were used. There were a total of 314 nodes, 27 of which were positioned on Γ. The nodal displace-

ment increments Δu_x, Δu_y for the interface nodes are represented by the vector array $\underline{\Delta U}^I$, while the remaining unknowns interior to I are denoted $\underline{\Delta U}^b$. The flaw surface and x-axis boundary conditions were treated routinely. The interface was considered as a surface with a variable traction distribution; variable in the sense that it is expressed in terms of the unknowns α_n, using equations such as (2) along with (11). The standard consistent load procedure was employed to define nodal loads from the traction distribution. This involves a numerical integration (2 point Gauss rule) over each element edge that lies on the interface. If we denote the vector array of the load components for nodes on I as $\underline{\Delta P}^I$ and if the coefficients are arranged in the vector array $\underline{\alpha}$, then

$$\underline{\Delta P}^I = \underline{c}\,\underline{\alpha} + \Delta T\,\underline{d} \tag{12}$$

where $\underline{c}$ is a rectangular array and $\underline{d}$ is a vector array of constants. For our problem, there were 53 nodal unknowns along I, so that $\underline{c}$ was of order 53 x 15.

The principle of virtual work provides the relationships between the nodal unknowns and $\underline{\Delta P}^I$. Using standard notation, the equations take the form

$$\begin{bmatrix} \underline{K}^{bb} & \underline{K}^{bI} \\ \underline{K}^{Ib} & \underline{K}^{II} \end{bmatrix} \begin{Bmatrix} \underline{\Delta U}^b \\ \underline{\Delta U}^I \end{Bmatrix} = \begin{Bmatrix} \underline{0} \\ \underline{\Delta P}^I \end{Bmatrix} \tag{13}$$

The continuity condition across I is satisfied by equating the components of $\underline{\Delta U}^I$ to the expressions involving ΔT and α_n suggested by the stress function and the equations for the displacement increment, equation (3). This results in a matrix equation of the form

$$\underline{\Delta U}^I = \underline{g}\,\underline{\alpha} + \Delta T\,\underline{h} \tag{14}$$

235

where g is rectangular and h is a vector array.

We arrive at a system of equations for a by eliminating ΔU^F from equation (13) and using equations (12) and (14). The resultant system takes the form

$$(k^* g - c) a = \Delta T (d - k^* h) \tag{15}$$

The matrix k^* is square and it is numerically established by a partial Gaussian elimination algorithm. Formally, it is given by

$$k^* = k^{TT} - k^{TF} (k^{FF})^{-1} k^{FT} \tag{16}$$

Equation (15) represents a system of 53 equations in the 15 coefficients a_n. A least square procedure was used to solve this system. With the coefficients determined, the nodal displacement increments, strain increments and finally stress increments are computed. We refer the reader to our earlier report [2] which describes the adaptive load incrementation algorithm which was utilized. It selects the magnitude of ΔT at each step of the load path according to a prescribed allowable, yield surface, deviatoric stress change.

NUMERICAL RESULTS. We considered a load spectrum which had the remote tension varying between values of 0 and 0.089 Y. The maximum value of 0.089 Y corresponds to the point in the initial monotonic loading when the plastic zone had extended to within one element of Γ. The spectrum was followed for 1 1/2 cycles, requiring a total of 68 load steps with the adaptive incrementation algorithm restricting the yield surface stress change at each step to $O(Y/20)$.

The nature of the stress variations which occur ahead of the flaw can be explained in terms of the elastic-plastic boundary movement. In

Figure (3), elastic plastic boundaries are drawn for key positions in the load spectrum. The spectrum is represented by the loading segment OAB, unloading segment BCD and reloading segment DEFG. States A, C, and E are approximately at one-half peak load. There is the expected gradual expansion of the plastic zone during loading, as indicated by the contours for states A and B. Upon load reversal, there is an interval of purely elastic response, corresponding to elastic unloading at all plastically deformed points, and this is followed by reverse yielding ($\sigma_{yy} < 0$) and an expansion of the plastic zone until the load minimum is reached. This behavior is indicated by the boundaries labeled C and D. With load increase from the zero load state D, elastic unloading occurs ($\Delta\sigma_{yy} < 0$), followed by an expanding zone of renewed forward plastic flow, as indicated by the boundaries labeled E, F, and G. There are two intriguing aspects of these results. First, the solution has a periodic character: the boundaries for peak load states B and G are nearly identical, and the same is true for the boundaries of states C and E. Secondly, a dramatic change in the plastic zone occurred during the last load step FG (ΔT had the small value of 0.0022 Y), as indicated by the outer two boundaries in the figure. During FG, the stress states of all elements beyond the middle boundary reached the yield surface - but no significant plastic flow occurred. Plastic straining was limited to the zone within the boundary labeled $D \approx F \approx A$, and thus this defines the cyclic plastic zone of our problem. As suggested by the labeling, there were minor differences in the elastic-plastic boundaries for states D, F, and A. Rice's [3] plastic superposition analysis predicts that the cyclic plastic zone should correspond to the plastic zone at the intermediate load state A. That analysis also predicts the periodic solution behavior that we observed. Noteworthy, however, is the fact that while the superposition

analysis is based upon assumptions of proportional plastic flow throughout the load history, we found what appears to constitute significant non-proportionality. Of course, these results and comparisons can serve to define just what is significant in this regard.

The stress distribution (σ_{yy}/σ_0 vs. x/c) is given in Figure (4) at the load positions A - G. The vertical divisions on the plot mark off the locations of the cyclic plastic boundary, the monotonic plastic boundary and the modeling interface Γ. As a first observation, curves A and B demonstrate the fact that the maximum stress value increases and the location at which the maximum occurs changes as the load level is increased. Curve C illustrates the compressive stress state that develops upon load reversal. There is a single element experiencing plastic flow at load position C. The reverse yield zone spreads and the hydrostatic compression increases as the load decreases, as shown by curve D. Upon reloading, the material at the flaw surface regains a tensile stress state and a zone of forward yield is developed, while the compressive stress field is overcome. Curve E illustrates how the form of the stress distribution (inflection points) is defined by the current, cyclic and monotonic plastic zones. The dip in the curve F shows the last evidence of the compressive hydro-static history. Consistent with the elastic-plastic boundary results of Figure (3), we see that the σ_{yy} stress distribution at G very closely agrees with that found at load state B. There are small differences between curves B and G that are apparent within the cyclic plastic zone.

The variation of the stress-strain state (σ_{yy}, ϵ_{yy}) during the load cycling is displayed in Figure (5) for six locations ahead of the flaw. The stress and strain values are normalized by the yield stress and yield strain, respectively. The spatial positions $(x-a)/c = 0.057, \cdots, 1.750$
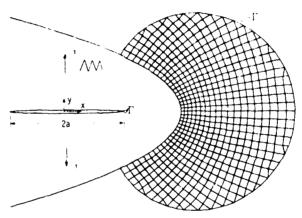
238

correspond to the centers of elements 1, 3, 5, 7, 9, and 11. Only the
first three of this group of elements are in the cyclic plastic zone, as
evidenced by the open loops between load states B, D, and G. It can be
seen that there are significant differences, from point to point, in stress
range, strain range, mean stress, and mean strain. Interestingly, the
mean stress throughout the cyclic zone is essentially zero. From the plot,
we see that each stress-strain history, normalized as it is by the yield
values, starts with a slope close to unity, and then at a stress level
above yield - the level increasing with distance from the flaw surface -
the slope drops drastically with a resultant large increase in strain with
further loading to peak load B. Upon load reversal, the stress and strain
decrease according to the point's initial elastic slope, until reverse
yield in the case of the first three points or attainment of the load free
state in the case of the other points. Although the strain decreases during
the unloading and there is a region that experiences compressive stress, the
strain remains tensile. Consistent with the behavior noted in Figure (4),
we see that the stress-strain state, after reloading to peak load, very
nearly coincides with that of state B. As we have mentioned, this type of
periodic stress-strain behavior is predictable using the assumption of
proportional plastic flow. The lack of proportionality in the solution can
be seen from Figure (6) which is a τ-plane plot of the $(\sigma_{xx}, \sigma_{yy}, \sigma_{zz})$
stress history of a point within element 1. Proportional flow would
require that the stress point not venture from a diameter of the yield
surface, yet there is a $32°$ variation on each side of the circle. Nonethe-
less, we see that states B, D, and G very nearly fall on a diameter. These
stress states are approximately those of fully plastic, biaxial, plane
strain extension, where $\sigma_{yy} = 2 \sigma_{zz} = \pm 2 Y/\sqrt{3}$. The development of these

fully plastic states at the flaw surface is perhaps the key to the realization of periodicity of the solution ahead of the flaw.
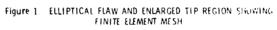
CONCLUSIONS. We commented earlier on the generic relevance of this elastic-plastic solution to crack growth criteria studies. There is little in the literature on the issue that was our primary concern here: namely, the very early stages of crack tip deformation when elastic strains are significant and plastic zone extent is very small in comparison to crack length. This lack of information is true not just for cyclic loadings, but for monotonic loadings as well. Of course, research needs are hardly limited to this early stage of elastic-plastic deformation and to the particular load cycle considered. Future work will consider the near tip field, as predicted from the blunt and sharp crack models, which develops at higher load levels and variable amplitude cycling.

REFERENCES.
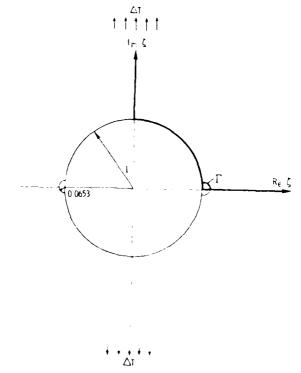
1. O. L. Bowie, "Solutions of Plane Crack Problems By Mapping Technique," in Mechanics of Fracture, v. 1, G. C. Sih, ed., Noordhoff, Leyden, 1973.

2. D. M. Tracey and C. E. Freese," Adaptive Load Incrementation in Elastic-Plastic Finite Element Analysis," to appear J. Computers and Structures, 1980.

3. J. R. Rice, "Mechanics of Crack Tip Deformation and Extension by Fatigue," in Fatigue Crack Propagation, ASTM STP 415, 1967.

Figure 1. ELLIPTICAL FLAW AND ENLARGED TIP REGION SHOWING
FINITE ELEMENT MESH

241
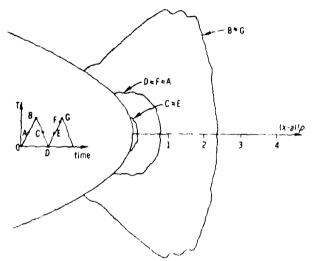
Figure 2 PROBLEM IN AUXILIARY PLANE

Figure 3. ELASTIC-PLASTIC BOUNDARIES FOLLOWING LOAD CYCLE
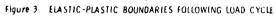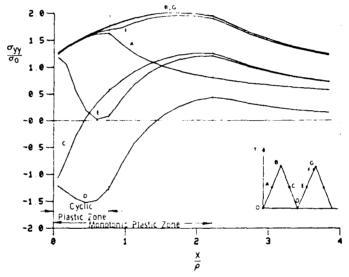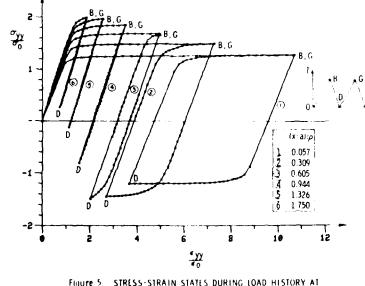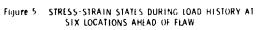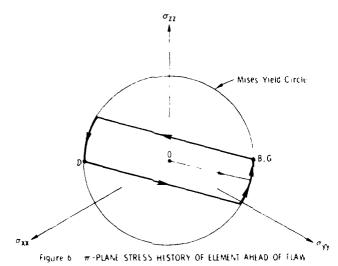
Figure 4. STRESS DISTRIBUTION AHEAD OF FLAW AT LOAD STATES A-G

Figure 5. STRESS-STRAIN STATES DURING LOAD HISTORY AT
SIX LOCATIONS AHEAD OF FLAW

Figure 6 π-PLANE STRESS HISTORY OF ELEMENT AHEAD OF FLAW

ELASTIC-PLASTIC ANALYSIS OF SCREW THREADS

G. P. O'Hara
U.S. Army Armament Research and Development Command
Large Caliber Weapon Systems Laboratory
Benet Weapons Laboratory
Watervliet, NY 12189

ABSTRACT. An elastic-plastic analysis method is suggested for screw thread teeth. In this method a single tooth is analyzed using boundary conditions to simulate a long chain of identical teeth. A set of five different loads are suggested to simulate pressure and shear on each flank along with a general stress field in the component. An example is worked out for a British Standard Buttress thread form. Data is presented from the example to show that friction is a very important parameter.

I. INTRODUCTION. The problem of stress concentrations in screw threads has long been obscured by the larger number of parameters involved and the lack of a systematic approach which could help to explain the variation that is in any experimental program. The object of this work is to try to cut through those problems and try to present a useful, organized approach which can encourage more work in this general area.

An example of the large number of parameters is the geometry description shown in Figure 1. While these dimensions may be of use to the designer to insure that the component will fit together, the stress analyst needs only a few of them. The major geometry parameters are the primary flank angle (α), secondary flank angle (β), and root radius (R). The primary loading parameters are the applied load (W), its angle (γ), and position (b). These last three parameters follow the convention of Heywood [1]. A further simplification is to nondimensionalize all linear dimensions to the pitch (P).

The very high performance requirements of military hardware have in the past produced a new thread form [2,3] for use on cannon breech components. During the development of the Watervliet Buttress thread, the Heywood equation [1] was used to choose geometry parameters for testing with good success. The Heywood stresses, however, were never correlated with test results. The reason for this was pointed out by this author [4] and it is simply that Heywood isolated his teeth so that only effects due to the load on the tooth would be present. In most experiments, the stress in the fillet of a thread is the result of load on the thread plus a stress concentration of the general stress field in the component.

In a recent paper [5] this author offered an elastic stress concentration approach to screw threads. In this work the overall loading on a thread is resolved into two forces parallel and normal to the pitch line. These are divided by the area on the pitch surface to produce two average stresses, radial stress and shear transfer rate. Of these, shear transfer rate is used to normalize all stresses, and the radial stress is used in a plot with the maximum fillet stress to produce a curve which is a characteristic of a particular thread form. This curve is usually generated as the coefficient of friction is varied from -1.0 to 1.0. where the sign denotes the direction of the friction vector, positive being away from the fillet. This sign convention gives the radial stress the same sign convention as all other stresses with tension positive. With this method an axial body stress in terms of a uniform remote tension can be easily added to produce a family of characteristic curves.

The above work is all elastic and certainly only looks at less than half of the overall problem. Elastic-plastic analysis adds a new set of problems to the analysis. First is that it is possible to identify five different plastic zones in a single tooth (Fig. 2), the axial stress zone, the Heywood zone, the secondary flank zone, the shear failure zone, and the bearing failure zone. It is difficult to imagine a problem in which only one of these is present and the usual case is where plasticity involves three or more of those zones working together with each starting at its own load.

The major factor that complicates elastic plastic analysis is that it is directly linked with the material stress-strain curve, and a general solution can be found only for materials with similarly shaped curves. For this report the assumed material will be 7075-T6 aluminum (Fig. 3) [6] with a proportional limit of 65 Ksi and 0.2 percent offset yielded of 72 Ksi. This is an engineering stress-strain curve defined out to 6% strain.

II. ELASTIC PLASTIC METHOD. The NASTRAN Rigid Format 6, Piecewise Linear Analysis is covered in the Theoretical Manual [7] and uses the triangular ring element (CTRIARG) which was implemented in a parallel program with the trapezoidal element reported by Chen [8]. In this method the number and size of the linear steps is selected by the user before the run. It is the duty of the user to select steps which produce adequate results within the limitations of the available computer time. The program then selects the slope off the stress-strain curve by extrapolating the change in effective strain for the current load step out to the end of the next load step and using an estimated elastic modulus (E)

$$E_i = \frac{\sigma_{i+1} - \sigma_i}{e_{i+1} - e_i} \tag{1}$$

Where σ_{i+1} and e_{i+1} are estimated values. This will be equal to the slope of one of the liner segments of the input stress-strain curve only when both points fall within the same liner segment of the curve. In the case of a zero modulus the element is assumed to have no increase in stiffness and a zero element stiffness matrix is generated.

The use of the stepped constraint input is not normally allowed because of the ambiguity that would exist if both forces and constraints were stepped together. This can be overcome using a small DMAP alter package in the executive control deck when only constraint input is to be used. Under these conditions it would seen that superior results could be expected because the extrapolation is done on the basis of strain.

The solutions in this report have been set up on the basis of 13 load steps, however, in one case the solution was truncated when a portion of the structure exceeded the defined stress strain curve and entered the zero slope region. When this has happened to all elements connecting any grid point a singular body stiffness results and the solution is stopped. This results when the modulus (E) is zero and the element stiffness matrix becomes zero.

III. BOUNDARY CONDITIONS. In this work a small finite element grid (Fig. 4) will be used to similate the behavior of a long chain of identical threads. This requires boundary conditions for the three surfaces where the model is cut out of the larger problem as well as applied loads. These surfaces are the two radial planes and an axial cylinder. These surfaces will be treated differently for axial load and the Heywood loads on the thread bearing surfaces.

The grid points on the axial cylinder must be constrained to replace the bulk of the body material. For the axial stress input these points are free in the axial direction and are constrained to a fixed displacement in the radial direction. This radial displacement accounts for Poisson contraction in the body. The grid points in the radial planes are generated at the same radial locations to allow them to be constrained in pairs, one in each radial plane. The radial displacement of each point of a pair is equal and the relative axial displacement of all pairs is the same. This forces the radial plane to conform to the same deformed shape while being free to distort out of the planes. In the elastic-plastic solution for axial stresses the constraint values for Poisson's constraint and relative elongation are stepped together to produce a piecewise constraint input condition.

In the solution for Heywood loads the object is to react the load out in shear, therefore the grid points on the axial cylinder are given a zero displacement in the axial direction. This zero displacement is also given to the radial displacements to similate a stiff structure. The two radial planes retain the same constraints as for the axial loads, however, the relative axial displacement is set to zero. In this case forces on the bearing surface are stepped to produce the piecewise loads.

Figure 5 shows the forces on a particular thread tooth. There is a primary and a secondary bearing flank with a pressure and a shear load on each. The primary flank is the one which is intended for force transfer. The secondary flank becomes loaded under reverse loading or when displacement removes the radial clearance. In this paper only uniform loads on the primary flank will be used. In this case the pressure and shear loads are added into an overall load W which is then resolved into a radial load (L_R) and axial load (L_a). These are the loads which are averaged over the area at the pitch line to produce the radial stress (σ_r) and the shear transfer rate (τ_R).

IV. EXAMPLES OF ELASTIC-PLASTIC ANALYSIS. In this paper four examples of elastic-plastic analysis will be shown for the thread form used, the British Standard Buttress. This form appears as a high strength thread in several Army structures such as the M68 cannon breech and some kinetic energy armor piercing projectiles where it seems to have been selected because of the low radial load component. The loadings are all uniform applied loads and include one axial stress load and three Heywood type loads. The finite element grid is shown in Figure 5 with the element shrunken to expose each side. This grid has a pitch diameter of 10.0 times the pitch length.

The first load is an axial stress in the body of the component with a peak of 65 Ksi. This is done by constraining the relative axial displacement of the radial planes to a fixed value and stepping that value in the piecewise solution. The axial cylinder is stepped in a similar way to produce the Poisson contraction. Figure 6 shows a shrunken element plot of those elements which have become nonlinear. In this plot all the elements shown are above the proportional limit stress of 65 Ksi. The elements shown doubled are above the conventional yield stress at the .2% offset point.
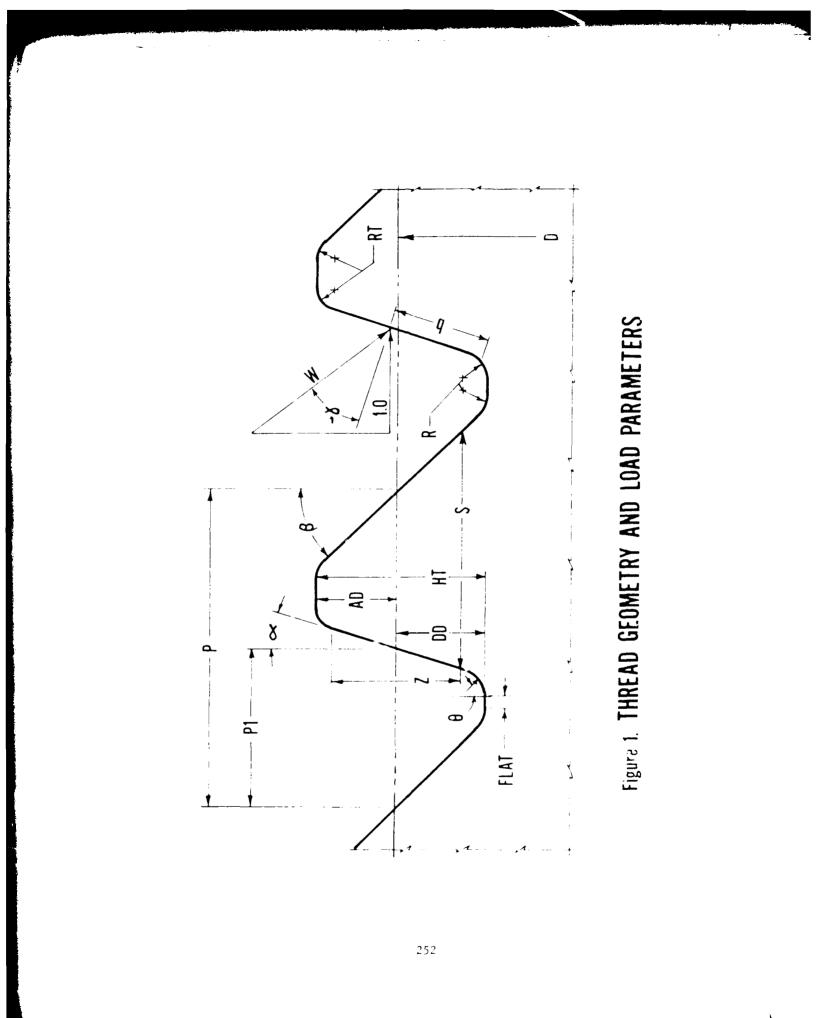
The three Heywood loads use three different values of friction coefficients -.5, 0.0, and +.5 where the sign on friction denotes the direction of the friction vector. Figures 7, 8, and 9 show the plots of nonlinear elements. It should be noted that the shear transfer rate for Figure 9 is lower than the other two. This is because that solution exceeded the 6% strain maximum of the stress-strain definition and the solution was stopped at that point. The arrows in these plots point out the element where the maximum stress occurs which is different in each of these solutions and the axial stress plot.

These plots of nonlinear element show one part of the overall picture. The next thing to look at is the maximum stress in the fillet. Figure 10 is the curve of fillet stress vs axial stress for the axial load. This solution was stopped at this point because the constraint input leaves some question about the nominal input stress when the bulk stress exceeds the yield point. The maximum fillet stresses for all three Heywood loads are plotted against shear transfer rate in Figure 11. The very high values for the stresses are the result of the multi-axial stress state in the fillet and other than that the plot speaks for itself.

NASTRAN uses the displacement method and displacements are often more useful in evaluating a problem than stresses so an example of displacement seems in order. Figure 12 shows the Z or axial displacement of grid point number 155 which is at the mid point of the primary bearing surface (on the pitch line) for Heywood loads. In this plot the displacements have been connected to reference the bottom of the fillet as the zero point. The difference here is well defined although not is marked as is the fillet stress case, probably because fillet stress is a much more localized effect than this displacement.

V. CONCLUSION. In conclusion this paper has attempted to define a method of elastic-plastic analysis of individual thread teeth. The problem of how to define reasonable loading condition for a specific practical problem has not been defined. Even with this limitation, an example has shown the relative magnitude of several loading effects. The reader should pay particular attention to the very definite effect of friction on the behavior of the thread.
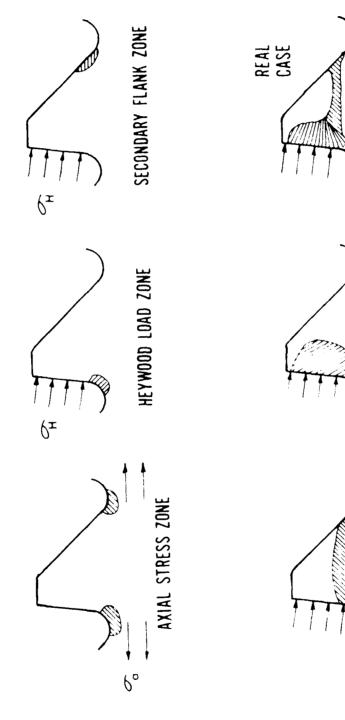
<div align="center">REFERENCES</div>

1. R. B. Heywood, "Tensile Fillet Stresses in Loaded Projections," Proceedings of the Institute of Mechanical Engineering, Vol. 160, p. 124, 1949.

2. R. E. Weigle, R. R. Lasselle and J. P. Purtell, "Experimental Investigation of the Fatigue Behavior of Thread-Type Projections," Experimental Mechanics, Number 5, Vol. 3, pp. 105-111, 1963.

3. R. E. Weigle and R. R. Lasselle, "Experimental Techniques for Predicting Fatigue Failure of Cannon-Breech Mechanisms," Experimental Mechanics, Feb. 1965.

4. G. P. O'Hara, "Finite Element Analysis of Treaded Connections," Proceedings of the Army Symposium on Solid Mechanics, AMMRC, MS-74-8, pp. 99-119, 1974.

5. G. P. O'Hara, "Stress Concentration in Screw Threads," ARRADCOM Technical Report, ARLCB-TR-80010, 1980.

6. "Aero Space Structural Metals Handbook," Mechanical Properties Data Center, Belfour Slaton Inc., Code 3207, pp. 15-16.

7. "The NASTRAN Theoretical Manual," R. H. MacNeal, Editor, NASA SP-221, April 1972.

8. S. Pepi, G. P. O'Hara, "Implementation of a Trapezoidal Ring Element in NASTRAN for Elastic-Plastic Analysis," ARRADCOM Technical Report, ARLCB-TR-79034, December 1979.

Figure 1. THREAD GEOMETRY AND LOAD PARAMETERS

PLASTIC ZONES

AXIAL STRESS ZONE

HEYWOOD LOAD ZONE

SECONDARY FLANK ZONE

SHEAR FAILURE ZONE

BEARING FAILURE ZONE

REAL CASE

COMBINATION

FIGURE 2

253

STRESS VS. STRAIN

FOR 7075-T6 ALIMINIUM

5.6% ZN.
2.5% MG.
1.6% CU.
0.3% CR.

FIGURE 3

LOADS

FIGURE 4

255

RADIAL PLANE

AXIAL CYLINDER

RADIAL PLANE

BRITISH STANDARD BUTTRESS

BASIC FINITE ELEMENT GRID

FIGURE 5

256

BRITISH STANDARD BUTTRESS

NON-LINEAR ELEMENTS FOR AN
AXIAL LOAD OF 65.0 KSI.

FIGURE 6

BRITISH STANDARD BUTTRESS

NON-LINEAR ELEMENTS FOR A
HEYWOOD LOAD WITH
FRICTION = -.5 AND
SHEAR TRANSFER = 21.0 KSI.

FIGURE 7

258

BRITISH STANDARD BUTTRESS

NON-LINEAR ELEMENTS FOR A
HEYWOOD LOAD WITH
FRICTION =0.0 AND
SHEAR TRANSFER = 31.0 KSI.

FIGURE 5

259

BRITISH STANDARD BUTTRESS

NON-LINEAR ELEMENTS FOR A
HEYWOOD LOAD WITH
FRICTION = +.5 AND
SHEAR TRANSFER = 25.7 KSI.

FIGURE 9

260

FILLET STRESS
VS.
AXIAL STRESS

FOR THE

BRITISH
STANDARD
BUTTRESS

FIGURE 10

261

FIGURE 11

CORRECTED
DISPLACEMENT
VS.
SHEAR TRANSFER

FOR THE

BRITISH
STANDARD
BUTTRESS

USING THREE
DIFFERENT
FRICTION
VALUES

$\dfrac{\Delta_2'^{ss}}{P} \times 10^3$

20

10

$f = 0.5$

$f = 0.0$

$f = -1.5$

T_R

10   20   30   KSI

FIGURE 12

263

GENERALIZED PLANE-STRAIN PROBLEMS IN AN ELASTIC-PLASTIC THICK-WALLED CYLINDER

P. C. T. Chen
U.S. Army Armament Research and Development Command
Benet Weapons Laboratory, LCWSL
Watervliet Arsenal, Watervliet, NY 12189

ABSTRACT. A new finite-difference approach has been developed for solving the generalized plane-strain problems of partially-plastic thick-walled cylinders made of strain hardening or ideally-plastic materials. The tube is assumed to obey the von Mises' criterion, the Prandtl-Reuss flow theory and the isotropic-hardening rule. The forces include internal pressure, external pressure, and end force. An incremental approach is used and no iteration is needed for each increment. The approach is simpler than othe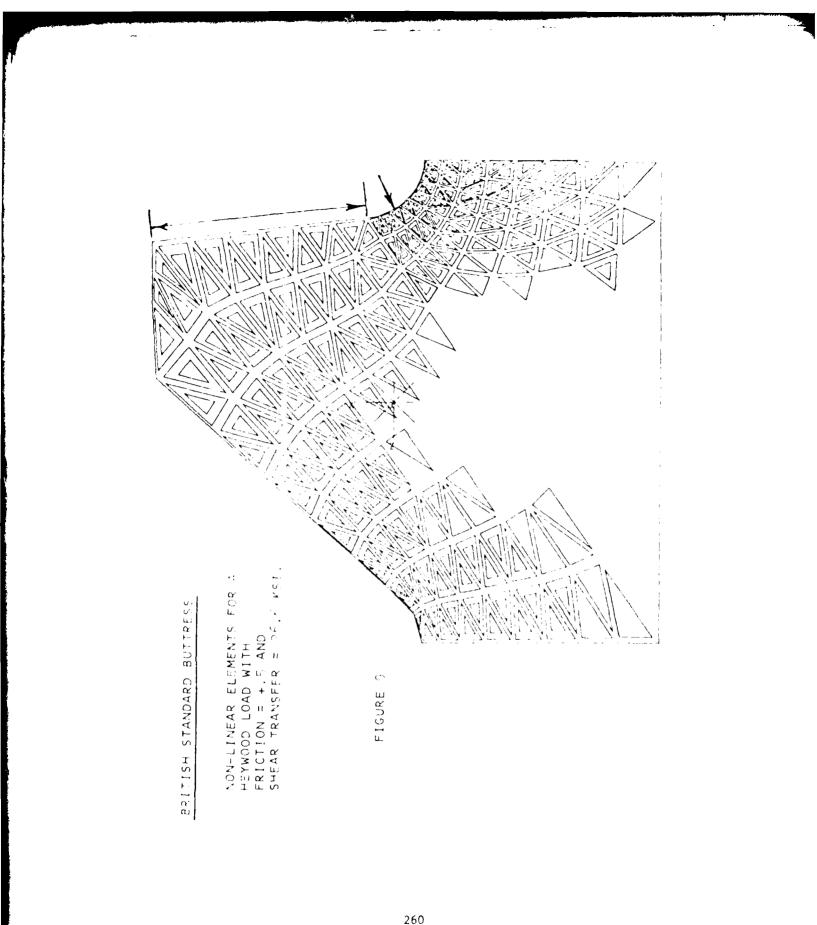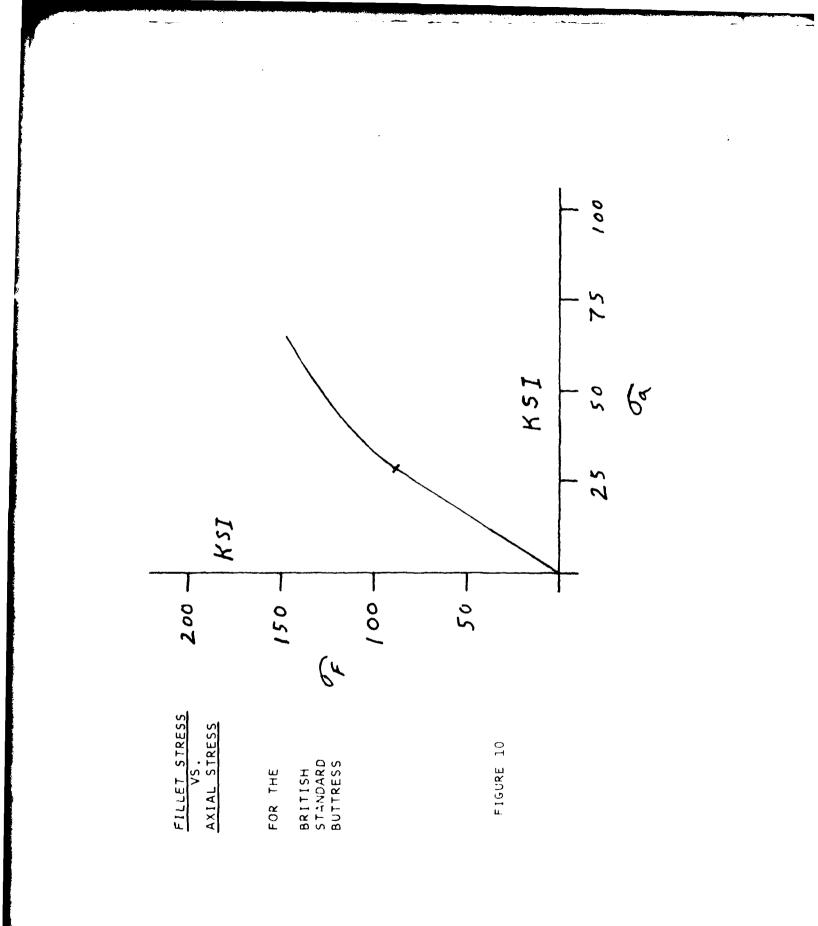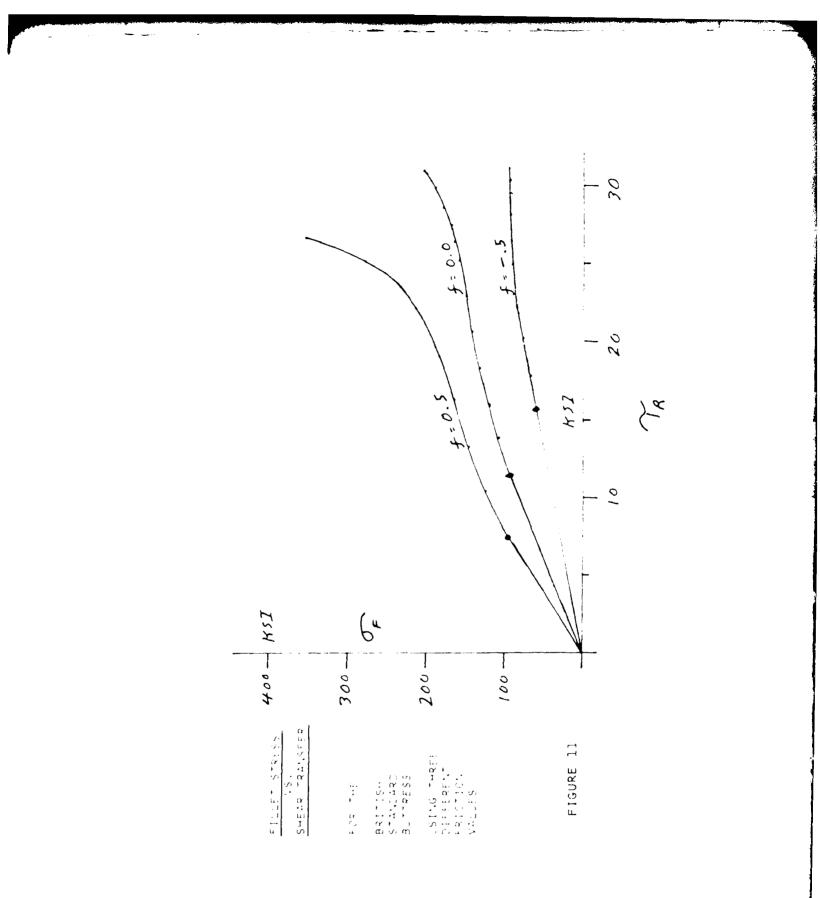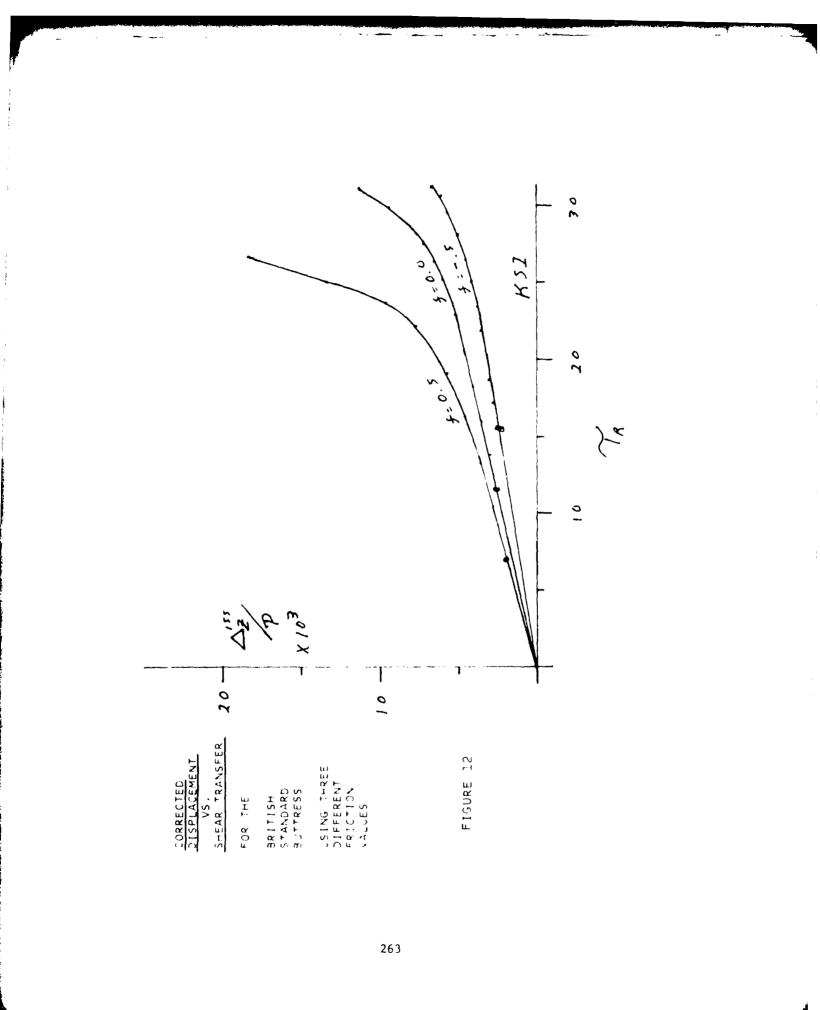rs yet quite general and accurate. The desired accuracy can be achieved by reducing the grid sizes and load increments. Some numerical results for the stresses and displacements in partially-plastic thick-walled cylinders with either open-end or closed-end conditions are presented.

I. INTRODUCTION. In a recent paper [1], a new finite-difference approach was developed for solving the axisymmetric plane-strain problems subjected to internal or external pressure beyond the elastic limit. The material was assumed to obey the von Mises' yield criterion, the Prandtl-Reuss flow theory and the isotropic hardening rule. The ideally-plastic material was treated as a special case. The new formulation was also used to determine the residual stresses in hollow cylinders due to quenching [2]. Since the plane-strain end condition was introduced only for simplicity, it is desirable to extend the approach to consider practical problems with either open-end or closed-end conditions.

In the present paper, the finite-difference approach is developed for solving the generalized plane-strain problems of thick-walled cylinders subjected to internal pressure, external pressure or end force beyond the elastic limit. The explicit equations between the incremental-stresses and incremental strains are used. The present approach is valid for ideally-plastic [3] as well as strain-hardening materials [4]. The approach is simpler than others [3,4] yet quite general and accurate. The desired accuracy can be achieved by reducing the grid size and load increments. No iteration is needed in each incremental loading step.

265

II. BASIC EQUATIONS. Assuming small strain and no body forces in the axisymmetric state of generalized plane strain, the radial and tangential stresses, σ_r and σ_θ, must satisfy the equilibrium equation,

$$r(\partial \sigma_r/\partial r) = \sigma_\theta - \sigma_r \quad ; \tag{1}$$

and the corresponding strains, ε_r and ε_θ, are given in terms of the radial displacement, u, by

$$\varepsilon_r = \partial u/\partial r \quad , \quad \varepsilon_\theta = u/r \quad . \tag{2}$$

It follows that the strains must satisfy the equation of compatibility

$$r(\partial \varepsilon_\theta/\partial r) = \varepsilon_r - \varepsilon_\theta \quad . \tag{3}$$

Whereas the differential equations (1), (2), and (3) hold throughout the tube regardless of the material properties, the constitution equations assume various forms according to the adopted form of yield function, hardening rule, total or incremental theory of plasticity. In the present paper, the material is assumed to be elastic-plastic, obeying the Mises' yield criterion, the Prandtl-Reuss flow theory and the isotropic hardening law. The complete stress-strain relations are [5]:

$$d\varepsilon_i' = d\sigma_i'/2G + (3/2)\sigma_i' d\sigma/(\sigma H') \tag{4}$$

$$d\sigma \geq 0 \quad \text{for} \quad i = r,\theta,z$$

$$d\varepsilon_m = E^{-1}(1-2\nu)d\sigma_m \tag{5}$$

where E, ν Young's modulus, Poisson's respectively,

$$2G = E/(1+\nu)$$

$$\varepsilon_m = (\varepsilon_r+\varepsilon_\theta+\varepsilon_z)/3 \quad , \quad \varepsilon_i' = \varepsilon_i - \varepsilon_m \quad ,$$

$$\sigma_m = (\sigma_r+\sigma_\theta+\sigma_z)/3 \quad , \quad \sigma_i' = \sigma_i - \sigma_m \quad ,$$

$$\sigma = (1/\sqrt{2})[(\sigma_r-\sigma_\theta)^2 + (\sigma_\theta-\sigma_z)^2 + (\sigma_z-\sigma_r)^2]^{1/2} \geq \sigma_o \quad , \tag{6}$$

and σ_o is the yield stress in simple tension or compression. For a strain-hardening material, H' is the slope of the effective stress/plastic strain curve

$$\sigma = H(\int d\varepsilon^p) \quad . \tag{7}$$

For an ideally-plastic material ($H' = 0$), the quantity $(3/2)d\sigma/(\sigma H')$ is replaced by $d\lambda$, a positive factor of proportionality. When $\sigma < \sigma_0$ or $d\sigma < 0$, the state of stress is elastic and the second term in equation (4) disappears. Following Yamada et al [6], equations (4) and (5) can be rewritten in an incremental form

$$d\sigma_i = d_{ij} d\epsilon_j \quad \text{for} \quad i,j = r,\theta,z$$

and

$$d_{ij}/2G = \nu/(1-2\nu) + \delta_{ij} - \sigma_i' \sigma_j'/S \quad ,$$

where

$$S = \frac{2}{3}(1 + \frac{1}{3} H'/G)\sigma^2 \quad , \quad H'/E = \alpha/(1-\alpha) \quad , \tag{8}$$

αE is the slope of the effective stress-strain curve, and δ_{ij} is the Kronecker delta.

This form was used in the finite-element formulation for solving elastic-plastic thick-walled tube problems [7]. In the following section, the incremental stress-strain matrix will be used in the finite difference formulation.

III. FINITE-DIFFERENCE FORMULATION. Consider an open-end or closed-end thick-walled cylinder of inner radius a and external radius b. The tube is subjected to inner pressure p, external pressure q, and end force f. The elastic solution for this problem is well-known and the pressure p^*, q^*, or f^* required to cause initial yielding can be determined by using the Mises' yield criterion. For loading beyond the elastic limit, an incremental approach of the finite-difference formulation is used. The analysis starts with the applied pressure p, q, or f and the loading path is divided into m increments with

$$\Delta p = (p-p^*)/m \quad , \quad \Delta q = (q-q^*)/m \quad , \quad \Delta f = (f-f^*)/m \quad . \tag{10}$$

The cross section of the tube is divided into n rings with

$$r_1 = a, r_2, \ldots, r_k = \rho, \ldots, r_{n+1} = b \quad , \tag{11}$$

where ρ is the radius of the elastic-plastic interface. At the beginning of each increment of loading, the distribution of displacements, strains and stresses is assumed to be known and we want to determine Δu, $\Delta \epsilon_r$, $\Delta \epsilon_\theta$, $\Delta \epsilon_z$, $\Delta \sigma_r$, $\Delta \sigma_\theta$, $\Delta \sigma_z$ at all grid points. Since the incremental stresses are related to the incremental strains by the incremental form (Eq. (8)) and $\Delta u = r\Delta \epsilon_\theta$, there exists only three unknowns at each station that have to be determined for each increment of loading. Accounting for the fact that the axial strain ϵ_z is independent of r, the unknown variables in the present formulation are $(\Delta \epsilon_\theta)_i$, $(\Delta \sigma_r)_i$, for $i = 1,2,\ldots,n,n+1$, and $\Delta \epsilon_z$.

The equation of equilibrium (1) and the equation of compatibility (3) are valid for both the elastic and the plastic regions of a thick-walled tube. The finite-difference forms of these two equations at $i = 1,\ldots,n$ are given in [4] by

$$(r_{i+1}-2r_i)(\Delta\sigma_r)_i - (r_{i+1}-r_i)(\Delta\sigma_\theta)_i + r_i(\Delta\sigma_r)_{i+1}$$

$$= (r_{i+1}-r_i)(\sigma_\theta-\sigma_r)_i - r_i[(\sigma_r)_{i+1} - (\sigma_r)_i] \qquad (12)$$

for the equation of equilibrium, and

$$(r_{i+1}-2r_i)(\Delta\varepsilon_\theta)_i - (r_{i+1}-r_i)(\Delta\varepsilon_r)_i + r_i(\Delta\varepsilon_\theta)_{i+1}$$

$$= (r_{i+1}-r_i)(\varepsilon_r-\varepsilon_\theta)_i - r_i[(\varepsilon_\theta)_{i+1} - (\varepsilon_\theta)_i] \qquad (13)$$

for the equation of compatibility. With the aid of the incremental stress-strain relations (Eq. (8)), equation (12) can be rewritten as

$$[(r_{i+1}-2r_i)(d_{12})_i + (-r_{i+1}+r_i)(d_{22})_i](\Delta\varepsilon_\theta)_i$$

$$+ [(r_{i+1}-2r_i)(d_{11})_i + (-r_{i+1}+r_i)(d_{21})_i](\Delta\varepsilon_r)_i$$

$$+ r_i(d_{12})_{i+1}(\Delta\varepsilon_\theta)_{i+1} + r_i(d_{11})_{i+1}(\Delta\varepsilon_r)_{i+1}$$

$$+ [(r_{i+1}-2r_i)(d_{13})_i + (-r_{i+1}+r_i)(d_{23})_i + r_i(d_{13})_{i+1}]\Delta\varepsilon_z$$

$$= (r_{i+1}-r_i)(\sigma_\theta-\sigma_r)_i - r_i[(\sigma_r)_{i+1} - (\sigma_r)_i] \quad . \qquad (14)$$

The boundary conditions for the problem are

$$\Delta\sigma_r(a,t) = -\Delta p \quad , \quad \Delta\sigma_r(b,t) = -\Delta q \quad ,$$

$$\pi \sum_{i=1}^{n} [r_i(\Delta\sigma_z)_i + r_{i+1}(\Delta\sigma_z)_{i+1}](r_{i+1}-r_i) = \mu\pi a^2\Delta p + \Delta f \quad , \qquad (15)$$

where μ is 0 for open-end tubes and 1, for closed-end tubes. Using the incremental relations (Eq. (8)), we rewrite equation (15) as

$$(d_{12})_1(\Delta\varepsilon_\theta)_1 + (d_{11})_1(\Delta\varepsilon_r)_1 + (d_{13})_1\Delta\varepsilon_z = -\Delta p \quad , \qquad (16)$$

$$(d_{12})_{n+1}(\Delta\varepsilon_\theta)_{n+1} + (d_{11})_{n+1}(\Delta\varepsilon_r)_{n+1} + (d_{13})_{n+1}\Delta\varepsilon_z = -\Delta q \quad , \qquad (17)$$

and

$$\sum_{i=1}^{n} (r_{i+1}-r_i)\{r_i[(d_{23})_i(\Delta\varepsilon_\theta)_i + (d_{13})_i(\Delta\varepsilon_r)_i] + r_{i+1}[(d_{23})_{i+1}(\Delta\varepsilon_\theta)_{i+1}$$

$$+ (d_{13})_{i+1}(\Delta\varepsilon_r)_{i+1}]\} + \sum_{i=1}^{n} (r_{i+1}-r_i)[r_i(d_{33})_i + r_{i+1}(d_{33})_{i+1}]\Delta\varepsilon_z$$

$$= \mu a^2 \Delta p + \Delta f/\pi \quad . \tag{18}$$

Now we can form a system of $2n+3$ equations for solving $2n+3$ unknowns, $(\Delta\varepsilon_\theta)_i$, $(\Delta\varepsilon_r)_i$, at $i = 1,2,\ldots,n,n+1$ and $\Delta\varepsilon_z$. Equations (16), (17), and (18) are taken as the first and last two equations, respectively, and the other $2n$ equations are set up at $i = 1,2,\ldots,n$ using (13) and (14). The final system is an unsymmetric matrix of arrow type with the nonzero terms appearing in the last row and column and others clustered about the main diagonal, two below and one above.

In the computer program which was developed, the dimensionless quantities r/a, $E\varepsilon_r/\sigma_0$, $E\varepsilon_\theta/\sigma_0$, $E\varepsilon_z/\sigma_0$, σ_r/σ_0, σ_θ/σ_0, σ_z/σ_0, p/σ_0, q/σ_0, $f/(\pi a^2 \sigma_0)$ were used in the formulation and the Gaussian elimination method was used to solve these equations. All calculations were carried out on IBM 360/Model 44 with double precision to reduce round-off errors.
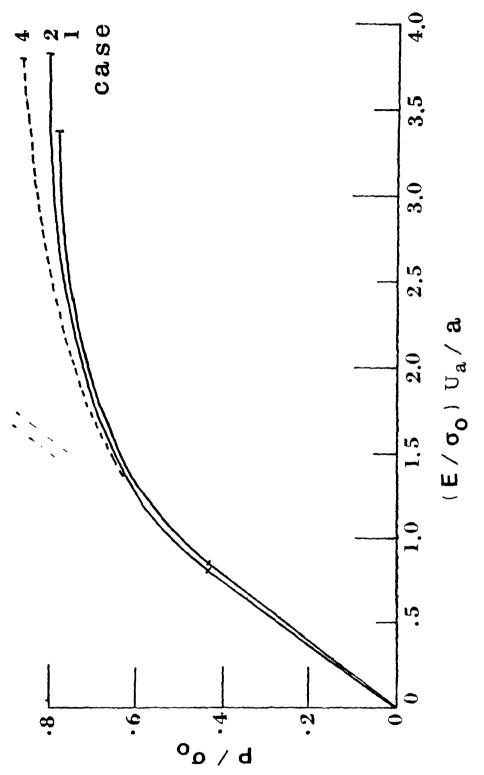
IV. NUMERICAL RESULTS AND DISCUSSIONS. The generalized plane-strain problems of thick-walled cylinders subjected to internal pressure p beyond the elastic limit were solved. The elastic-perfectly-plastic as well as strain-hardening materials were considered for open-end or closed-end conditions. The numerical result were based on the following parameters: $b/a = 2$, $\nu = 0.3$, $\alpha = 0.05$, $\mu = 0$ or 1. Various values of m and n were used to test the convergence of the numerical results. The incremental loadings were applied until the fully plastic state was reached. The value for p corresponding to this final state was denoted by p^{**}. It was found that the results of these values for all four cases converge by increasing m and/or n. For $n = 100$, $\Delta p/\sigma_0 = 0.0004$, we have $p^{**}/\sigma_0 = 0.7802$ ($\alpha = 0$, $\mu = 0$ as case 1); 0.8038 ($\alpha = 0$, $\mu = 1$ as case 2); 0.8222 ($\alpha = 0.05$, $\mu = 0$ as case 3); 0.8618 ($\alpha = 0.05$, $\mu = 1$ as case 4). Additional results are shown in Figures 1 to 5. Figure 1 shows the bore radial displacements as functions of internal pressure for cases 1, 2, and 4. Figure 2 shows the relations between internal pressure and elastic-plastic boundary for cases 1, 2, and 4. The effects of end conditions and strain hardening can also be seen in these two figures. The distributions of radial and tangential stresses for $\rho/a = 1.0$, 1.2, 1.4, 1.6, and 1.8 are shown in Figure 3 for case 1 and in Figure 4 for case 2. Finally the distributions of axial stress for $\rho/a = 1.0$, 1.4, and 1.8 are shown in Figure 5 for cases 1 and 2. The effect of end conditions and elastic-plastic boundary on the axial stress is quite significant.
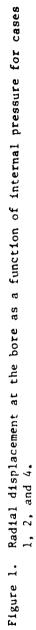
The present approach determines $\Delta\epsilon_z$ directly for each step of incremental loadings whereas in [4], many iterations were needed because a value of $\Delta\epsilon_z$ was assumed. In addition, the computer storage needed in this approach was only 35% of that in [4], and much larger n can be used to yield better results. The present approach is simpler yet more general than the other finite-difference approaches because both ideally-plastic [3] and strain-hardening materials [4] can be considered in a unified manner. Furthermore, very accurate numerical results can be obtained and used to verify the accuracy of [7,8] the finite-element programs.

REFERENCES

1. Chen, P. C. T., "A Finite-Difference Approach to Axisymmetric Plane-Strain Problems Beyond the Elastic Limit," Transactions Twenty-Fifth Conference of Army Mathematicians, pp. 455-466, January 1980.

2. Vasilakis, J. D. and Chen, P. C. T., "Thermo-Elastic-Plastic Stresses in Hollow Cylinders Due to Quenching," Transactions Twenty-Fifth Conference of Army Mathematicians, pp. 661-674, January 1980.

3. Elder, A. S., Tomkins, R. C., and Mann, T. L., "Generalized Plane Strain in an Elastic, Perfectly Plastic Cylinder, With Reference to the Hydraulic Autofrettage Process," Tran. 21st Conference of Army Mathematicians, 1975, pp. 623-659.

4. Chu, S. C., "A More Rational Approach to the Problem of an Elastoplastic Thick-Walled Cylinder," J. of the Franklin Institute, Vol. 294, 1972, pp. 57-65.

5. Hill, R., Mathematical Theory of Plasticity, Oxford University Press, 1950.

6. Yamada, Y., Yoshimura, N., and Sakurni, T., "Plastic Stress-Strain Matrix and Its Application For the Solution of Elastic-Plastic Problems by the Finite Element Method," Int. J. Mech. Sci., Vol. 10, 1968, pp. 343-354.

7. Chen, P. C. T., "The Finite Element Analysis of Elastic-Plastic Thick-Walled Tubes," Proc. of Army Symposium on Solid Mechanics, 1972, The Role of Mechanics in Design-Ballistic Problems, pp. 243-253.

8. Chen, P. C. T., "Elastic-Plastic Solution of a Two-Dimensional Tube Problem by Using the Finite Element Method," Trans. 19th Conf. of Army Math, 1973, pp. 703-734.

Figure 1. Radial displacement at the bore as a function of internal pressure for cases 1, 2, and 4.
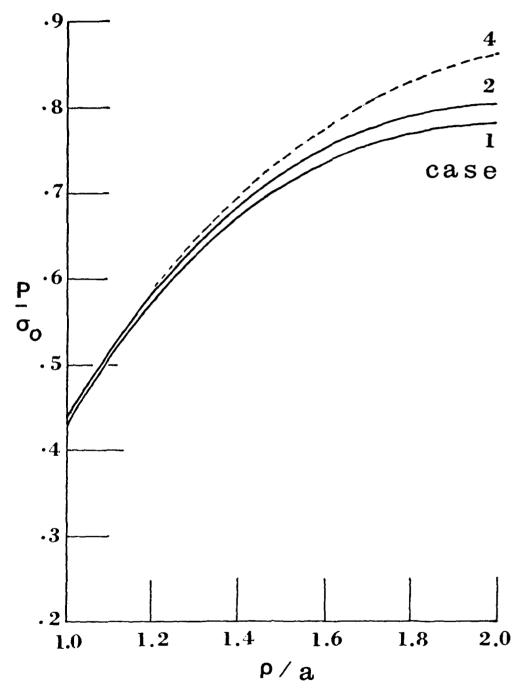
Figure 2. Elastic-plastic boundary as a function of internal
pressure for cases 1, 2, and 4.

272
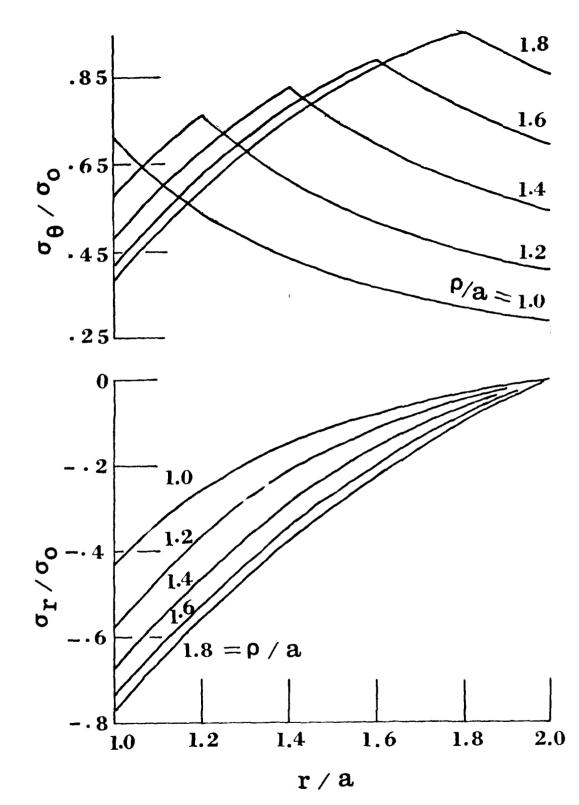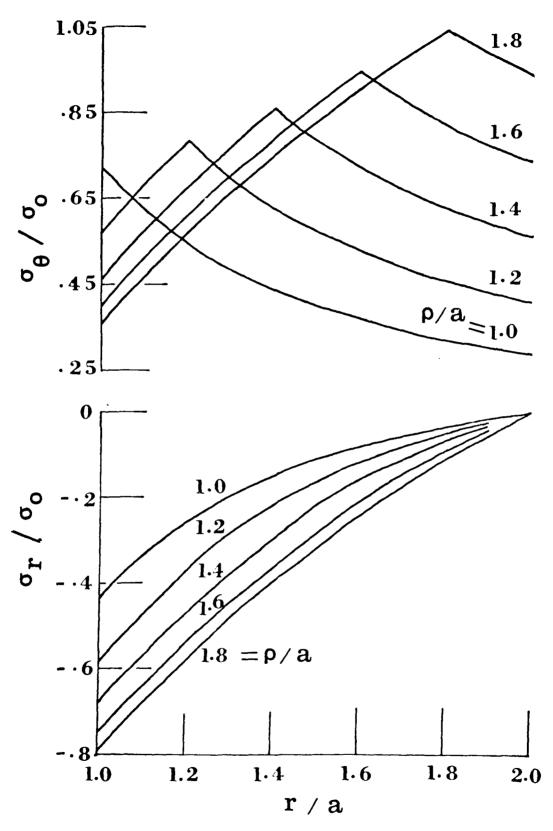
Figure 3. Distributions of radial and tangential stresses for case 1 ($\alpha = 0$, $\mu = 0$, $b/a = 2$).

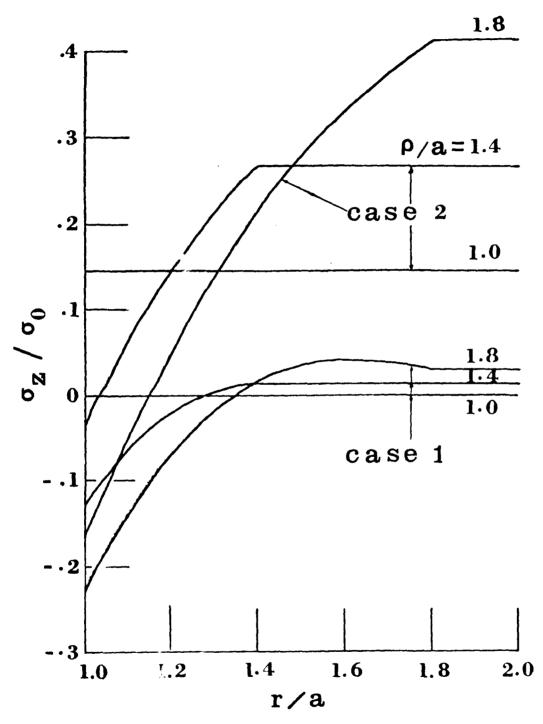Figure 4. Distributions of radial and tangential stresses for case 2 ($\alpha = 0$, $\mu = 1$, $b/a = 2$).

Figure 5. Distributions of axial stress for cases 1 and 2.

QUADRATIC AND CUBIC TRANSITION ELEMENTS

M. A. Hussain, J. D. Vasilakis, and S. L. Pu
U.S. Army Armament Research and Development Command
Benet Weapons Laboratory, LCWSL
Watervliet, NY 12189

ABSTRACT. Based on the investigations of Barsoum [1], Henshell and
Shaw [2], quarter-point quadratic elements have been successfully used as
crack tip elements in fracture mechanics. This concept of singular ele-
ment was extended to cubic isoparametric elements [3]. Recently it was
discovered by Lynn and Ingraffea [4] that under special configuration,
transitional elements improve the accuracy of stress intensity factor
computations. These transitional elements are located in the immediate
vicinity of the singular elements with the mid-side nodes so adjusted as
to reflect or extrapolate the square root singularity on the stresses and
strains at the tip of the crack.

In this paper, we have obtained the locations of mid-side nodes of
these transitional elements for the quadratic as well as cubic elements.
Explicit computations for a typical element are symbolically carried out
using MACSYMA* [5]. These computations reveal that in addition to the
desired square root singularities, the crack tip senses a stronger
singularity, i.e., of order one. Further, the strength of this singu-
larity cannot be controlled, as was possible for the cubic and quadratic
collapsed elements, where, by tying the collapsed nodes together, we
could easily annihilate this strong singularity.

These cubic elements also have Hibbit-type [6] singularities. The
locations of mid-side nodes for these singularities have also been deter-
mined.

The cubic transitional elements were used for double-edge crack prob-
lem, and it was found that there was improvement in accuracy for a config-
uration which consisted only of singular and transitional elements. How-
ever, for a well-laid out grid, the improvement was only marginal. MACSYMA
has proved to be an indispensible tool for the present investigation.

---

*MACSYMA is a large program for symbolic manipulation at MIT.

<u>SECTION I.</u> Consider a quadratic quadrilateral isoparametric element,

$$x = \sum_{i=1}^{8} N_i X_i \quad , \quad y = \sum_{i=1}^{8} N_i Y_i \tag{1}$$

$$u = \sum_{i=1}^{8} N_i U_i \quad , \quad v = \sum_{i=1}^{8} N_i V_i \tag{2}$$

where N_i are the shape function of 'Serendipity' family [6], and are given by,

$$\text{CORNER NODES} \quad : \quad N_1 = \frac{1}{4} (1-\xi)(1-\eta)(-\xi-\eta-1) \quad , \quad \text{etc.} \tag{3}$$

$$\text{MID-SIDE NODES} \quad : \quad N_5 = \frac{1}{2} (1-\xi^2)(1-\eta) \quad , \quad \text{etc.} \tag{4}$$

Without loss of generality consider the sectorial element, together with the mapped unit element in the transformed plane, shown in Figure 1. For simplicity, considering the one dimensional case along line 1-2 in Figure 1 (i.e., $\eta = -1$) we have from (1)

$$x = \frac{1}{2} \xi(\xi-1) + \frac{1}{2} \xi(1+\xi)L + (1-\xi^2)\beta L \tag{5}$$

The condition for the coalescence of roots of (5) at $x = 0$, together with the condition that $\beta L > 1$ gives

$$\beta L = \frac{L+2\sqrt{L}+1}{4} \tag{6}$$

This is the result, in a slightly different form, obtained by Lynn and Ingraffea [4]. With this location of mid-side nodes, the mapping of the general element of Figure 1 becomes, from (1) and (2),

$$x = \frac{1}{8} \{(\eta+1) \cos \alpha + (1-\eta)\}\{\xi(\sqrt{L}-1) + (\sqrt{L}+1)\}^2 \tag{7}$$

$$y = \frac{1}{8} (\eta+1)\{\xi(\sqrt{L}-1) + (\sqrt{L}+1)\}^2 \sin \alpha \tag{8}$$

The Jacobian of the transformation (1) and (2) is then given by

$$J = \frac{\partial(x,y)}{\partial(\xi,\eta)} = \frac{1}{16}(\sqrt{L}-1)\{\xi(\sqrt{L}-1) + (\sqrt{L}+1)\}^3 \sin \alpha \tag{9}$$

As can be seen from (7), (8), and (9) that the Jacobian has a third order zero while x and y have second zeroes at

$$\xi = - \frac{\sqrt{L}+1}{\sqrt{L}-1} \tag{10}$$

Using the inverse of the Jacobian matrix, the strain component can be written as

$$\frac{\partial u}{\partial x} = \frac{1}{J} \left\{ \frac{\partial u}{\partial \xi} \frac{dy}{d\eta} - \frac{\partial u}{\partial \eta} \frac{dy}{d\xi} \right\} \tag{11}$$

Substituting the various derivatives and collecting terms we get

$$\frac{\partial u}{\partial x} = \frac{A_1}{(\xi(\sqrt{L}-1) + \sqrt{L}+1)^2} + \frac{A_2}{(\xi(\sqrt{L}-1) + \sqrt{L}+1)} + A_3 \tag{12}$$

where A_1, A_2, and A_3 are given in the Appendix.

Comparing (12) with (7) and (8) it is seen that the strain component not only has singularity of order one half but also of order one. Similarly we have

$$\frac{\partial u}{\partial y} = \frac{1}{J} \left\{ - \frac{\partial u}{\partial \xi} \frac{dx}{d\eta} + \frac{\partial u}{\partial \eta} \frac{dx}{d\xi} \right\} = \frac{A_4}{(\xi(\sqrt{L}-1) + \sqrt{L}+1)^2} + \frac{A_5}{(\xi(\sqrt{L}-1) + \sqrt{L}+1)} + A_6 \tag{13}$$

where A_4, A_5, A_6 are given in the Appendix.

SECTION II. Consider now the cubic, 12 node, quadrilateral isoparametric element,

$$x = \sum_{i=1}^{12} N_i X_i \quad , \quad y = \sum_{i=1}^{12} N_i Y_i \tag{14}$$

and displacements

$$u = \sum_{i=1}^{12} N_i U_i \quad , \quad v = \sum_{i=1}^{12} N_i V_i \tag{15}$$

where the shape functions are given by

$$\text{CORNER NODES} : \quad N_1 = \frac{1}{32} (1-\xi)(1-\eta)\{9(\xi^2+\eta^2)-10\} \text{ , etc.} \tag{16}$$

279

$$\text{MID-SIDE NODES}: \quad N_2 = \frac{9}{32}(1-3\xi)(1-\xi^2)(1-\eta) \quad , \quad etc. \tag{17}$$

The general transitional element together with its map in ξ-η plane is given in Figure 2. For simplicity consider the one dimensional case along line 1-2-3-4 (i.e., $\eta = -1$),

$$x = \frac{1}{16} \left\{ \begin{array}{l} \xi^3(-9+27\beta_1 L-27\beta_2 L+9L) + \xi^2(9-9\beta_1 L-9\beta_2 L+9L) \\ + \quad (1-27\beta_1 L+27\beta_2 L-L) + (-1+9\beta_1 L+9\beta_2 L-L) \end{array} \right\} \tag{18}$$

The requirement that (18) be quadratic in ξ, together with the condition of coalescence of roots gives the following, physically possible solution for locations of mid-side nodes for all L,

$$\beta_1 L = \frac{L+4\sqrt{L}+4}{9} \quad ,$$
$$\beta_2 L = \frac{4L+4\sqrt{L}+1}{9} \quad . \tag{19}$$

With the above values the general mapping of the element shown in Figure 2 then becomes

$$x = \frac{1}{8}((\eta+1)\cos\alpha - (\eta-1))\{\xi(\sqrt{L}-1) + (\sqrt{L}+1)\}^2 \tag{20}$$

$$y = \frac{1}{8}(\eta+1)\{\xi(\sqrt{L}-1) + (\sqrt{L}+1)\}^2 \sin\alpha \tag{21}$$

and the Jacobian of the transformation becomes

$$J = \frac{\partial(x,y)}{\partial(\xi,\eta)} = \frac{1}{16}(\sqrt{L}-1)(\xi(\sqrt{L}-1) + (\sqrt{L}+1))^3 \sin\alpha \tag{22}$$

These expressions are the same as for quadratic elements (compare eqs. (7), (8), and (9)), and hence the Jacobian has third order zeroes and x,y have second order zeroes, at

$$\xi = -\frac{\sqrt{L}+1}{\sqrt{L}-1} \tag{23}$$

Following the procedure outlined before, the strain components can be obtained from the following,

$$\frac{\partial u}{\partial x} = \frac{B_1}{(\xi(\sqrt{L}-1) + \sqrt{L}+1)^2} + \frac{B_2}{(\xi(\sqrt{L}-1) + \sqrt{L}+1)} + B_3 \quad , \tag{24}$$

$$\frac{\partial u}{\partial v} = \frac{B_4}{(\xi(\sqrt{L}-1) + \sqrt{L}+1)^2} + \frac{B_5}{(\xi(\sqrt{L}-1) + \sqrt{L}+1)} + B_6 \quad , \tag{25}$$

where $B_1 - B_6$ are given in the Appendix. Similar expression hold for derivatives of v. Equation (24) and (25) again reveal the same kinds of singularities as (12) and (13).

SECTION III. In the cubic elements there is an additional set of locations of mid-side nodes which give Hibbit-Type [6] singularity. This is obtained from the condition that all the three roots of (18) coalesce. The location of nodes is given by

$$\beta_1 L = (\frac{L^{1/3} + 2}{3})^3$$
$$\beta_2 L = (\frac{2L^{1/3} + 1}{3})^3 \tag{26}$$

and the transformations become

$$x = \frac{1}{16} \{(\eta+1)\cos\alpha - (\eta-1)\}\{\xi(L^{1/3}-1) + L^{1/3}+1\}^3$$
$$y = \frac{1}{16} \{(\eta+1)\sin\alpha\}\{\xi(L^{1/3}-1) + L^{1/3}+1\}^3 \tag{27}$$

and the Jacobian becomes

$$J = \frac{\partial(x,y)}{\partial(\xi,\eta)} = \frac{3\sin(\alpha)}{128} (L^{1/3}-1)\{\xi(L^{1/3}-1) + L^{1/3}+1\}^5 \tag{28}$$

Following the procedure outlined before it can be shown that

$$\frac{\partial u}{\partial x} = \frac{C_1}{(\xi(L^{1/3}-1) + L^{1/3}+1)^3} + \frac{C_2}{(\xi(L^{1/3}-1) + L^{1/3}+1)^2} +$$

$$+ \frac{C_3}{(\xi(L^{1/3}-1) + L^{1/3}+1)} + C_4 \tag{29}$$

281

The above equation indicates that in this case the singularities are of order 1, 2/3, and 1/3. This combination is of no immediate interest in linear fracture in homogeneous media.

SECTION IV. The sample problem of a double-edge-cracked plate of [4] is selected for numerical assessment of transition elements when they are used with 12-node collapsed singular elements. Figure 3 is an idealization we usually take for such a mode I crack problem. The distance ρ between the crack tip and the nearest node in a collapsed element is often taken in the range of 0.5% to 3% of the crack length a. The ratios a/b and b/c are usually in the range of 2 to 10. Stress intensity factors for several values of ρ, b/c, and a/b with and without the use of transition elements are tabulated in Table I. Comparing to the reference value, $K_I = \sigma\sqrt{\pi a}F(a/2a)$, where $F(1/2) = 1.184$ [7], the percentage errors $\Delta\%$ are also shown in the table. The result with the use of transition elements is better only when a very large ratio of b/c ($\approx$20) is used.
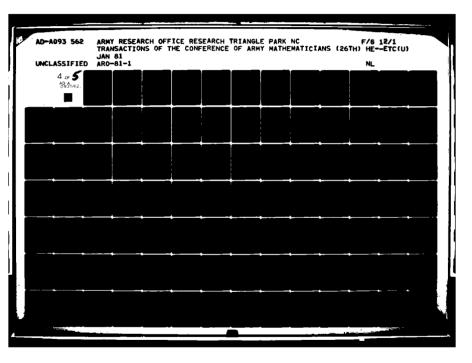
TABLE I.   STRESS INTENSITY FACTOR AND PERCENTAGE ERROR FOR A DOUBLE-EDGE-CRACKED PLATE USING 12-NODE COLLAPSED SINGULAR ELEMENTS WITH AND WITHOUT TRANSITION ELEMENTS. FINITE ELEMENT IDEALIZATION OF FIGURE 3.

ρ	b/c	a/b	Without Transition Elements		With Transition Elements	
			SIF	$\Delta\%$	SIF	$\Delta\%$
0.005	4	10	2.8808	2.31	2.8736	2.06
	10	4	2.8376	0.78	2.7831	-1.16
	20	2	2.9863	6.06	2.7851	-1.09
0.01	4	5	2.7986	-0.61	2.7926	-0.82
	10	2	2.8334	0.63	2.7813	-1.22

TABLE II.   STRESS INTENSITY FACTOR AND PERCENTAGE ERROR FOR A DOUBLE-EDGE-CRACKED PLATE USING 12-NODE COLLAPSED SINGULAR ELEMENTS WITH AND WITHOUT TRANSITION ELEMENTS. FINITE ELEMENT IDEALIZATION OF FIGURE 4.

ρ	a/c	Without Transition Elements		With Transition Elements	
		SIF	$\Delta\%$	SIF	$\Delta\%$
0.005	40	3.325	18.09	2.7658	-1.77
0.01	20	2.963	5.23	2.7654	-1.79
0.02	10	2.8115	~ 0.15	2.7650	-1.80
0.04	5	2.7632	~ 1.86	2.655	-5.71

Another idealization, Figure 4, similar to the one used by Lynn and Ingraffea [4] is used to recompute stress intensity factors for various values of a/c to see whether the transition elements in cubic isoparametric elements can give as good improvement in accuracy as reported in [4] in the quadratic isoparametric case. These results are tabulated in Table II. It shows again the result obtained from the use of transition elements is better only when a very large ratio of a/c is used.

In this paper the stress intensity factors are calculated from the normal component of displacement of the node on the crack surface and nearest to the crack tip. It usually gives better results than the average value computed from nodal displacements along the rays from the crack tip at various angles [8].

For elastic crack problems, the correct order of singularity at the crack tip is taken care by the collapsed singular elements. The use of transition elements does not practically improve the accuracy.

   CONCLUSIONS. In this paper we have been able to obtain explicit expressions for singularities the crack tip senses from a transitional element. The application of these elements for a few practical problems of fracture mechanics as well as stress concentration factors has been partially successful. It is believed that this is due to the fact that the crack tip senses not only the square root singularity but also a stronger singularity. The strength of this singularity cannot be controlled as was possible for collapsed singular elements, where the strong singularity was essentially eliminated by tying the nodes together.

REFERENCES

1. Barsoum, Roshdy S., "On the Use of Isoparametric Finite Elements in Linear Fracture Mechanics," Int. J. Num. Meth. Engrg., Vol. 10, 1976, pp. 25-76.

2. Henshell, R. D., and Shaw, K. G., "Crack Tip Elements are Unnecessary," Int. J. Num. Meth. Engrg., Vol. 9, 1975, pp. 495-507.

3. Pu, S. L., Hussain, M. A., and Lorensen, W. E., "The Collapsed Cubic Isoparametric Element as a Singular Element for Crack Problems," Int. J. Num. Meth. Engrg., Vol. 12, 1978, pp. 1727-1742.

4. Lynn, P. P., Ingraffea, A. R., "Transition Elements to be Used with Quarter-Point Crack-Tip Elements," Int. J. Num. Meth. Engrg., Vol. 12, 1978, pp. 1031-1036.

5. "MACSYMA Reference Manual," The Mathlab Group, Laboratory for Computer Science, MIT, Cambridge, Massachusetts.

6. Hibbit, H. D., "Some Properties of Singular Isoparametric Elements," Int. J. Num. Meth. Engrg., Vol. 11, 1977, pp. 180-184.

7. Tada, H., Paris, P., and Irwin, G., The Stress Analysis of Cracks Handbook, Del Research Corp., 1973.

8. Pu, S. L., Hussain, M. A., and Lorensen, W. E., "Collapsed 12-Node Triangular Elements as Crack Tip Elements for Elastic Fracture," Technical Report ARLCB-TR-77047, December 1977.

APPENDIX

In this appendix we give the explicit expressions for the coefficients of the various terms in the strain components given in the text.

$$R = \sqrt{L}$$

$$A_1 = \frac{2(\eta+1)}{(R-1)^2} \{4\eta(R-1)[Ru_8-u_6] + 4R[u_7-u_5]$$

$$+ (2\eta R+R-2\eta+1)[u_2-Ru_4] + (2\eta R-R-2\eta-1)[u_3-Ru_1]\}$$

$$A_2 = -\frac{1}{2(R-1)^2} \{2(3\eta^2+4\eta+1)(R-1)(u_8-u_6) + 4(\eta+1)(R+1)u_7$$

$$- 4(\eta+3)(R+1)u_5 + (R(3\eta^2+7\eta+4) - 3\eta(\eta+1))u_4$$

$$+ (3R\eta(\eta+1) - (3\eta^2+7\eta+4))u_3 + (R(3\eta^2+5\eta+4) - (3\eta^2+\eta-8))u_2$$

$$- (R(3\eta^2+\eta-8) - (3\eta^2+5\eta+4))u_1\}$$

$$A_3 = -\frac{2}{(R-1)^2} (2u_5-u_2-u_1)$$

$$A_4 = -\frac{2((\eta+1)\cos\alpha - (\eta-1))}{(R-1)^2\sin\alpha} \{4R\eta(R-1)u_8 + 4R(u_7-u_5) - 4\eta(R-1)u_6$$

$$- (R(2\eta+1) - (2\eta-1)(Ru_4-u_2) + (R(2\eta-1) - (2\eta+1))(u_3-Ru_1)\}$$

$$A_5 = \frac{1}{2(R-1)^2\sin\alpha} \{2(R-1)[\cos\alpha(3\eta^2+4\eta+1) - (3\eta^2-4\eta+1)](u_8-u_6)$$

$$+ 4(R+1)(\cos\alpha(\eta+1) - (\eta-3))u_7 - 4(R+1)(\cos\alpha(\eta+3) - (\eta-1))u_5$$

$$- (\cos\alpha[R(3\eta^2+7\eta+4) - 3\eta(\eta+1)] - [R(3\eta^2-\eta-8) - (3\eta^2-5\eta+4)])u_4$$

$$+ (\cos\alpha[3R\eta(\eta+1) - (3\eta^2+7\eta+4)] - [R(3\eta^2-5\eta+4) - (3\eta^2-\eta-8)])u_3$$

$$+ (\cos\alpha[R(3\eta^2+5\eta+4) - (3\eta^2+\eta-8)] + [-3R\eta(\eta-1) + (3\eta^2-7\eta+4)])u_2$$

$$- (\cos\alpha[R(3\eta^2+\eta-8) - (3\eta^2+5\eta+4)] + [-R(3\eta^2-7\eta+4) + 3\eta(\eta-1)])u_1\}$$

$$A_6 = -\frac{2}{(R-1)^2\sin\alpha} \{\cos\alpha(-2u_5+u_2+u_1) + (2u_7-u_4-u_3)\}$$

$$B_1 = \frac{\eta+1}{(R-1)^3} \{ \frac{1}{4} [R^2(27\eta^2-18\eta-1) - R(54\eta^2-36\eta-38) + 27\eta^2-18\eta-1](Ru_1-u_4)$$

$$+ 9R(2R+1)(u_9-u_2) + 9R(R+2)(u_3-u_8)$$

$$+ \frac{9}{4} (R-1)^2(9\eta^2-2\eta-3)(u_5-Ru_{12}) - \frac{9}{4} (R-1)^2(9\eta^2+2\eta-3)(u_6-Ru_{11})$$

$$+ \frac{1}{4} [R^2(27\eta^2+18\eta-1) - R(54\eta^2+36\eta-38) + 27\eta^2+18\eta-1](u_7-Ru_{10}) \}$$

$$B_2 = \frac{1}{(R-1)^3} \{ - \frac{1}{16} [R^2(45\eta^3+27\eta^2-\eta+105) - R(90\eta^3+54\eta^2-146\eta-222)$$

$$+ 45\eta^3+27\eta^2-37\eta-3]u_1 + \frac{9}{4} (2R^2+6R+1)[(\eta+3)u_2 - (\eta+1)u_9]$$

$$- \frac{9}{4} (R^2+6R+2)[(\eta+3)u_3 - (\eta+1)u_8] +$$

$$+ \frac{1}{16} [R^2(45\eta^3+27\eta^2-37\eta-3) - R(90\eta^3+54\eta^2-146\eta-222) +$$

$$+ 45\eta^3+27\eta^2-\eta+105]u_4$$

$$- \frac{9}{16} (R-1)^2(\eta+1)[(15\eta^2-7)u_5 + (15\eta^2+6\eta-5)u_{11}]$$

$$+ \frac{9}{16} (R-1)^2(\eta+1)[(15\eta^2+6\eta-5)u_6 + (15\eta^2-7)u_{12}]$$

$$- \frac{1}{16} (\eta+1)[R^2(45\eta^2+36\eta-1) - 2R(45\eta^2+36\eta-37) + 45\eta^2+36\eta+35]u_7$$

$$+ \frac{\eta+1}{16} [R^2(45\eta^2+36\eta+35) - 2R(45\eta^2+36\eta-37) + 45\eta^2+36 -1]u_{10} \}$$

$$B_3 = \frac{9}{2} \frac{1}{(R-1)^3} \{ (2R+1)u_1 - (3R+4)u_2 + (4R+5)u_3 - (R+2)u_4 \}$$

$$B_4 = \frac{((\eta+1)\cos \alpha - \eta + 1)}{\sin \alpha(R-1)^3} \{ \frac{1}{4} [R^2(27\eta^2-18\eta-1) - R(54\eta^2-36\eta-38) +$$

$$+ 27\eta^2-18\eta-1](-Ru_1+u_4) + 9R(2R+1)(u_2-u_9) +$$

$$+ 9R(R+2)(-u_3+u_8) + \frac{9}{4} (R-1)^2(9\eta^2-2\eta-3)(-u_5+Ru_{12})$$

$$+ \frac{9}{4} (R-1)^2(9\eta^2+2\eta-3)(u_6-Ru_{11}) +$$

$$+ \frac{1}{4} [R^2(27\eta^2+18\eta-1) - R(54\eta^2+36\eta-38) + 27\eta^2+18\eta-1](-u_7+Ru_{10}) \}$$

$$B_5 = \frac{1}{\sin \alpha (R-1)^3} \{ \frac{1}{16} [R^2(45\eta^3(\cos \alpha-1) + 27\eta^2(\cos \alpha+3) - \eta(\cos \alpha+71)$$

$$+ 35(3 \cos \alpha+1)) + R(90\eta^3(-\cos \alpha+1) - 54\eta^2(\cos \alpha+3) +$$

$$+ 2\eta(73 \cos \alpha-1) + 74(3 \cos \alpha+1)) + 45\eta^3(\cos \alpha-1) +$$

$$+ 27\eta^2(\cos \alpha+3) - \eta(37 \cos\alpha +35) - 3 \cos \alpha-1]u_1$$

$$- \frac{9}{4} ((\eta+3)\cos \alpha-\eta+1)[(2R^2+6R+1)u_2 - (R^2+6R+2)u_3]$$

$$- \frac{1}{16} [R^2(45\eta^3(\cos \alpha-1) + 27\eta^2(\cos \alpha+3) - \eta(37 \cos \alpha-35)$$

$$- (3 \cos \alpha+1)) + R(90\eta^3(-\cos \alpha+1) - 54\eta^2(\cos \alpha+3)$$

$$+ 2\eta(73 \cos \alpha-1) + 74(3 \cos \alpha+1)) + 45\eta^3(\cos \alpha-1) +$$

$$+ 27\eta^2(\cos \alpha+3) - \eta(\cos \alpha+71) + 35(3 \cos \alpha+1)]u_4$$

$$+ \frac{9}{16} (R-1)^2(15\eta^3(\cos \alpha-1) + 3\eta^2(5 \cos \alpha+7) - \eta(7 \cos \alpha+1)$$

$$- 7 \cos \alpha-5)u_5 - \frac{9}{16} (15\eta^3(\cos \alpha-1) + 3\eta^2(7 \cos \alpha+5)$$

$$+ \eta(\cos \alpha+7) - 5 \cos \alpha-7)(R-1)^2u_6 +$$

$$+ \frac{1}{16} [R^2(45\eta^3(\cos \alpha-1) + 27\eta^2(3 \cos \alpha+1) + \eta(35 \cos \alpha+37)$$

$$- (\cos \alpha+3)) + R(90\eta^3(-\cos \alpha+1) - 54\eta^2(3 \cos \alpha+1)$$

$$+ 2\eta(\cos \alpha-73) + 74(\cos \alpha+3)) + 45\eta^3(\cos \alpha-1) +$$

$$+ 27\eta^2(3 \cos \alpha+1) + \eta(71 \cos \alpha+1) + 35(\cos \alpha+3)]u_7$$

$$+ \frac{9}{4} ((\eta+1)\cos \alpha-\eta+3)[-(R^2+6R+2)u_8 + (2R^2+6R+1)u_9] -$$

$$- \frac{1}{16} [R^2(45\eta^3(\cos \alpha-1) + 27\eta^2(3 \cos \alpha+1) + (71 \cos \alpha+1) +$$

$$+ 35(\cos \alpha+3))+ R(90\eta^3(-\cos \alpha+1) - 54\eta^2(3 \cos \alpha+1)$$

$$+ 2\eta(\cos \alpha-73) + 74(\cos \alpha+3)) + 45\eta^3(\cos \alpha-1) +$$

$$+ 27\eta^2(3 \cos \alpha+1) + \eta(35 \cos \alpha+37) - (\cos \alpha+3)]u_{10} +$$

$$+ \frac{9}{16} (R-1)^2[15\eta^3(\cos \alpha-1) + 3\eta^2(7 \cos \alpha+5) + \eta(\cos \alpha+7)$$

$$- (5 \cos \alpha+7)]u_{11} -$$

$$- \frac{9}{16} (R-1)^2[15\eta^3(\cos \alpha-1) + 3\eta^2(5 \cos \alpha+7) - \eta(7 \cos \alpha+1)$$

$$- (7 \cos \alpha+5)]u_{12}$$

$$B_6 = \frac{9}{2 \sin \alpha (R-1)^3} \{ (2R+1)[-\cos \alpha u_1 + u_{10}] + (5R+4)[\cos \alpha u_2 - u_9]$$

$$- (4R+5)[\cos \alpha u_3 - u_8] + (R+2)[\cos \alpha u_4 - u_7]\}$$

FIGURE 1. QUADRATIC QUADRILATERAL ISOPARAMETRIC
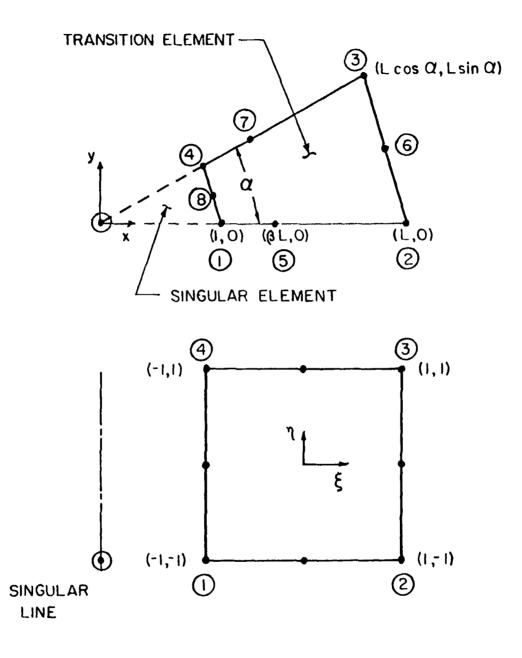
ELEMENT AS TRANSITION ELEMENT
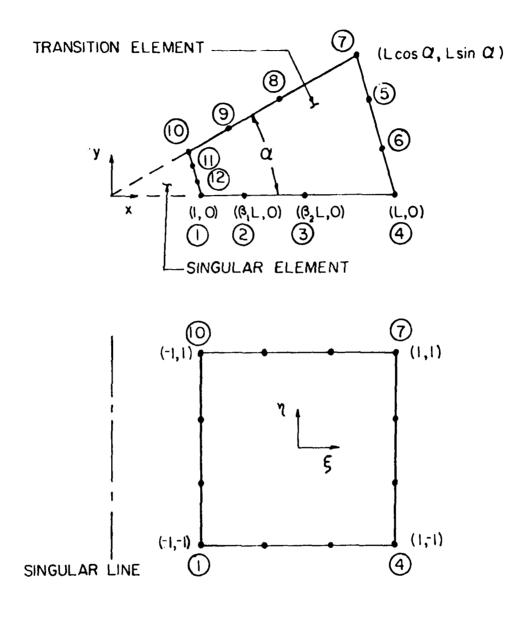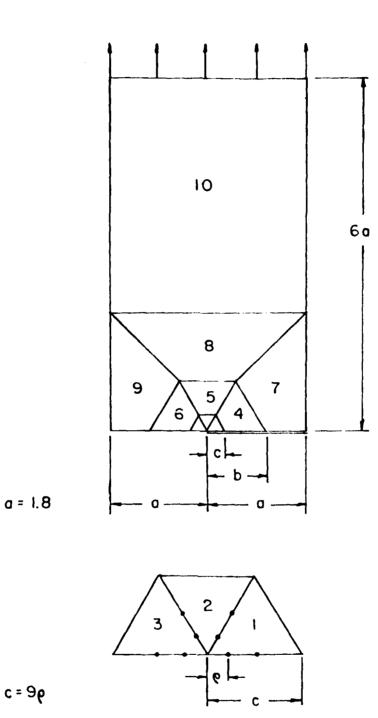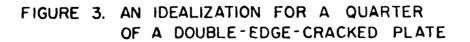
FIGURE 2. CUBIC QUADRILATERAL ISOPARAMETRIC
ELEMENT AS TRANSITION ELEMENT

FIGURE 3. AN IDEALIZATION FOR A QUARTER
OF A DOUBLE-EDGE-CRACKED PLATE
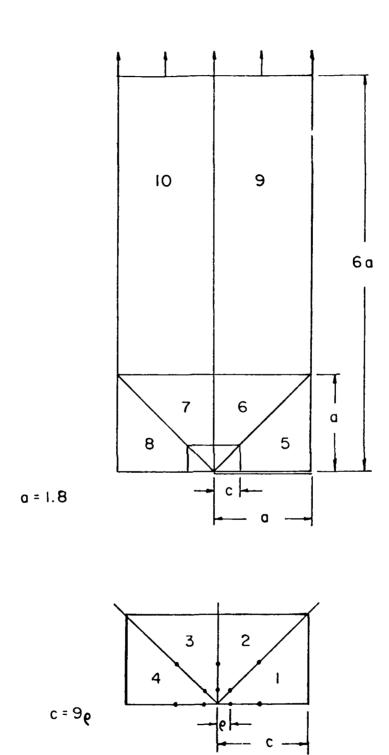
FIGURE 4. A SIMILAR IDEALIZATION USED IN [4] FOR A
QUARTER OF A DOUBLE-EDGE-CRACKED PLATE

PERTURBATION AND BIFURCATION
IN A FREE BOUNDARY PROBLEM

Roger K. Alexander[*] and Bernard A. Fleishman[†]
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, New York   12181

   Abstract. We study the equation with a discontinuous non-linearity:

$$-\Delta u = \lambda H(u-1)$$

(H is Heaviside's unit function) in plane domains with various boundary conditions. We expect to find a curve dividing the harmonic ($\Delta u = 0$) region from the superharmonic ($\Delta u = -\lambda$) region, defined by the equation

$$u(x,y) = 1.$$

This curve is called the free boundary since its location is determined by the solution to the problem.

   We use the implicit function theorem to study the effect of perturbation of the boundary conditions on known families of solutions. This justifies rigorously a formal scheme derived previously. Our method also discovers bifurcations from previously known solution families. Finally, numerical methods for this problem are discussed.

   1. Introduction Let Ω be the unit square $\{(x,y) \mid 0 < x, y < 1\}$ in the x,y-plane, Γ_0 its left edge $\{(0,y) \mid 0 \le y \le 1$ and $\Gamma_1 = \partial\Omega\backslash\Gamma_0$ the rest of the boundary. Let H denote the Heaviside unit function

$$H(t) = \begin{cases} 0 & t < 0, \\ 1 & t > 0. \end{cases}$$

Fleishman and Mahar [FM] have posed a boundary value problem
with discontinuous nonlinearity almost equivalent to the
following:

$$
(1) \qquad
\begin{cases}
-\Delta u = \lambda H(u-1) & \text{in } \Omega, \\[2ex]
u\big|_{\Gamma_0} = h(y) \\[2ex]
\dfrac{\partial u}{\partial n}\Big|_{\Gamma_1} = 0,
\end{cases}
$$

where λ is a real parameter, and h is a given function. (When
in the second equation h(y) is replaced by $\lambda h(y)$, problem (1)
is equivalent to that in [FM]; see the remarks in Section 3
following the proof of Proposition 2.)

Equations resembling the PDE in (1) have been proposed for
models in plasma physics and thermal conduction problems [K].
Further references to work on differential equations with dis-
continuous nonlinearities may be found in [Ch].

Our problem is a free boundary problem: a typical solu-
tion u may be expected to define by the equation $u(x,y) = 1$ a
curve (across which u and its first derivatives are continuous)
which separates the region where $u < 1$ (and $\Delta u = 0$) from the
region where $u > 1$ (and $\Delta u = -\lambda$). The location of this separa-
ting curve, however, is not known beforehand; it is determined
by the solution itself.

In the next section we shall specify precisely what we mean
by a solution of (1). For the moment, we proceed informally.
To begin, we specialize to the boundary condition $h \equiv 0$, and
record the results of [FM], which motivated the present work.

The problem (1) with $h \equiv 0$ is called the reduced problem.
It always has the trivial solution $u \equiv 0$. When $\lambda > 4$ positive
solutions appear which depend on x only: for x_0 a solution of
the quadratic equation

$$
x_0(1-x_0) = \lambda^{-1}
$$

(that is $x_0 = x_0^{\pm}(\lambda) = \frac{1}{2} \pm \frac{1}{2}\sqrt{1-4\lambda^{-1}}$) we have either one ($\lambda = 4$) or two ($\lambda > 4$) solutions of the form

$$(2) \qquad u_0(x) = u_0^{\pm}(x) = \begin{cases} \lambda(1-x_0)x, & 0 \le x < x_0 \\ \\ \lambda[x-\frac{1}{2}(x^2+x_0^2)] & x_0 \le x \le 1, \end{cases}$$

whose graphs are shown in Figure 1. Note that $u_0^-(x) > u_0^+(x)$ for $0 < x \le 1$. Thus u_0^- (resp. u_0^+) corresponds to points on the upper (resp. lower) branch of the λ,u-curve in Figure 2.

The line $x = x_0$ is the free boundary where $u_0 = 1$. The solution and its gradient are continuous in the whole square, and the differential equation is satisfied in the classical sense in the regions where $u_0 < 1$ or $u_0 > 1$. It is not asserted that these are the only solutions.

We turn now to problem (1) with $h(y) \not\equiv 0$, which will be called the perturbed problem. In this case no closed-form solution is known. Under the assumption that solutions exist, a formal scheme was developed in [FM] to calculate first-order approximations to the solution (and associated free boundary) close to a given reduced solution u_0^-. The questions of existence of the solution, and the range of validity of the perturbation scheme, were left open.

In attempting to answer these questions, we have established the following theorem, to the proof of which the remainder of this paper is devoted.

THEOREM 1: __For__ $\lambda > 4$, __consider the solution__ u_0^- __of the reduced__ __problem__ (1) __corresponding to__ $x_0 = x_0^-(\lambda)$, __that is, on the "upper branch" of the curve in Figure 2.__ For all boundary data $h(y)$ __sufficiently close to zero in an appropriate function space, there is a unique solution u of (1) which depends continuously on h; this solution determines through the equation__ $u(x,y) = 1$ __a unique curve whose equation may be written__ $x = x_0^- + b(y)$.

__The perturbation__ $b(y)$ __of the free boundary depends continuously (Fréchet-) differentiably on the function h, and is given to first order by the perturbation scheme of__ [FM].

__For the "lower branch"__ $x_0 = x_0^+(\lambda)$, __all of the above assertions hold, provided that__ x_0 __does not belong to a certain sequence of exceptional values having__ $x_0 = 1$ __as their only limit point.__ For __each__ $n = 1,2,\ldots$ __there is an exeptional value of__ x_0, x_n, __at which the reduced problem has a bifurcation:__ __there are solutions of the reduced problem having free boundaries of the form__ $x = x_n + \alpha \cos n\pi y + o(\alpha)$ __for all__ α __in some neighborhood of zero.__

In the next section we specify the class of admissible boundary values $h(y)$; Section 3 gives the sequence of exceptional values of x_0 for which bifurcation occurs in the lower branch of solutions of the reduced problem.

The theorem asserts that, aside from the bifurcations, which surprised us, the perturbed problem has solutions which may be approximated by the scheme proposed in [FM]. In a way, the assertion is actually stronger: the admissible boundary value functions $h(y)$ have uniformly convergent Fourier cosine series, and it will follow from the proof of the theorem that the perturbation of the free boundary may be computed (to first order) term by term.

Note that continuously differentiable dependence on h is established for the free boundary, not for the solution u. This is because our method of proof is to reformulate problem (1) as a nonlinear integral equation for $b(y)$, the perturbation of the free boundary. While using the Green's function (see Section 2) to transform (1) into an integral equation for u leads formally to the same results, we have not been able to establish the estimates needed for a proof by this route.

In Section 2 of this paper we formulate a nonlinear integral equation for $b(y)$ and prove that by solving it we can solve problem (1). In Section 3 we use the implicit function theorem to solve our nonlinear integral equation, and show that bifurcation occurs when the implicit function theorem fails. To apply the implicit

function theorem we need to know that certain linearized expressions are Fréchet derivatives; the justifying estimates are provided in Section 4. Section 5 concludes with a discussion of current work on numerical methods for this problem, other geometries, and open questions.

2. Reformulation of Problem (1) as a Nonlinear Integral Equation.

For reasons that will emerge below, we restrict our attention to boundary data h in $C^{1,\alpha}([0,1])$ which satisfy $h'(0) = h'(1) = 0$. By a <u>solution</u> u of problem (1), for such an h, we mean a function

$$u \in C^1(\Omega) \cap C^2(\Omega \setminus \{(x,y) \mid u(x,y) = 1\}).$$

satisfying the boundary conditions in (1). (If one is content with a less regular solution, a milder assumption can be made on the data, e.g., that h be merely continuous.)

Let us suppose now that u is a solution of (1) but we have forgotten everything about u except the location of the free boundary. We ought to be able to recover u by replacing the Heaviside function in (1) by λ times the characteristic function of the region to the right of the free boundary, then solving the resulting linear Poisson problem. With a few technical assumptions this idea works, as we show presently.

It also suggests an appropriate space for the boundary data h. The source term in the Poisson equation belongs to every L^p space, so we choose any $p > 2$ and seek a solution in $W^{2,p}(\Omega)$. Such a function has a trace on the left boundary Γ_0 which is in $W^{2-\frac{1}{p},p}(\Gamma_0)$ (see [A]), so we require that h belong to this space.

To show that h is continuously differentiable, let

$$\sigma = 1 - \frac{1}{p}.$$

The norm of h (see [A]) is

$$\left\{ ||h||_{1,p,\Gamma_0}^p + \int_0^1 \int_0^1 \frac{|h'(x) - h'(y)|^p}{|x-y|^{1+\sigma p}} \, dx \, dy \right\}^{1/p} < \infty.$$

Hence $W^{2-\frac{1}{p},p} \subset W^{1,p}$, and it may be shown that h is absolutely continuous. Since $1 + \sigma p = p$ the integrand in the double integral is $|(h'(x)-h'(y))/(x-y)|^p$, and a result of A. Garsia [Ga] shows that

h' satisfies a uniform Hölder condition with exponent $1 - 2/p$.

The only difficulty in solving the Poisson problem to recover $u \in W^{2,p}$ is that singularities can appear in its derivatives at the corner of the square unless certain conditions are satisfied by the data. The method of [M] may be used to show that the compatibility condition in this case is

$$h'(0) = h'(1) = 0,$$

which makes sense because h' satisfies a uniform Hölder condition.

Finally, we do not want to "pull the free boundary around the corner", so we suppose that $h < 1$. We can now state conditions under which a solution of (1) can be recovered from knowledge of the free boundary only.

Let

$$A = \{h \in W^{2-\frac{1}{p},p}(\Gamma_0) \mid h < 1 \text{ and } h'(0) = h'(1) = 0\},$$

(3)

$$B = \{(\lambda,b) \in \mathbb{R} \times C(0,1) \mid \lambda > 4 \text{ and } 0 < x_0 + b(y) < 1$$
$$\text{for } 0 < y \leq 1\},$$

where x_0 may be either $x_0^+(\lambda)$ or $x_0^-(\lambda)$ but fixed.

PROPOSITION 1: <u>For</u> $h \in A$ <u>and</u> $(\lambda,b) \in B$ <u>the problem</u>

(4)
$$\begin{cases} -\Delta u = \lambda \chi_{\{(x,y) \mid x \geq x_0 + b(y)\}} \\ \\ u\big|_{\Gamma_0} = h, \ \frac{\partial u}{\partial n}\big|_{\Gamma_1} = 0 \end{cases}$$

<u>has a unique solution</u> $u \in W^{2,p}(\Omega) \subset C^{1,\alpha}(\overline{\Omega})$, $\alpha = 1 - \frac{2}{p}$ (see [A]). <u>More-over, if</u> h <u>and</u> b <u>are small enough and</u>

$$u(x_0+b(y), y) \equiv 1$$

<u>then</u> u <u>is a solution of</u> (1) <u>and actually</u> $b \in C^{1,\alpha}$.

PROOF: The existence, uniqueness and regularity of the solution of (4) follow from the conditions on the data, by the theory of the Poisson equation in a rectangle (see [Gr, M]). Thus we have only to show that if $u = 1$ on the putative free boundary, then u is actually a solution of (1) and $b \in C^{1,\alpha}$.

Suppose $u(x_0+b(y), y) = 1$. To establish u as a solution of the free boundary problem (1), it is enough to show that

$x < x_0 + b(y)$ (resp. $x > x_0 + b(y)$) implies $u < 1$ (resp. $u > 1$).

First consider the region $x < x_0 + b(y)$ to the left of the curve. Since u is a harmonic function there, its maximum must occur on the boundary; it cannot occur in the interior because u is not identically constant. (Recall that $u(0,y) = h(y) < 1$, $u(x_0 + b(y),y)=1$.) Obviously the maximum does not occur on the left edge of the square, nor on the top ($y=1$) or bottom ($y=0$), because $\partial u/\partial n = 0$ there; hence the maximum is 1 and is taken on only at the free boundary.

We use a similar argument for the part of the square lying to the right of the curve $x = x_0 + b(y)$; there u is a superharmonic function, and it may be seen that its minimum occurs only on the free boundary curve. The previous argument applies except at the corners (1,0) and (1,1). For these points we note that if h and b are small enough, u will be close in $W^{2,p}$ (hence in $C^{1,\alpha}$) to the solution of the problem (4) with $h \equiv 0$ and $b \equiv 0$, and it follows that $u(1,0) > 1$, $u(1,1) > 1$.

Finally, the solution of the reduced problem has $\partial u/\partial x$ bounded away from 0 near the free boundary. Hence if h and b are small enough, the solution of (4) has the same property. Since b satisfies $u(x_0+b(y),y) = 1$, it follows from the classical implicit function theorem that $b \in C^{1,\alpha}$.

Let us pursue further the observation with which we began this section. Given $h \in A$ and $\lambda > 4$, Proposition 1 shows that a solution of (1) may be obtained by finding $b \in C(0,1)$ such that $(\lambda,b) \in B$ and the solution u of (4) satisfies

$$u(x_0+b(y),y) = 1, \quad 0 \le y \le 1.$$

We now use Green's representation formula for u to obtain a nonlinear integral equation for the free boundary b.

297

To obtain the Green's function, we note first that the linear problem

$$\Delta u = \lambda u \quad \text{in } \Omega,$$

$$u\big|_{\Gamma_0} = 0, \quad \frac{\partial u}{\partial n}\Big|_{\Gamma_1} = 0,$$

has eigenvalues

$$\lambda_{mn} = -\pi^2 [(m+\tfrac{1}{2})^2 + n^2], \quad m, n \geq 0,$$

with corresponding eigenfunctions

$$u_{mn} = \begin{cases} 2^{\frac{1}{2}} \sin (m+\tfrac{1}{2})\pi x, & m \geq 0, \ n = 0, \\[2mm] 2 \sin (m+\tfrac{1}{2})\pi x \cos n\pi y, & m \geq 0, \ n \geq 1. \end{cases}$$

The Green's function for the Laplacian with these boundary conditions is given by the bilinear formula

$$G(x,y,\xi,\eta) = -\frac{2}{\pi^2} \sum_{m=0}^{\infty} \frac{\sin(m+\tfrac{1}{2})\pi x \ \sin(m+\tfrac{1}{2})\pi\xi}{(m+\tfrac{1}{2})^2}$$

(5)

$$-\frac{4}{\pi^2} \sum_{\substack{m>0 \\ n\geq 1}}^{\infty} \frac{\sin(m+\tfrac{1}{2})\pi x \ \cos n\pi y \ \sin(m+\tfrac{1}{2})\pi\xi \ \cos n\pi\eta}{(m+\tfrac{1}{2})^2 + n^2}$$

Now we write the solution of (4) with the aid of Green's representation formula

$$u(x,y) = \int_{\partial\Omega} (u \frac{\partial G}{\partial n} - G \frac{\partial u}{\partial n})ds + \int_{S} G \Delta u \ d\xi \ d\eta,$$

where n is the outward unit normal. Since $\partial/\partial n = - \partial/\partial\xi$ on the left boundary Γ_0 of the square and $\Delta u = 0$ to the left of $\xi = x_0 + b(\eta)$, the solution of problem (4) may be written

$$u(x,y) = - \int_0^1 G_\xi (x,y,0,\eta)h(\eta)d\eta$$

(6)

$$-\lambda \int_0^1 \int_{x_0+b(\eta)}^1 G(x,y,\xi,\eta)d\xi \ d\eta,$$

so that our constraint on the free boundary, $u(x_0+b(y),y)-1 = 0$, takes the form

298

$$-\int_0^1 G_\xi(x_0+b(y),y,0,\eta)h(\eta)\,d\eta \; - \; \lambda \int_0^1 \int_{x_0+b(\eta)}^1 G(x_0+b(y),y,\xi,\eta)\,d\xi\,d\eta \; - \; 1 \; = \; 0,$$

(7) $\hspace{7cm} 0 \le y \le 1.$

If $b(y)$ is a solution of this nonlinear integral equation then $x = x_0+b(y)$ is a free boundary for a solution of (1). In the next section we use the implicit function theorem to solve this equation for $b(y)$.

3. Solutions of the Equation for the Free Boundary.

Denote the left side of (7) by $F(h,\lambda,b)$. We regard F as an operator from $A \times B$, defined in (3), into $C(0,1)$, and we seek solutions $b(y)$ of the operator equation

(7)' $\hspace{4cm} F(h,\lambda,b) = 0,$

for each λ, in the neighborhood of the known solution $b(y) \equiv 0$ of

$$F(0,\lambda,b) = 0, \qquad \lambda > 4.$$

The results are described in the following theorem, from which Theorem 1 follows immediately.

THEOREM 2. For $\lambda > 4$, let $x_0 = x_0^-(\lambda)$ be the free boundary for the "upper solution" of the reduced problem. Then there is an open neighborhood U of $(0,\lambda)$ in $A \times \mathbb{R}$ and a unique continuously differentiable mapping $g: U \to C(0,1)$ such that $g(0,\lambda) = 0$ and $F(h,\lambda,g(h,\lambda)) = 0$ for $(h,\lambda) \in U$. The partial derivative $D_1 g$ matches the expression given by the perturbation scheme of [FM].

If $x_0 = x_0^+(\lambda)$, the free boundary for the "lower solution", then all the above results hold provided x_0 is not a solution of one of the equations

(8.n) $\quad 1 - x_0 = \dfrac{1}{n\pi} \dfrac{\sinh n\pi x_0 \cosh n\pi(1-x_0)}{\cosh n\pi} \qquad (n = 1,2,\ldots).$

Each of the equations (8.n) has a unique solution x_0 in the open interval $(\tfrac{1}{2},1)$. The solutions to these equations form a monotone sequence tending to 1, and at each such point x_0 a bifurcation occurs.

To be specific, fix $n \ge 1$ and let x_0 be the solution of equation (8.n) in $(\tfrac{1}{2},1)$. Let $\lambda_n = 1/x_0(1-x_0)$, and let Z be any complement of the linear span of $\cos n\pi y$ in $C(0,1)$. Then we find

i) <u>an interval</u> $I = (-\delta, +\delta)$,

ii) <u>continuous functions</u> $\phi: I \to \mathbb{R}$ <u>and</u> $\psi: I \to Z$ <u>with</u> $\phi(0) = \lambda_n$, $\psi(0) = 0$, and

iii) <u>a neighborhood</u> V <u>of</u> $(\lambda_n, 0)$ <u>in</u> $\mathbb{R} \times C(0,1)$,

<u>such that for all</u> t <u>in</u> I <u>the following pairs are solutions of</u> $\overline{F(0,\lambda,b) = 0}$ <u>in</u> V:

a) $(\lambda, b) = (\lambda_n + t, 0)$ (corresponding to symmetric solutions $u_0(x)$ of the reduced problem),

b) $(\cdot, b) \sim (\phi(t), t \cos n\pi(.) + t\psi(t))$ (the bifurcated solutions);

<u>and every solution in</u> V <u>has one of these forms.</u>

We give the proof as a sequence of lemmas; the proof of the perturbation result comes first, followed by the bifurcation proof.

For the perturbation result we use the implicit function theorem, which requires the existence, continuity and invertibility of $D_3 F$, the partial (Fréchet-) derivative of F with respect to its third argument. The information needed is established in the next three lemmas.

LEMMA 1. <u>For</u> $h \in A$ <u>and</u> $(\lambda, b) \in B$, <u>let</u> u <u>be the corresponding solution</u> <u>of</u> (4) <u>given by</u> (6). <u>Then</u> $D_3 F(h, \lambda, b)$ <u>is the linear operator in</u> $C(0,1)$ <u>given by</u>

(9) $D_3 F(h, \lambda, b) \cdot \beta(y) \sim \frac{\partial u}{\partial x}(x_0 + b(y), y) \cdot \beta(y)$

$$+ \int_0^1 G(x_0 + b(y), y, x_0 + b(\eta), \eta) \beta(\eta) \, d\eta$$

PROOF: We show here only that the expression given by (9) defines a bounded linear operator in $C(0,1)$. The estimates necessary to show that it is the (Fréchet-) derivative are given in Section 4. Incidentally, the same argument shows that $D_2 F$ is continuous in $A \times B$ as well.

For each fixed (h, λ, b), the operator (9) is a multiplication operator by a bounded continuous function, plus an integral operator. The multiplication part is evidently a bounded operator. The integral operator is even compact, as we now show. Call the kernel of the integral operator $k(y, \eta)$. It is of the form

$$k(y,\eta) = \frac{1}{2\pi} \log \frac{1}{[(b(y)-b(\eta))^2+(y-\eta)^2]^{\frac{1}{2}}} + \text{analytic function.}$$

If we also denote the integral operator by k we have

$$|k\beta(y') - k\beta(y)| = |\int_0^1 [k(y',\eta) - k(y,\eta)]\beta(\eta)d\eta|$$

$$\leq ||\beta||_\infty \int_0^1 |k(y',\eta) - k(y,\eta)|d\eta$$

$$\leq ||\beta||_\infty \omega(y'-y),$$

where $\omega(x) \to 0$ as $x \to 0$. Thus k takes bounded sequences into uniformly bounded and equicontinuous sequences.

LEMMA 2. D_3F is a continuous mapping of $A \times B$ into the space of bounded linear operators in $C(0,1) = C^0$.

PROOF: The continuous dependence of $D_3F(h,\lambda,b)$ on its arguments jointly is easily seen as follows. First, the solution u of (4) depends continuously in the norm of $C^{1,\alpha}$ on the data h,λ,b. Next, the mapping

$$(u,\lambda,b) \in C^{1,\alpha} \times \mathbb{R} \times C^0 \longmapsto \frac{\partial u}{\partial x}(x_0(\lambda) + b(y),y) \in C^0$$

is evidently continuous, proving the continuity of the multiplication part of D_3F. For the continuity of the integral operator, we observe that the norm of the difference of the integral operators in $D_3F(h,\lambda,b)$ and $D_3F(h',\lambda',b')$ is dominated by

$$\max_y \int_0^1 |G(x_0+b(y),y,x_0+b(\eta),\eta)-G(x_0'+b'(y),y,x_0'+b'(\eta),\eta)|d\eta,$$

where $x_0' = x_0(\lambda')$. The integrand is continuous as a function of all its arguments outside an arbitrarily small neighborhood of $\eta = y$, over which the integral is as small as we please because the singularity is only logarithmic. This proves continuous dependence.

LEMMA 3. $D_3F(0,\lambda,0) = \lambda[(1-x_0)I-K]$, where K is the compact integral operator in $C(0,1)$ with kernel

$$K(y,\eta) = \sigma_0 + 2 \sum_{n=1}^\infty \sigma_n \cos n\pi y \cos n\pi\eta.$$

301

The eigenvalues σ_n are given by

$$\sigma_0 = x_0$$

$$\sigma_n = \frac{1}{n\pi} \frac{\sinh n\pi x_0 \cosh n\pi(1-x_0)}{\cosh n\pi}, \qquad n = 1,2,\ldots.$$

PROOF: When $h \equiv 0$ and $b \equiv 0$ we find from (2) and (9) that

$$D_3 F(0,\lambda,0) = \lambda[(1-x_0)I - K],$$

where K is the integral operator with kernel

$$K(y,\eta) = -G(x_0, y, x_0, \eta).$$

By substituting into the Green's function (5) we find that the eigenfunctions are $\{\cos n\pi y: n \geq 0\}$ and that the eigenvalues are given by

$$\sigma_0 = \frac{2}{\pi^2} \sum_{m=0}^{\infty} \frac{\sin^2(m+\tfrac{1}{2})\pi x_0}{(m+\tfrac{1}{2})^2} = x_0,$$

$$\sigma_n = \frac{2}{\pi^2} \sum_{m=0}^{\infty} \frac{\sin^2(m+\tfrac{1}{2})\pi x_0}{(m+\tfrac{1}{2})^2 + n^2}$$

$$= \frac{1}{n\pi} \frac{\sinh n\pi x_0 \cosh n\pi(1-x_0)}{\cosh n\pi}, \quad n \geq 1.$$

Interchange of the order of summation in this calculation is justified by the fact that outside any neighborhood of $(x,y) = (\xi,\eta)$ the series for the Green's function converges uniformly; the resulting series for $K(y,\eta)$ converges uniformly away from $y = \eta$.

In order to compute the linear approximation of the perturbed free boundary and compare it with the calculation of [FM], we also need $D_1 F$.

LEMMA 4. $D_1 F(0,\lambda,0)$ is the integral operator from

$$X = \{h \in W^{2-\frac{1}{p},P}(\Gamma_0) | h'(0) = h'(1) = 0\} \text{ into } C(0,1) \text{ with the kernel}$$

$$1 + \sum_{n=1}^{\infty} \rho_n \cos n\pi y \cos n\pi\eta,$$

where

$$\rho_n = \cosh n\pi (1-x_0)/\cosh n\pi, \quad n = 1,2,\ldots.$$

PROOF: The operator F depends affinely on h, so D_1F is the integral
operator from X into $C(0,1)$ given by

$$D_1F(h,\lambda,b)\cdot\hat{h}(y) = -\int_0^1 G_\zeta (x_0+b(y),y,0,\eta)\hat{h}(\eta)\,d\eta.$$

Since $x_0+b(y) > 0$ the kernel is analytic, so D_1F is compact. The
computation of the eigenvalues proceeds exactly as for D_3F.

The results up to now have established that F is a continuously
differentiable mapping of $A\times B$ into $C(0,1)$. We now wish to apply the
implicit function theorem to solve the equation

$$F(h,\lambda,b) = 0$$

for b as a function of h and λ, say $g(h,\lambda)$, such that $g(0,\lambda) = 0$.
This can be done if $D_3F(0,\lambda,0)$ is invertible. From Lemma 3 it
follows that this is the case whenever $1 - x_0$ is not equal to any
of the σ_n, eigenvalues of K. Now it is easy to see that
$\sigma_n < \frac{1}{2}$, $n = 1,2,\ldots$; hence if we take $x_0 = x_0^-(\lambda)$, corresponding to
the upper solution (see Fig. 2), $1 - x_0 > \frac{1}{2} > \sigma_n$ and $D_3F(0,\lambda,0)$ is
always invertible. This proves the perturbation result for the
upper solutions.

For the lower solutions $x_0 = x_0^+(\lambda) > \frac{1}{2}$; so $1 - x_0 < \frac{1}{2}$,
and $1 - x_0$ can coincide with an eigenvalue of K. In fact this
happens just once for each n, as we now show.

PROPOSITION 2. For each $n = 0,1,2,\ldots$ there is a $\lambda_n \geq 4$ such
that

$$1 - x_0 = \sigma_n,$$

x_0 being taken with the "+" sign. The λ_n form an increasing
sequence with no finite point of accumulation.

PROOF: Since $\sigma_0 = x_0$, we have $\sigma_0 = 1 - x_0$ when $x_0 = \frac{1}{2}$, $\lambda_0 = 4$.
This corresponds to the appearance of the nontrivial solution,

followed by its splitting, for $\lambda > 4$, into an upper and a lower solution. For $n \geq 1$, we use the formula

$$\sigma_n = \frac{1}{n\pi} \frac{\sinh n\pi}{\sinh 2n\pi} (\sinh n\pi + \sinh n\pi (2x_0 - 1))$$

to see that as x_0 increases from $\frac{1}{2}$ to 1, σ_n increases from $\frac{1}{2n\pi} \tanh n\pi$ to $\frac{1}{n\pi} \tanh n\pi$, while $1 - x_0$ decreases from $\frac{1}{2}$ to 0. Hence there is a unique solution of the equation $1 - x_0 = \sigma_n$. The inequality $x_0 > 1 - \frac{1}{n\pi}$ for the solution of this equation shows that $x_0 \to 1$, therefore λ_n tends to infinity, as $n \to \infty$.

We may now verify that the scheme of [FM] gives b correctly to first order in h, for those points (λ, u_0), $\lambda > 4$, on the graph of Figure 2 for which $D_3F(0,\lambda,0)$ is invertible; namely, for the entire upper branch and for all points (λ, u_0^+) on the lower branch except the ones covered by Proposition 2. Since

$$F(h,\lambda,g(h,\lambda)) = 0$$

it follows that

$$D_1g = -(D_3F)^{-1}D_1F.$$

Now any $h \in A$ has a uniformly convergent Fourier cosine series

$$h(y) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\pi y.$$

Hence the linear approximation to $b = g(h,\lambda)$ is

$$D_1g(0,\lambda) \cdot h = -\frac{a_0}{2(1-2x_0)} - \sum_{n=1}^{\infty} \frac{\rho_n}{1-x_0^{-\sigma_n}} a_n \cos n\pi y$$

This matches the calculation in [FM] when h is replaced by λh; see the remark following equation (1).

Now we return to the situation when $1-x_0 = \sigma_n$ for some $n = 1,2,\ldots$ We do not attempt to describe what happens at $x_0 = \frac{1}{2}$.

To show that bifurcation occurs, and thus establish the second part of Theorem 2, we apply a bifurcation theorem of Crandall and Rabinowitz [CR]. The hypotheses of that theorem are established in

the three lemmas which follow.

LEMMA 5. <u>Let</u> x_0 <u>satisfy the equation</u> $1-x_0 = \sigma_n$ <u>for some n and</u> <u>let</u> $\lambda_n = 1/x_0(1-x_0)$ <u>Then</u> $D_3F(0,\lambda_n,0)$ <u>has a one-dimensional null</u> <u>space, spanned by</u> $\cos n\pi y$, <u>while its range has codimension one,</u> <u>coinciding with the null space of the continuous linear functional</u>

$$\Phi_n(f) = \int_0^1 f(y) \cos n\pi y \, dy.$$

PROOF: From the form of $K(y,\eta)$ (see Lemma 3) $D_3F(0,\lambda_n,0)$ annihi-
lates $\cos n\pi y$ when $1-x_0 = \sigma_n$; and since the σ_n's are distinct,
$\cos m\pi y$ is not in the null space for $m \neq n$. From the Schauder
theory for compact operators the range has codimension one. For
the characterization of the range, observe that
$\Phi_n(D_3F(0,\lambda_n,0)\cdot\beta) = 0$ for any $\beta \in C(0,1)$; since the range of
$D_3F(0,\lambda_n,0)$ is a hyperplane it coincides with the null space of Φ_n.

LEMMA 6. D_2D_3F <u>exists and is continuous in a neighborhood of</u>
$(0,\lambda_n,0)$. (We give the proof of this lemma in Section 4.)

LEMMA 7. $D_2D_3F(0,\lambda_n,0)\cdot\cos n\pi(.)$ <u>does not belong to the range of</u>
$D_3F(0,\lambda_n,0)$.

PROOF: The proof of Lemma 6 shows that $\cos n\pi(\cdot)$ is an eigenfunction
of $D_2D_3F(0,\lambda_n,0)$. The eigenvalue is

$$\left[\frac{d}{d\lambda}(\lambda(1-x_0-\sigma_n))\right]_{\lambda=\lambda_n} = \left[-\lambda\frac{dx_0}{d\lambda}(1+\frac{d\sigma_n}{dx_0})\right]_{\lambda=\lambda_n}$$

$$= \left[-\lambda\frac{dx_0}{d\lambda}(1+\frac{\cosh n\pi(2x_0-1)}{\cosh n\pi})\right]_{\lambda=\lambda_n} \neq 0.$$

This completes the proof of Lemma 7, and of Theorem 2.

4. Estimates for the Derivative Calculations. In this section
we prove estimates to show that D_3F and D_2D_3F have the analytic forms
given in Lemma 1 and Lemma 6, respectively.

We begin with D_3F. Let $(h,\lambda,b) \in A\times B$ and let $\beta \in C(0,1)$ such that
$(h,\lambda,b+\beta) \in A\times B$. Denoting by L the operator on the right side of (9),
and using (6) to express $\partial u/\partial x$, we have

$$[F(h,\lambda,b+\beta) - F(h,\lambda,b) - L\beta](y) =$$

$$-\int_0^1 G_\xi(x_0+b(y)+\beta(y),y,0,\eta)h(\eta)\,d\eta - \lambda\int_0^1 \int_{x_0+b(\eta)+\beta(\eta)}^1 G(x_0+b(y)+\beta(y),y,\xi,\eta)\,d\xi\,d\eta$$

$$+\int_0^1 G_\xi(x_0+b(y),y,0,\eta)h(\eta)\,d\eta + \lambda\int_0^1 \int_{x_0+b(\eta)}^1 G(x_0+b(y),y,\xi,\eta)\,d\xi\,d\eta$$

$$+\left[\int_0^1 G_{x\xi}(x_0+b(y)y,0,\eta)h(\eta)\,d\eta + \lambda\int_0^1 \int_{x_0+b(y)}^1 G_x(x_0+b(y),y,\xi,\eta)\,d\xi\,d\eta\right]\cdot\beta(y)$$

$$-\lambda\int_0^1 G(x_0+b(y),y,x_0+b(\eta),\eta)\beta(\eta)\,d\eta.$$

We need to show that the norm in $C(0,1)$ of this difference, divided by $||\beta||_\infty$, may be made arbitrarily small by choosing $||\beta||_\infty$ sufficiently small. Rearranging terms, as in the usual proof of the Leibnitz Rule, we find that the above expression is equal to

$$\int_0^1 h(\eta) \, d\eta \left[- G_\xi(x_0+b(y)+\beta(y),y,0,\eta) + G_\xi(x_0+b(y),y,0,\eta) \right.$$

$$\left. + \beta(y) \, G_{x\xi}(x_0+b(y),y,0,\eta) \right]$$

$$+ \lambda \int_0^1 d\eta \int_{x_0+b(\eta)}^1 d\xi \left[-G(x_0+b(y)+\beta(y),y,\xi,\eta) + G(x_0+b(y),y,\xi,\eta) \right.$$

$$\left. + \beta(y) \, G_x(x_0+b(y),y,\xi,\eta) \right]$$

$$+ \lambda \int_0^1 d\eta \left[\int_{x_0+b(\eta)}^{x_0+b(\eta)+\beta(\eta)} G(x_0+b(y)+\beta(y),y,\xi,\eta) d\xi - \beta(\eta) \, G(x_0+b(y),y,x_0+b(\eta),\eta) \right] \; .$$

Each of the first two integrals is of the form

$$f(x_0+b(y)+\beta(y),y) - f(x_0+b(y),y) - \beta(y) \, f_x(x_0+b(y),y)$$

for a $C^{1,\alpha}$ function f, and from Taylor's Theorem it follows that for any $\varepsilon > 0$ there is a $\delta > 0$ so small that

$$\max_{0 \le y \le 1} \left| f(x_0+b(y)+\beta(y),y) - f(x_0+b(y),y) \right.$$

$$\left. - \beta(y) \, f_x(x_0+b(y),y) \right| < \varepsilon ||\beta||_\infty$$

provided $||\beta||_\infty < \delta$.

For the last integral we add and subtract a term to obtain

$$(10) \quad \lambda \int_0^1 d\eta \left[\int_{x_0+b(\eta)}^{x_0+b(\eta)+\beta(\eta)} G(x_0+b(y),y,\xi,\eta) d\xi - \beta(\eta) G(x_0+b(y),y,x_0+b(\eta),\eta) \right]$$

$$+ \lambda \int_0^1 d\eta \int_{x_0+b(\eta)}^{x_0+b(\eta)+\beta(\eta)} d\xi \left[G(x_0+b(y)+\beta(y),y,\xi,\eta) - G(x_0+b(y),y,\xi,\eta) \right].$$

The first of these may be written

$$\lambda \int_0^1 d\eta \int_{x_0+b(\eta)}^{x_0+b(\eta)+\beta(\eta)} d\xi \left[G(x_0+b(y),y,\xi,\eta) - G(x_0+b(y),y,x_0+b(\eta),\eta) \right].$$

Let $\varepsilon > 0$ be given. We show that the integral above is bounded in maximum norm (considered as a function of y), by a constant times ε times the maximum norm of β. We choose $||\beta||_\infty$ so small that

$$||\beta||_\infty^{1/4} < \varepsilon, \text{ and } ||\beta||_\infty^{1/2} \log (1/||\beta||_\infty) < \varepsilon,$$

and let $\gamma = ||\beta||_\infty^{3/4}$.

For each fixed y, $0 < y < 1$, let $D(y)$ denote the disk of radius γ about the point $(\xi,\eta) = (x_0+b(y),y)$. Letting S denote the region of integration and $I(\xi,\eta)$ the integrand we have

$$\iint_S I(\xi,\eta) \, d\xi d\eta = \iint_{S \cap D(y)} I \, d\xi d\eta = \iint_{S \setminus D(y)} I \, d\xi d\eta$$

There are constants C and C' independent of y such that

$$\left| \iint_{S \cap D} I \, d\xi d\eta \right| \leq C' \int_0^\gamma \log (1/r) \, r dr \leq C \gamma^2 \log (1/\gamma),$$

where we have taken $r^2 = [\xi - (x_0+b(y))]^2 + (\eta-y)^2$. Outside $D(y)$ we estimate the integrand by the mean value theorem to obtain, with constants C_1 and C_1', independent of y,

$$\left| \iint_{S \setminus D(y)} I \, d\xi d\eta \right| \leq C' \iint_{S \setminus D} (1/r) \, ||\beta||_\infty \, d\xi d\eta \leq C(1/\gamma) \, ||\beta||_\infty^2$$

from the fact that $\iint_S 1 \leq ||\beta||_\infty$. Combining these estimates with (11) we obtain the desired estimate for the integral.

The second integral in (10) is estimated the same way. This completes the proof of Lemma 1.

Finally we turn to the proof of Lemma 6. Let us write the Green's function in the form

$$\frac{1}{2\pi} \log \frac{1}{r} + Q(x,y,\xi,\eta),$$

308

where $r = [(x-\xi)^2 + (y-\eta)^2]^{1/2}$, and Q has no singularities in the square. Then $D_3 F$ may be put in the form

$$D_3 F(h,\lambda,b) = M + \lambda L_1,$$

where M is the operator of multiplication by the function

(12)
$$\frac{\partial u}{\partial x}(x_0 + b(y), y)$$

and L_1 is the integral operator with kernel

$$L_1(y,\eta) = \frac{1}{2\pi} \log \frac{1}{\sqrt{(b(y)-b(\eta))^2 + (y-\eta)^2}} + N(x_0 + b(y), y, x_0 + b(\eta), \eta),$$

where N is analytic.

The dependence on λ is only through x_0. Thus the singularity does not depend on λ, and $\partial(\lambda L_1)/\partial\lambda$ may be computed by differentiating the kernel. The operator $\partial(\lambda L_1)/\partial\lambda$ depends continuously on λ and b. To see that $D_2 D_3 F$ exists and is continuous, it remains to be shown that the function (12) may be differentiated with respect to λ. (Recall that x_0 is determined by λ.) Since

$$\frac{\partial u}{\partial x}(x_0 + b(y), y) = - \int_0^1 G_{x\xi}(x_0 + b(y), y, 0, \eta) h(\eta) \, d\eta$$

$$- \lambda \int_0^1 \int_{x_0 + b(\eta)} G_x(x_0 + b(y), y, \xi, \eta) \, d\xi d\eta,$$

the only problem is the differentiated logarithm in the second integral. This term is

$$\int_0^1 \int_{x_0 + b(\eta)}^1 \frac{d}{dx} \log \frac{1}{r} \Big|_{(x,y) = (x_0 + b(y), y)} d\xi d\eta = - \int_0^1 \log \frac{1}{r} \Big|_{\substack{\varepsilon = 1 \\ \xi = x_0 + b(\eta)}} d\eta$$

$$= - \int_0^1 \log \left(\frac{1}{\sqrt{(x_0 + b(y) - 1)^2 + (y-\eta)^2}} \right) d\eta + \int_0^1 \log \left(\frac{1}{\sqrt{(b(y) - b(\eta))^2 + (y-\eta)^2}} \right) d\eta$$

showing that differentiation by x_0 may be performed under the integral sign.

5. Further Work and Open Problems.

The obvious iteration scheme for numerical solution of our problem (1) is probably

$$- \Delta u_{n+1} = \lambda H(u_n - 1).$$

But a simple analysis of the one-dimensional problem

$$- u_{xx} = \lambda H(u-1), \quad u(0) = h, \quad u'(1) = 0$$

shows that this scheme converges only to the upper solutions. This suggests that to find the lower solutions of the two-dimensional problem, some other scheme must be tried. A further problem is that the solution u is necessarily not globally smooth. This implies that some kind of adaptive scheme must be used to refine the mesh near the free boundary. Our work on this approach is continuing.
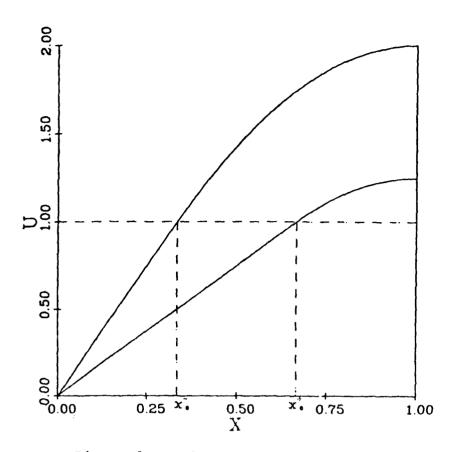
Our method may also be applied to the following problem, considered in [FM2]. In polar coordinates (r, θ),
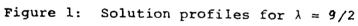
$$- \Delta u = \lambda H(u-1) \text{ in } D, \text{ the unit disk in the plane,}$$
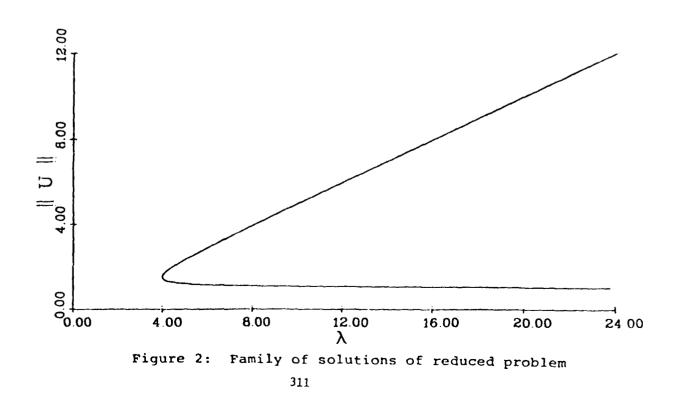
$$u(1, \theta) = h(\theta), \quad 0 \leq \theta \leq 2\pi,$$

with $h \geq 0$ and small. Classical solutions are constructed in [FM2] by means of a monotone iteration scheme. When the method of the present work is used, the assumption $h \geq 0$ may be dropped; it turns out further that there are no bifurcations from the family of radially symmetric solutions of the reduced problem, $h \equiv 0$. Details of these results will appear elsewhere.

We turn to some questions left unanswered by our present approach. First, the smoothness of perturbed or bifurcated free boundaries is of interest. Our method yields a free boundary of class $C^{2-\varepsilon}$. We conjecture that it is analytic. A second problem is the effect of perturbation at the bifurcation points: if the boundary values $h(y)$ are nonzero, what happens to the bifurcated solutions? Some of the ideas of [S], in particular a coordinate transformation to "straighten out the free boundary," may lead to answers to both of these problems.

Figure 1: Solution profiles for $\lambda = 9/2$

Figure 2: Family of solutions of reduced problem

311

References

[A] Robert A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[Ch] Kung-ching Chang, On the multiple solutions of the elliptic differential equations with discontinuous nonlinear terms, *Scientia Sinica* XXI, No. 2 (March-April, 1978) 139-158.

[CR] Michael G. Crandall and Paul H. Rabinowitz, Bifurcation from Simple Eigenvalues, *Journal of Functional Analysis* 3 (1971) 321-340.

[D] Jean Dieudionné, *Foundations of Modern Analysis*, Academic Press, New York, 1960.

[FM] Bernard A. Fleishman and Thomas J. Mahar, Analytic Methods for approximate solution of elliptic free boundary problems, *Nonlinear Analysis*, I No. 5 (1977) 561-569.

[FM2] _____ , On the existence of classical solutions to an elliptic free boundary problem, *Differential Equations and Applications*, W. Eckhaus and E. M. deJager, eds., North-Holland Publishing Co., 1978.

[Ga] Adriano M. Garsia, Combinatorial inequalities and smoothness of functions, Bull. AMS 82 No. 2 (March, 1976) 157-170.

[Gr] Pierre Grisvard, Behavior of the solutions of an elliptic boundary value problem in a polygonal or polyhedral domain, in *Numerical Solution of Partial Differential Equations* - III, Bert Hubbard, ed., Academic Press, New York, 1976.

[K] Hendrick J. Kuiper, Eigenvalue Problems for Noncontinuous Operations associated with Quasilinear Equations, *Arch. Rat. Mech. Anal.* 53 (1974) 178-186.

[M] M. Merigot, Étude du probléme $\Delta u = f$ dans un polygone plan. Inégalités á priori, Bollettino U.M.I. (4) 10 (1974) 577-597.

[S] Jan Sijbrand, Bifurcation Analysis for a class of problems with a free boundary, *Nonlinear Analysis* Vol. 3 No. 6 (1979) 723-753.

HOPF BIFURCATION COMPUTATIONS FOR
DISTRIBUTED PARAMETER SYSTEMS

R. F. Heinemann and A. B. Poore
Mathematics Research Center
University of Wisconsin-Madison
Madison, Wisconsin 53706

ABSTRACT. The development of numerical procedures for the computation of the Hopf bifurcation formulas are described for systems of parabolic partial differential equations whose time-independent solutions are defined by two-point boundary value problems. This class of problems was motivated by our efforts to understand the oscillatory dynamics of problems in combustion theory and chemical reactor theory. However, the Hopf formulas that are used are derived from the theory for evolution equations and are thus more widely applicable.

I. INTRODUCTION. The operation to bring about the ignition and extinction processes in chemical reactor and in combustion problems can be achieved by varying one or more of the physical parameters. This variation of a system parameter often leads to an exchange in the stability of a steady state resulting from a real eigenvalue of an appropriate linearized problem passing through zero or a pair of complex conjugate eigenvalues crossing the imaginary axis. In chemical reactor and combustion theory, the former case often gives rise to limit point bifurcations [1] and the ignition and extinction processes, and the second, to oscillatory instabilities or the appearance of bifurcating stable oscillatory states. The determination of the ensuing dynamics in this latter case can be based on the Hopf bifurcation theory. (See, e.g. [2, 3, 4] and the references therein.) The application of that theory to problems in chemical reactor and combustion theory in all but the most trivial cases is, in our opinion, best achieved through numerics, and it is this aspect of the problem that we investigate here. Specifically, we numerically treat systems of parabolic partial differential equations whose time-independent solutions are defined by systems of two-point boundary value problems. What makes these computations even more worthwhile is that the computed information yields a systematic method of investigating the stable oscillatory states. After discussing the general class of problems and the numerical techniques, we present an example from chemical reactor theory to demonstrate the types of information that are obtainable from the Hopf bifurcation theory. A more extensive discussion of this particular problem will appear elsewhere.

II. NUMERICAL METHODS. We now present the numerical techniques for generating the steady state response curves and the Hopf bifurcation information for systems whose mathematical description is given by a distributed parameter model. The coupling of two methods is used to systematically locate all the possible steady and periodic states exhibited by these systems. We begin with a discussion of the general forms of the equations that we are capable of treating.

Many of the finite systems that arise in the modeling of these systems can be written as a system of nonlinear partial differential equations of the form

$$\frac{\partial u}{\partial t} = A\left(\;\right)\frac{\partial^2 u}{\partial x^2} + f(u, u_x, \;)$$

$$G_l(u(a), u_x(a), \;) = 0 \tag{1}$$

$$G_r(u(b), u_x(b), \;) = 0$$

where u may represent, for example, the concentration and temperature in the reactor. G_l and G_r denote the left and right boundary conditions (possibly nonlinear) and $\;$ is the bifurcation parameter. The functions f may depend nonlinearly on u, u_x, and $\;$ and may contain the space variable x explicitly but not the time variable t. For brevity we write this system as

$$\frac{du}{dt} = F(u, \;), \qquad G(u, \;) = 0. \tag{2}$$

The corresponding two-point boundary value problem is denoted by

$$F(v, \;) = 0, \qquad G(v, \;) = 0. \tag{3}$$

III. STEADY STATE NUMERICAL TECHNIQUES. The steady state problem (3) is solved by combining Keller's modification of the Euler-Newton continuation method [1] with a Bador and Wiess' spline collocation code [5] and a fourth-order finite difference scheme due to Stepleman [6]. Both the collocation and finite difference methods perform quite well with the collocation technique capable of higher accuracy.

The objective of the steady state methods is to compute the solution v of equation (3) as the parameter $\;$ assumes all its possible values. The Euler-Newton continuation method can be used to solve this problem with Euler's method (4) serving as a predictor for Newton's method (5)

$$v^0(\; + \;) = v(\;) + \;\frac{du(\;)}{d\;} \tag{4}$$

$$v^{i+1}(\; + \;) = v^i(\; + \;) - F_v^{-1}(v^i(\; + \;), \; + \;)F(v^i(\; + \;), \; + \;). \tag{5}$$

However, the technique fails near transition between steady states since the Jacobian matrix, F_v, cannot be inverted near singularities such as limit or bifurcation points.

Keller has modified the above method by imposing an additional normalization on the solution which enables entire solution branches to be traced, skipping over any singular points. The imposition of this constraint allows the specification of a new parameter, s, which replaces $\;$ as the continuation parameter in the Euler-Newton technique. The reparameterized problem becomes

$$P(x, s) = 0 \tag{6}$$

where

$$x(s) = (v(s), \mu(s)) \qquad (7)$$

and

$$P(x(s), s) = \begin{bmatrix} F(v(s), \mu(s)) \\ N(v(s), \mu(s), s) \end{bmatrix}.$$

It is convenient to choose the normalization $N(v, \mu, s)$ so that s approximates the arc length of the solution branch for some parameter $\alpha \in (0, 1)$

$$N(v, \mu, s) = \alpha \|v(s) - v(s_0)\|^2 + (1 - \alpha)|\mu(s) - \mu(s_0)|^2 - (s - s_0)^2. \qquad (8)$$

When the Euler-Newton technique is applied to (6), computational difficulties near singularities are eliminated since the Jacobian matrix remains non-singular near such points.

These techniques and an algorithm for their implementation are presented in detail by Keller [1] and have previously been applied to laminar flame problems [7-10] and catalysts problems [11] which exhibit multiple steady states. Therefore, we forego a discussion of the exact computational procedure.

IV. HOPF BIFURCATION FORMALISM. We use the term formalism here since the presentation will be stripped of the technical mathematical assumptions. A proper mathematical framework can be found in the work of Crandall and Rabinowitz [3] and the references therein. Our work follows closely the presentation of Iooss and Joseph [4], though modified somewhat to account for the nonzero steady state problem and the form of the model equations (2). Since this bifurcation theory is most effectively used in a study of the dynamics associated with an exchange of stability, we begin with a brief discussion of steady state stability.

The stability of time-independent solutions can in principle be resolved by examining the eigenvalues of the linearized boundary value problem. If the eigenvalues all have negative real parts, the steady state is stable; whereas, if an eigenvalue has a positive real part, the solution is unstable. Exceptions to this principle occur when the linearized problem has a zero eigenvalue or a pair of complex conjugate, purely imaginary eigenvalues. In the current reactor problem, a zero eigenvalue gives rise to a limit point bifurcation (a point of vertical tangency) on the response curves. The bifurcation of a periodic solution (Hopf bifurcation) occurs when a pair of complex conjugate eigenvalues $\sigma(\mu)$ and $\bar{\sigma}(\mu)$ become purely imaginary. We assume that this crossing of the imaginary axis occurs at μ_0 so that $\sigma(\mu_0) = +i\omega$ with ω positive. It is also assumed that Re $\sigma_\mu(\mu_0) \neq 0$ where $\sigma_\mu = d\sigma/d\mu$. This ensures a strict crossing of the axis and is nearly always satisfied in these problems.

When the Hopf point is accompanied by a change in stability, different types of periodic behavior may be encountered. If the periodic orbit is stable, a small amplitude oscillation is observed near the Hopf point, but if the orbit is unstable, the solution will jump to either a large amplitude, stable oscillation or to another stable steady state. Examination of the orbit stability near the Hopf point suggests a systematic procedure for locating the stable oscillations in the reactor.

To present the Hopf bifurcation formulas, we linearize the initial value problem and write it as

$$L_\mu w = 0, \quad G_v(v^\mu, \mu|w) = 0 \qquad (\;)$$

where

$$L_\mu w = F_v(v^\mu, \mu|w) = \frac{\partial}{\partial \delta}F(v^\mu + \delta w, \mu)|_{\delta=0}$$

and

$$G_v(v^\mu, \mu|w) = \frac{\partial}{\partial \delta}G(v^\mu + \delta w, \mu)|_{\delta=0}.$$

The essential requirements for Hopf bifurcation without their technical assumptions may be summarized as follows. Assume $\pm i\omega$ are simple eigenvalues of L_{μ_0}, that $ni\omega$ is not an eigenvalue for $n = 0, 2, 3, \ldots,$ and that the real part of $\xi(\mu_0)$ is nonzero. Then one can construct a bifurcating periodic solution of (1) with frequency $\omega(\epsilon)$ via a perturbational expansion which can be shown to take the form [4]

$$u(x, t) = v^{\mu_0} + 2\epsilon \text{Re}\{\zeta_0 e^{is}\} + \frac{\epsilon^2}{2}\{w_1 - \mu_2\frac{dv^{\mu_0}}{d\mu} + 2\text{Re}\{w_2 e^{2is}\}\} + o(\epsilon^2), \quad (\;)$$

$$\mu = \mu_0 + \frac{\epsilon^2}{2}\mu_2 + o(\epsilon^2) \qquad (10)$$

$$w(\epsilon) = \omega_0 + \frac{\epsilon^2}{2}\omega_2 + o(\epsilon^2) \qquad (11)$$

$$t = \omega(\epsilon)s \qquad (12)$$

where ϵ is an auxiliary parameter representing the amplitude of the orbit close to the Hopf point. The vector function ζ_0 is the eigenvector corresponding to the eigenvalue $+i\omega$; w_1 and w_2 are solutions of certain linear nonhomogeneous boundary value problems discussed in the next section. We note that the sign of μ_2 yields the sign of $\mu - \mu_0$ for ϵ sufficiently small and therefore determines the direction of bifurcation. Similarly, ω_2 determines the change in the frequency of the bifurcating solution. The perturbational expansion (9) provides a good approximation to the periodic solution for computational purposes.

The stability of the bifurcating oscillation can be based on a study of the Floquet exponents as discussed by Iooss and Joseph [3]. The essential result is that the periodic solution will be locally stable near the bifurcation point if the eigenvalues of L_{μ_0}, other than $\pm i\omega$, have negative real parts and if $\mu_2 \text{Re}\, \xi(\mu_0)$ is positive.

316

The relevant [illegible] functions [illegible] that can be [illegible]
$_\mu(\ldots)$ are computed. The algorithm for computing each [illegible] is [illegible]
next.

V.   AN ALGORITHM [illegible] THE [illegible] BIFURCATION [illegible] The nonlinear
two-point boundary value problem

$$F(v^{\prime\prime}, \ldots) = 0, \quad G(v^{\prime}, \ldots) = 0 \tag{13}$$

must first be solved at [illegible] where the linearized [illegible] [illegible] a
pair of purely imaginary eigenvalues [illegible] [illegible] the most [illegible] [illegible] [illegible]
by finding a root [illegible] of the [illegible]. This [illegible] uses [illegible] [illegible] [illegible]
algorithm to compute the eigenvalues at each [illegible] point on the [illegible] [illegible]
response curves and then applying the [illegible], [illegible] [illegible] [illegible]
method to locate [illegible].

Let $L^*_{\ldots}$ denote the adjoint differential operator and $G^*_{\ldots}$ [illegible] the
adjoint boundary conditions. The eigenvectors [illegible] and [illegible] are then computed
from the eigenvalue problems

$$L_{\ldots}\phi = i\omega_0\phi, \quad G_v(v^{\prime\prime}, \ldots) = 0 \tag{14}$$

$$L^*_{\ldots}\phi^* = -i\ldots\phi^*, \quad G_v(v^{\prime\prime}, \ldots, \phi^*) = 0 \tag{15}$$

These eigenvectors are then normalized by requiring

$$\langle \phi, \phi \rangle = 1 \quad \text{and} \quad \langle \phi, \phi^* \rangle = 1. \tag{16}$$

Here we have introduced the complex [illegible] inner product

$$\langle \ldots, \ldots \rangle = \int_a^b \ldots \cdot \bar{\ldots} \, dx \tag{17}$$

where $\ldots \cdot \bar{\ldots}$ denotes the dot product of the vector functions [illegible] and [illegible].

A sequence of three linear nonhomogeneous boundary value problems
must next be solved. These are

$$L_{\mu_0}\frac{dv^{\prime\prime}}{d\mu} = -F_\mu(v^{\prime\prime}, \ldots)$$

$$G_v(v^{\prime\prime}, \ldots, \frac{dv^{\prime\prime}}{d\mu}) = -G_\mu(v^{\prime\prime}, \ldots) \tag{18}$$

$$L_{\mu_0}w_1 = -2F_{vv}(v^{\prime\prime}, \ldots, \ldots)$$

$$G_v(v^{\prime\prime}, \ldots, w_1) = -2G_{vv}(v^{\prime\prime}, \ldots, \ldots) \tag{19}$$

$$(L_{\mu_0} - 2i\omega_0 I)w_- = -F_{vv}(v^{\mu 0}, \omega_0 s_0, s_0)$$

$$\tag{20}$$

$$G_v(v^{\mu 0}, \omega_0 | w_-) = -G_{vv}(v^{\mu 0}, \omega_0 s_0, s_0).$$

The derivatives F_{vv} and G_{vv} appearing in the above problems are computed by the rule

$$F_{vv}(v^{\mu}, \omega|s, \omega) = \frac{\partial^2 F}{\partial\delta_1\partial\delta_2}(v^{\mu} + \delta_1 s + \delta_2 s, \omega)\Big|_{\delta_1=\delta_2=0}. \tag{21}$$

With these computations complete, we now compute $\sigma_\mu(\omega_0)$, ω_2, and ω_2 from

$$\sigma_\mu(\omega_0) = -F_{vv}(v^{\mu 0}, \omega_0|s_0, \frac{dv^{\mu 0}}{d\mu}) + F_{v_\mu}(v^{\mu 0}, \omega_0|s_0), s_0^* \tag{22}$$

and

$$(-i\omega_2 + \omega_2\sigma_\mu(\omega_0)) = -F_{vvv}(v^{\mu 0}, \omega_0|s_0, s_0, \bar{s}_0), s_0^*$$

$$-F_{vv}(v^{\mu 0}, \omega_0|s_0, w_+), s_0^* \tag{23}$$

$$-F_{vv}(v^{\mu 0}, \omega_0|\bar{s}_0, w_-), s_0^*$$

The first author evaluated the above expressions by using the Stephenson finite difference code to solve equations (16-20) and using Simpson's rule to numerically integrate the inner products. The second author solved the boundary value problem with the deBoor and Weiss' collocation code and integrated the inner products via Gauss quadrature quadratures. Both methods worked well and yielded quite comparable results for $\sigma_\mu(\omega_0)$, ω_2, and ω_2. A program for the computations for the general problem (6) will be available from the second author.

We note that for computations where rapid sensitivities were observed, the sensitivity may be eliminated by more accurate solution of the steady state solutions, the eigenvector, and the parameter value at which the onset occurs. This sensitivity is not removed by increasing the accuracy of these calculations to machine precision as in the case for ordinary differential equations obtained from a survey are an orthogonality relation for ... ...); and, perhaps more important, the condition for the eigenvalues of the linearized problem (12) and its adjoint.

VI. AN EXAMPLE FROM CHEMICAL REACTOR THEORY. To illustrate the computational procedure discussed above, we consider the onset of a tubular chemical reactor problem (1, 13), in which there is a single chemical reaction $A \to B$. The equations describing the concentration of reactant A and energy for the nonadiabatic tubular reactor with axial mixing appear below in dimensionless form:

$$\frac{\partial y}{\partial t} = \frac{1}{Pe_m}\frac{\partial^2 y}{\partial x^2} - \frac{\partial y}{\partial x} - Dy\, e^{-\gamma/T} \tag{24}$$

$$\frac{\partial \theta}{\partial \tau} = \frac{1}{Pe_h} \frac{\partial^2 y}{\partial s^2} - \frac{\partial \theta}{\partial s} - \beta(\theta - \theta_0) + B\, Dy\, e^{\gamma - \gamma/\theta}. \qquad (25)$$

The boundary and initial conditions are:

$$\frac{\partial y}{\partial s} = Pe_m(y - 1) \quad \text{at } s = 0,\ \tau > 0$$

$$\frac{\partial \theta}{\partial s} = Pe_h(\theta - 1) \quad \text{at } s = 0,\ \tau > 0 \qquad (26)$$

$$\frac{\partial y}{\partial s} = \frac{\partial \theta}{\partial s} = 0 \qquad \text{at } s = 1,\ \tau > 0 \qquad (27)$$

$$y = y_{in}, \quad \theta = \theta_{in} \text{ at } \tau = 0. \qquad (28)$$

In writing these equations, we have defined the following dimension-less quantities:

$$y = c/c_0 \qquad\qquad \theta = T/T_0$$

$$s = x/L \qquad\qquad \tau = tv/L$$

$$Pe_m = vL/D_e \qquad Pe_h = \rho C_p vL/k_e$$

$$B = \Delta Hc_0/\rho C_p T_\rho \qquad \beta = UPL/av_\rho C_p$$

$$D = Ae^{-\gamma}L/v \qquad \gamma = E/RT_0.$$

The above model describes an exothermic A → B reaction occurring in a homogeneous tube under the assumptions that the velocity profile is flat with constant velocity v; the variables y and θ depend only on one space dimension and time; the diffusion of reactant A is governed by Fick's Law with an effective diffusivity, D_e; heat conduction is described by Fourier's Law with an effective thermal conductivity, k_e; the heat loss at any point is proportional to (θ - θ); and the reaction rate is describable by an Arrhenius expression. The dimensional predecessors of the above equations and the applicability of this formulation are discussed in detail in two review articles [12, 13].

VII. RESULTS. We now illustrate the utility of our numerical procedures by applying the above methods to the tubular reactor model (24-28) for a couple of parameter combinations. A more extensive treatment will be given elsewhere [14]. We trace the steady state solution branches, determine their stability, and isolate the bifur-cation points. The Hopf bifurcation computations are then performed to determine the direction of bifurcation, the stability of the oscilla-tion close to the bifurcation point, and the asymptotic solutions for the orbits. These asymptotic solutions (9) are then used to start the time-dependent computations using PDECOL, a general code based on the method of lines and B-splines [15]. The stable periodic solutions are computed and then traced as the Damköhler number is varied.

If one thinks of operating the chemical reactor by varying the Damköhler number, the objective is to operate at a high temperature so
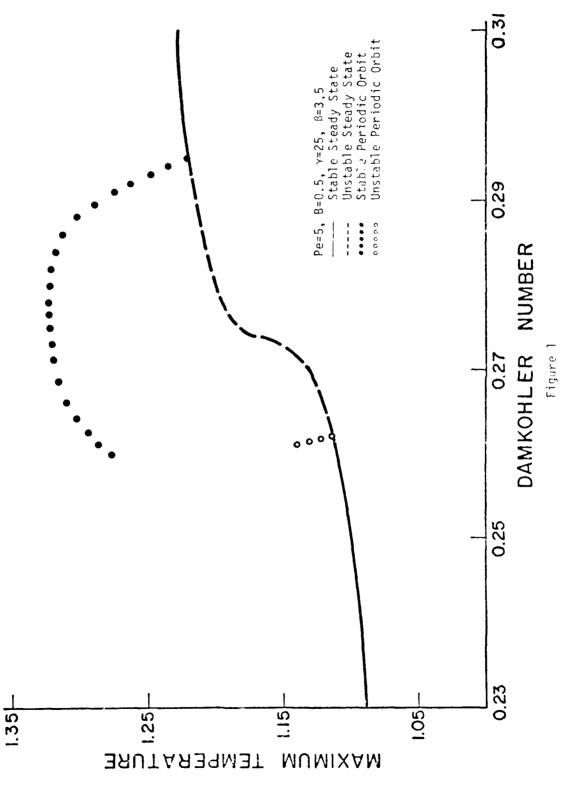
that a more complete conversion of the chemical A into B is achieved.
Thus one tries to operate the reactor at one of the higher steady states
in Figures 1 and 2. These higher steady states are achieved, however,
not by starting the reactor out at the high stable steady state, but by
varying a physical parameter such as the flow velocity, feed temperature,
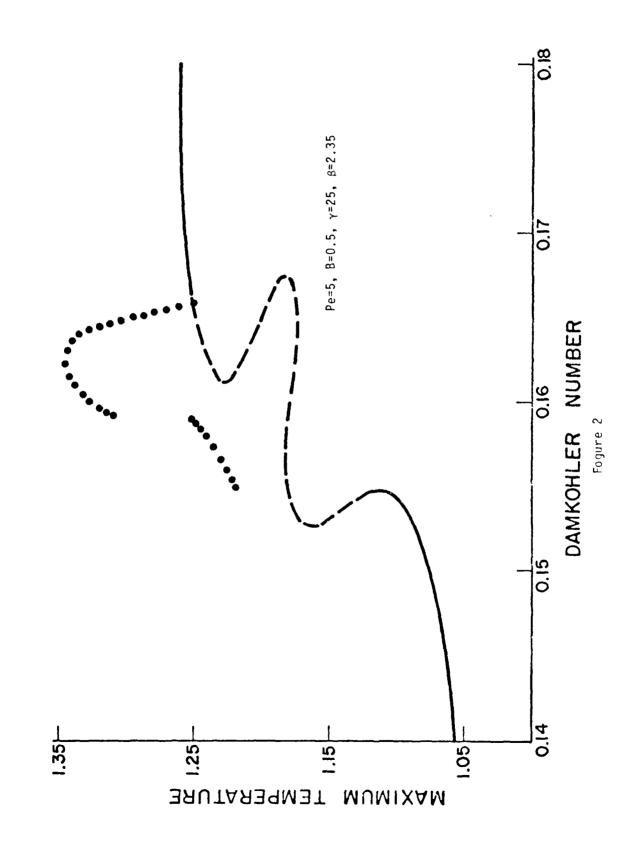or the Damköhler, which is used here.

The first example is illustrated in Figure 1 which corresponds to the
parameter values $Pe_h = Pe_m = 5$, $B = 0.5$, $\gamma = 25$, $\beta = 3.5$ and $\omega = 1$. There
is a unique steady state for all values of the Damköhler D with exchanges
in the stability at $D = 0.262$ and $D = 0.295$. At $D = 0.262$ the bifurcation
is to the left and unstable. At the value $D = 0.295$ the bifurcation is
also to the left but is not stable. By using AUTOL the stable oscillation
was traced from $D = 0.295$ down to about $D = 0.26$ where the stable oscilla-
tions cease to exist and the time-dependent solutions go to the stable
steady state directly below. (We conjecture that the stable branch of
periodic solutions connect with the unstable branch emanating from
$D = 0.262$.)

The response curve dynamics associated with varying the Damköhler
number D can now be explained for the case depicted in Figure 1. For D
close to zero, the reactor begins operating in a stable but low temperature
ture and thus low conversion of the chemical A into B. As D is increased
the steady state remains stable and the steady state temperature continues
to rise. As D passes through $D = 0.262$, a jump occurs in the temperature
and concentration profiles into the stable oscillation directly above,
which does not cease until D reaches $D = 0.295$ after which the reactor
operates in a stable steady state condition. To extinguish the reaction
the Damköhler is now decreased. At $D = 0.295$ a small stable oscillation
in the temperature and concentration profiles begins to appear. These
oscillations continue to grow but cease through a jump back down to the
stable steady state at $D = 0.26$. Both of these jumps may be thought of
as ignition and extinction processes, respectively.

The second example represented in Figure 2 corresponds to the
parameter values $Pe_h = Pe_m = 5$, $B = 0.5$, $\gamma = 25$, $\beta = 2.35$ and $\omega = 1$.
This case demonstrates a 1 - 3 - 1 - 3 - 1 multiplicity pattern in the
steady states. The first stability exchange is found at the lower limit
and all intermediate steady states remain unstable until a Hopf bifurca-
tion point is encountered at $D = 0.166$. A stable periodic orbit bifurca-
tes to the left at this value of the Damköhler.

If D starts out close to zero and is increased, the temperature
increases and remains in a stable steady state until the lower limit
point is encountered at which point the temperature jumps into the much
higher stable oscillation. This stable oscillation grows steadily until
a jump to an even higher stable oscillation at about $D = 0.159$. The
amplitude of the temperature oscillation continues to grow, peaks out,
and then rapidly decays back to the steady state at $D = 0.166$. To extinguish
the reaction the Damköhler number can now be decreased. The rapid growth of
the amplitude of a periodic orbit begins at $D = 0.166$ which peaks out,
decays, jumps down, continues to decay, and then disappears as the time-
dependent solutions decay to the stable steady state at the
lower turning point.

320

VIII. CONCLUSIONS. The bifurcation techniques presented enable one to determine the possible steady states and bifurcating stable and unstable oscillations. Thus systematic numerical methods are established for investigating the response curve dynamics, including jump phenomena, and the oscillatory dynamics for a broad array of models found in chemical reactor theory, combustion theory, and even mathematical biology.

Pe=5, B=0.5, γ=25, β=3.5
——— Stable Steady State
— — — Unstable Steady State
●●●●● Stable Periodic Orbit
○○○○○ Unstable Periodic Orbit

DAMKOHLER NUMBER

MAXIMUM TEMPERATURE

Figure 1

MAXIMUM TEMPERATURE

DAMKOHLER NUMBER

Pe=5, B=0.5, γ=25, β=2.35

Figure 2

NOTATION

a	cross-sectional area of the reactor, m^2
A	frequency factor, s^{-1}
B	dimensionless heat of reaction, $\Delta H c_0 / \rho C_p T_0$
c	concentration, mol/m^3
C_0	inlet concentration, mol/m^3
C_p	specific heat, $J/mol°K$
D	Damkohler number, $Ae^{-\gamma}L/v$
D_e	effective diffusivity, m^2/s
E	activation energy, J/mol
ΔH	heat of reaction, J/mol
k_e	effective thermal conductivity J/s $m°K$
L	reactor length, m
P	reactor perimeter, m
Pe_m	Peclet number for mass transfer vL/D_e
Pe_h	Peclet number for heat transfer $\rho C_p vL/k_e$
R	universal gas constant
t	time, s
T	temperature, $°K$
T_0	inlet temperature, $°K$
s	dimensionless axial distance, x/L
U	heat transfer coefficient, J/m^2 $s°K$
v	velocity, m/s
x	axial distance, m
y	dimensionless concentration, c/c_0

Greek symbols

β	dimensionless heat transfer coefficient, $UPL/av\rho C_p$
γ	dimensionless activation energy, E/RT_0
θ	dimensionless temperature, T/T_0
ρ	density, kg/m^3
τ	dimensionless time, tv/L

REFERENCES

1. Keller, H.B., In *Applications of Bifurcation Theory* (Edited by P.H.Rabinowitz), p. 359. Academic Press, New York, 1977.

2. Poore, A.B., *Arch. Rational Mech. Anal.*, 1976, 60, 371.

3. Crandall, M.G. and Rabinowitz, P.H., MRC Tech. Summary Report #1604, University of Wisconsin, Madison, Wisconsin, 1976.

4. Iooss, G. and Joseph, D.D., *Elementary Stability and Bifurcation Theory*, University of Minnesota Lecture Notes, 1979.

5. de Boor, C. and Weiss, R., MRC Tech. Summary Report #1625, University of Wisconsin, Madison, Wisconsin, 1976.

6. Stepleman, R.S., *Math. Comp.*, 1976, 30, 92.

7. Heinemann, R.F., Overholser, K.A. and Reddien, G.W., *Chem. Engng. Sci.* 1979, 35, 833.

8. Heinemann, R.F., Overholser, K.A. and Reddien, G.W., *A. I. Ch. E. J.* in press, 1980.

9. Peterson, J. Overholser, K.A. and Heinemann, R.F., *Chem. Engng. Sci.* in press, 1980.

10. Overholser, K.A. and Heinemann, R.F., Submitted for publication, 1980.

11. Bissett, E. and Cavendish, J.C., 72nd Annual A. I. Ch. E. Meeting, San Francisco, 1979.

12. Schmitz, R.A., *Adv. Chem. Ser.*, 1975, 148, 156.

13. Varma, A. and Aris, R., In *Chemical Reactor Theory* (Edited by L. Lapidus and N.R.Amundson), p. 79. Prentice-Hall, Englewood Cliffs, NJ, 1977.

14. Heinemann, R.F. and Poore, A.B., submitted for publication, 1980.

15. Sincovec, R.F. and Madsen, N.K., *ACM-TOMS*, 1975, 1, 232.

THERMOELASTIC WAVE PROPAGATION

J. L. Davis

US Army Armament Research and Development Command
Dover, N. J. 07801

Yu Chen
Dept. of Mechanics and Materials Science
Rutgers, The State University of New Jersey
P.O. Box 909
Piscataway, N. J. 08854

ABSTRACT. The coupled dynamic thermoelastic problem was formulated
as a fourth-order partial differential equation in temperature (or stress).
The fourth order Laplace-transformed operator was decomposed into a "wave
like" and a "diffusion like" operator. Boggio's theorem was extended vis-a-
vis this decomposition. As a result of this extension two functions were
defined, one satisfies a "wave like" equation, the other a "diffusion like"
equation. The boundary value problem for the fourth order PDE in temperature
was formulated in a finite medium and a method of solution was obtained
through Boggio's theorem and a perturbation techniuqe.

I. INTRODUCTION. The dynamic thermoelasticity problem has been
studied quite extensively since its beginning in 1838 when Duhamel derived
equations for the strain field in an elastic medium containing temperature
gradient. A comprehensive review of the literature up to 1960 was given
by Chadwick [1]. A more recent treatise on the subject was the 1975
edition of the book "Dynamic Problem of Thermoelasticity" by Nowacki [2].
Other articles which have close relevance to the subject matter of thermo-
elastic wave propagation are listed in references [3-18]. Since the
literature on the subject is so vast, the reference list contains only those
which the authors have some faimilarity with.

To understand the nature of thermoelastic wave, it is instructive to
look into the characteristics of the uncoupled waves. For a hypothetical
medium with the thermal expansion coefficient $\alpha = 0$ the pair of equations of
thermoelasticity are uncoupled to give the one-dimensional wave and heat
equation respectively. On inserting the plane wave solution of the form

$$\{\sigma,T\} = \{\sigma^0,T^0\} \exp \{i(kx-\omega t)\} \tag{1}$$

into each of the following uncoupled equations

$$c^2_{xx} = u_{tt} \qquad (2)$$

and

$$T_{xx} = c^{-1}T_t \qquad (3)$$

one obtains two relations between ω and k,

$$\omega^2 = c^2 k^2 \qquad (4)$$

and

$$i\omega = \kappa k^2. \qquad (5)$$

For waves of assigned frequency by letting ω be a real constant, the solutions obtained are:

$$u = u_+^0 \exp\{-i\omega(t-\tfrac{x}{c})\} + u_-^0 \exp\{-i\omega(t+\tfrac{x}{c})\},$$

$$T = T_+^0 \exp\{-x\sqrt{\tfrac{\omega}{2\kappa}} - i\omega(t-\tfrac{x}{\sqrt{2\kappa\omega}})\} \qquad (6)$$

$$+ T_-^0 \exp\{x\sqrt{\tfrac{\omega}{2\kappa}} - i\omega(t + \tfrac{x}{\sqrt{2\kappa\omega}})\},$$

which represent progressive waves travelling along the x-axis. The thermal wave is subject to dispersion as the phase velocity $\sqrt{2\kappa\omega}$ is a function of the frequency.

On the other hand, if one assigns k to be a real constant, waves of assigned length are represented by

$$\sigma = \sigma_+^0 \exp\{ik(x-ct)\} + u_-^0 \exp\{ik(x+ct)\},$$

$$T = T^0 \exp\{-\kappa k^2 t + ikx\}. \qquad (7)$$

The elastic waves have the same character as before whereas the temperature is a standing wave, the amplitude decaying exponentially with time.

For the propagation of the coupled plane harmonic waves one assumes the same wave form as represented by equation (1) and substitutes the solutions in the pair of coupled equations such as equations (18) and (19) of reference 17. One can either study waves of assigned frequency or waves of assigned

length. Chadwick [1] discussed these solutions in great length in his article. In essence, the purely elastic and thermal waves which are the solutions of the completely uncoupled equations are modified. Chadwick calls these modified waves quasi-elastic and quasi-thermal waves. For waves of assigned frequency, a thermoelastic wave of displacement or temperature consist of a quasi-elastic and a quasi-thermal mode. This phenomenon represents a coupling of elastic and thermal effects, the strength of which depends on the frequency and the coupling constant ϵ. The quasi-elastic mode, in contradiction to purely elastic wave, is subject to damping and dispersion. On the other hand, both the purely thermal and its modification, the quasi-thermal mode, exhibit damping and dispersion.

Waves of assigned length, as would be expected, display properties of modification and coupling just as waves of assigned frequency. The quasi-thermal mode is also a standing wave, like its counterpart in the uncoupled wave. A comprehensive discussion of the modified waves as outlined here can be found in Chadwick's article [1].

The subject of this paper is on waves in a bounded medium. Some discussion of the boundary value problem in one dimension can be found in [16] and [17]. A different approach will be adopted in this paper. This approach is based on the idea contained in Ignaczak's paper [15] in which he showed how the solutions to the coupled problem can be approached via Boggio's theorem [19]. Boggio's theorem indicates that the displacement solution, as well as the temperature solution to the coupled problem can be constructed by superimposing two solutions, each satisfying a "wave like" and a "diffusion like" equation respectively. This fact that the solutions to the coupled partial differential equations can be decomposed mathematically into two solutions as described above will be demonstrated in the following sections through the Laplace transform method. In the next section the fundamental equations will be displayed.

II. MATHEMATICAL MODEL. Since the following discussion will be restricted to one-dimensional wave propagation, all the equations will be given in one-dimensional form. Let the bounded region R be given by $0 \leq x \leq \ell$. The equation of motion is

$$\sigma_x = \rho u_{tt} \tag{8}$$

where ρ is the density, σ is the stress and u is the displacement of a field point at x. Equation (8) can be expressed in the strain e as

$$\sigma_{xx} = \rho e_{tt} \tag{9}$$

The coupled energy equation is

$$T_{xx} - \kappa^{-1} T_t = \eta e_t \tag{10}$$

where κ is the diffusivity ($\kappa = \rho c / K$), $\eta = \gamma T_0 / \rho c k$ and $\gamma = (3\lambda + 2\mu)/\alpha$, α being the coefficient of thermal expansion.

The Duhamel-Newmann constitutive equation is

$$\sigma = Ee - \gamma T \tag{11}$$

where E is the Young's modulus of the materials.

The boundary conditions are as follows:

$$x = 0, \quad \sigma(0,t) = f(t)$$
$$T_x(0,t) - \beta_1 T(0,t) = q(t) \tag{12}$$

$$x = \ell, \quad \sigma(\ell,t) = c$$
$$T_x(\ell,t) + \beta_2 T(\ell,t) = 0 \tag{13}$$

where $\beta_1 = \dfrac{h_1}{\kappa}$, $\beta_2 = \dfrac{h_2}{\kappa}$, $f(t) = -p(t)$, $q(t) = \beta_1 T_g(t)$, h_1 and h_2 are heat transfer coefficients at the inner and outer surfaces respectively.

DIMENSIONLESS VARIABLES. We define dimensionless variables x, t, T, σ, by the barred quantities as follows:

$$\bar{x} = \frac{x}{\ell}, \quad \bar{t} = \frac{c_1^2 t}{\kappa}, \quad \bar{T} = \frac{T}{T_0}, \quad \bar{\sigma} = \frac{\sigma}{E}, \quad \bar{c} = \alpha T_0 \tag{14}$$

Two dimensionless parameters ϵ and r are defined

$$\epsilon = r\bar{\alpha}, \quad r = \frac{\kappa}{\ell c_1}, \quad c_1^2 = \frac{E}{\rho} \tag{15}$$

Inserting equations (13) and (14) in equations (9) and (10) with the additional equation (11) to reduce the PDE's to dimensionless form in the variables T and σ, we have

$$r^2 T_{xx} - (1+a_1\epsilon^2)T_t = a_2\epsilon\sigma_t \qquad (16)$$

$$r^2\sigma_{xx} - \sigma_{tt} = a_3\epsilon T_{tt} \qquad (17)$$

where $a_1 = (3\lambda+2\mu)^2/T_0\rho c(\lambda+2\mu)r^2$

$\quad a_2 = (3\lambda+2\mu)/\rho c T_0 r$

$\quad a_3 = (3\lambda+2\mu)/(\lambda+2\mu)r.$

The boundary conditions in dimensionless form are:

$$x = 0, \quad \sigma(0,t) = \bar{f}(t)$$

$$T_x - \bar{\beta}_1 T = \bar{g}(t) \qquad (18)$$

where $\bar{f}(t) = f(t)/E$, $\bar{\beta}_1 = \beta_1\ell$, $\bar{g} = g(t)\ell/T_0$.

$$x = 1, \quad \sigma(1,t) = 0,$$

$$T_x + \bar{\beta}_2 T = 0 \qquad (19)$$

where $\quad \bar{\beta}_2 = \beta_2\ell.$

<u>FOURTH-ORDER PDE's</u>. The pair of second-order PDE's can be reduced to a single fourth-order PDE either in T or in σ. By simple algebraic manipulations we obtain

$$(r^2 T_{xx}-T_{tt})_{xx} - (T_{xx}-T_{tt})_t = a_1\epsilon^2 T_{xxt} \qquad (20)$$

and

$$(r^2\sigma_{xx}-\sigma_{tt})_{xx} - (\sigma_{xx}-\sigma_{tt})_t = a_1\epsilon^2 \sigma_{xxt}. \qquad (21)$$

It is noted that equations (20) and (21) can be put in a more compact form by introducing two operators as follows:

$$\delta \equiv \partial^2/\partial x^2, \quad D^2 = \partial^2/\partial t^2. \qquad (22)$$

Thus we can write

$$[\delta^2(r^2\delta^2-D^2) - D(r^2\delta^2-D^2)]T = a_1\epsilon^2 r^2\delta^2 DT$$

or

331

$$[(\sigma^2 - \beta)(r^2 \sigma^2 - \sigma^2) - a_1 \ldots] \ldots = 0$$

III. DECOMPOSITION THEOREM. Ignaczak [15] extended the theorem to to Boggio [19] to the thermoelasticity problem. The decomposition was accomplished for the PDE equation (23). In this paper a decomposition theorem will be proven for the Laplace transformed PDE with respect to t. The proof of the theorem will be given in two steps, first for the uncoupled and then for the coupled equations.

UNCOUPLED EQUATIONS. Letting $\ldots$ 0 in equation (23) we obtain

$$(\sigma^2 - \beta)(r^2 \sigma^2 - \sigma^2) T = 0 \tag{24}$$

Assuming homogeneous initial conditions on T, T_t, T_{tt} and T_{ttt}, we will show that there exist two functions $u(x,t)$ and $v(x,t)$ such that

$$(r^2 \sigma^2 - \sigma^2) u = 0, \quad (r^2 \sigma^2 - \beta) v = 0 \text{ and } T = u + v. \tag{25}$$

Laplace transforming equation (25) yields

$$(r^2 \sigma^2 - s^2) u = 0, \quad (r^2 \sigma^2 - s) v = 0, \quad T = u + v \tag{26}$$

where s is the transform variable, u, v and T are transformed functions for shortness equation (26) is rewritten in symbolic forms, the symbols being used to replace the operators on u and v. Thus,

$$\square u = 0, \quad H v = 0 \tag{27}$$

and equation (24) after being Laplace transformed becomes

$$H \square T = 0. \tag{28}$$

The decomposition theorem states that if $\square u = 0$, $K(S)u = HT$ where $K(S) = s^2 - s$, then $Hv = 0$ and $T = u + v$.

PROOF. Define $v(x,s)$ such that $v = T-u$, since $H = \square + s^2 - s$, we have

$$Hv = HT - Hu = HT - \square u - (s^2 - s)u$$

but by hypothesis $\square u = 0$, thus

333

$$Hv = HT - (s^2-s)u = k(s)u-(s^2-s)u \tag{30}$$

hence $Hv = 0$ if and only if $K(s) = s^2-s$ which is given in the hypothesis.

It is noted that this theorem can be stated for the system of PDE's in equation (24) and (25) by simply applying inversion to the above. Also, a parallel theorem can be stated by interchanging the roles played by u and v. That is, if $Hv = 0$, $(D-D^2)v = \square T$, then $\square u = 0$ and $T = u + v$.

COUPLED EQUATIONS. Taking the Laplace transform of equation (24) and writing the resulting equation in symbolic form

$$LT = 0 \tag{31}$$

where

$$L = (r^2 \ell^2-s^2)(r^2 k^2-s) - a_1 r^2 \ell^2 s^2$$

We shall attempt to decompose L as follows. Let

$$L = (r^2 \ell^2-p)(r^2 \ell^2-q) \tag{32}$$

where p and q are functions of s to be determined. Expanding equation (32) and equating coefficients of like powers of ℓ^2 with that from equation (31), we get

$$p + q = s^2 + s + a_1 \ell^2 s \quad \text{and} \quad pq = s^3. \tag{33}$$

Therefore p and q are the roots of the following quadratics

$$z^2 - (s^2+s+a_1 \ell^2 s)z + s^3 = 0. \tag{34}$$

Now we can write equation (31) in the form

$$L_1 L_2 T = 0 \tag{35}$$

where $L_1 = r^2 \ell^2-p$, $L_2 = r^2 \ell^2-q$.

The decomposition theorem states that given equation (31), $L_1 u = 0$ and $k(s)u = L_2 T$ where $k(s) = p-q$, then $L_2 v = 0$, $T = u + v$.

PROOF. Define $v = T-u$. Since $L_2 = L_1 + p-q$ we have
$$L_2 v = L_2 T - L_2 u = L_2 T - L_1 u - (p-q)u = L_2 T - (p-q)u = 0.$$

It is noted that the theorem can be stated in a similar way by inter-changing the roles played by u and v.

IV.   BOUNDARY VALUE PROBLEM AND ITS SOLUTION BY A PERTURBATION TECHNIQUE.
The boundary value problem of the fourth order PDE in equation (33) can be stated, for example, in $T(x,t)$ as

$$LT = 0, \quad 0 \leq x \leq 1, \quad t \geq 0 \tag{36}$$

where $L = (r^2 \partial^2 - \partial^2)(r^2 \partial^2 - \partial^2) - a_1^2 r^2 \partial^2 \partial$ and the associated boundary conditions

$$x = 0, \quad [\partial - \partial_1]T(0,t) = \bar{a}(t) \tag{37}$$

$$[\partial^2 - (1 + \partial^2)\partial]T(0,t) = \partial \bar{f}(t)$$

$$x = 1, \quad [\partial + \partial_2]T(1,t) = 0 \tag{38}$$

$$[\partial^2 - (1 + \partial^2)\partial]T(1,t) = 0$$

The second conditions in equation (37) and (38) are stress boundary conditions. They are expressed, through the use of the Duhamel-Newmann equation (11) and the energy equation (10) and by simple calculation, in terms of the temperature.

To solve the boundary value problem we shall first use Laplace transform in all the equation (36), (37) and (38). The system of the transformed equation are as follows:

$$LT = 0, \tag{39}$$

where $L = (r^2 \partial^2 - s^2)(r^2 \partial^2 - s) - a_1^2 r^2 \partial s,$

$$x = 0, \quad [\partial - \partial_1]T(0,s) = \bar{a}(s) \tag{40}$$

$$[\partial^2 - (1 + \partial^2)s]T(0,s) = s \bar{f}(s)$$

$$x = 1, \quad [\partial + \partial_2]T(1,s) = 0 \tag{41}$$

$$[\partial^2 - (1 + \partial^2)s]T(1,s) = 0.$$

According to the decomposition theorem proven before, in terms of the

functions u and v, we have

$$[r^2 \varepsilon^2 - p(s)]\tilde{u}(x,s) = 0$$
$$[r^2 \varepsilon^2 - q(s)]\tilde{v}(x,s) = 0.$$

(42)

After u and v are solved as solutions of the ODE, the solution to the fourth order ODE, equation (39), can be obtained as the sum of these two functions. However, the boundary conditions in equation (40) and (41) are applied to a linear combination of these functions. A perturbation technique is proposed.

PERTURBATION TECHNIQUE. Let

$$T(x,s,\varepsilon) = \sum_{j=0}^{\infty} \tilde{T}_j(x,s)\varepsilon^j$$

$$u(x,s,\varepsilon) = \sum_{j=0}^{\infty} \tilde{u}_j(x,s)\varepsilon^j$$

(43)

$$v(x,s,\cdot) = \sum_{j=0}^{\infty} v_j(x,s)\varepsilon^j$$

where T_j, $\tilde{u}_j$ and v_j are the expansion coefficients of the respective functions.

The boundary value problems for the expansion coefficients are obtained by inserting equation (43) and in equations (39), (40) and (41). Thus we have the ODE

$$(r^2\varepsilon^2-q)(r^2\varepsilon^2-s)\tilde{T}_n(x,s) \quad d_1\varepsilon^2 s\tilde{T}_{n-2}(x,s)$$

(44)

and the boundary conditions

$$x \quad \vdots, \quad \varepsilon = {}_1 \tilde{T}_0(0,s) \quad {}_{0n}\tilde{f}(s), \quad {}_{0n} = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

(45)

$$[r^2-s]T_0(0,s) \quad {}_{1n}s\tilde{f}(s) + sT_{n-2}(0,s) \quad {}_{1n} = \begin{cases} 1, & n = 1 \\ 0, & \text{others} \end{cases}$$

$$x \quad 1, \quad \vdots \quad \tilde{T}_n(1,s) \quad 0$$

(46)

$$[r^2-q]\tilde{T}_n(1,s) \quad sT_{n-2}(1,s)$$

We also need to expand $p(s)$ and $q(s)$ into infinite series of even powers in ϵ. Thus,

$$p = \sum_{j=0}^{\infty} p_{2j}(s)\epsilon^{2j}, \qquad q = \sum_{j=0}^{\infty} q_{2j}(s)\epsilon^{2j} \tag{47}$$

Inserting equations (43) and (47) into equation (42) yields

$$(r^2\delta^2 - p_0)\tilde{u}_{2\ell}(x,s) = \sum_{k=0}^{2\ell} p_{2\ell-k}\tilde{u}_k(x,s)$$

$$\tag{48}$$

$$(r^2\delta^2 - q_0)\tilde{v}_{2\ell}(x,s) = \sum_{k=0}^{2\ell} q_{2\ell-k}\tilde{v}_k(x,s) \ .$$

The ODE's in equation (48) represent a set of recursion relationships amongst the expansion coefficients $\tilde{u}_j$ and $\tilde{v}_j$ respectively. Even though $p(s)$ and $q(s)$ have expansions in even powers of ϵ, $\tilde{u}(x,s,\epsilon)$ and $\tilde{v}(x,s,\epsilon)$ must be expanded in both odd and even powers of ϵ because the boundary conditions depend on ϵ as well as ϵ^2.

The boundary conditions for $\tilde{u}_j$ and $\tilde{v}_j$ are coupled and are given by the following system

$$x = 0, \quad \delta u_n - \bar{\beta}_1\tilde{u}_n + \delta v_n - \bar{\beta}_1\tilde{v}_n = \lambda_{0n}\bar{q}(s)$$

$$\sum_{j=0}^{n} p_j\tilde{u}_{r-j} + \sum_{j=0}^{n} q_j\tilde{v}_{n-j} - s\tilde{u}_n - s\tilde{v}_n - s\tilde{u}_{n-2} \tag{49}$$

$$- s\tilde{v}_{n-2} = \lambda_{1n}s\bar{f}(s) \ .$$

The λ usage has been explained in equation (45).

$$x = 1, \quad \delta u_n + \bar{\beta}_2\tilde{u}_n + \delta v_n + \bar{\beta}_2\tilde{v}_n = 0$$

$$\sum_{j=0}^{n} p_j\tilde{u}_{n-j} + \sum_{j=0}^{n} q_j\tilde{v}_{n-j} - s\tilde{u}_n - s\tilde{v}_n \tag{50}$$

$$- s\tilde{u}_{n-2} - s\tilde{v}_{n-2} = 0.$$

Uncoupled case, $\epsilon = 0$. In this case the system of equations becomes $n = 0$ by letting

$$(r^2 s^2 - p_0)\tilde{u}_0(x,s) = 0, \quad p_0 = s^2 \tag{51}$$

$$\tilde{u}_0(0,s) = 0, \quad \tilde{u}_0(1,s) = 0$$

and

$$(r^2 s^2 - q_0)\tilde{v}_0(x,s) = 0 \tag{52}$$

$$[\delta - \bar{\varepsilon}_1]\tilde{v}_0(0,s) = \bar{g}(s), \quad [\delta + \varepsilon_1]\tilde{v}_0(1,s) = 0 .$$

This shows that $u_0(x,t)$ is identically zero, signifying the fact that in the fundamental temperature response there is no contribution from the stress input. On the other hand, $\tilde{v}_0(x,t)$ is the uncoupled temperature distribution in the bounded region $0 \leq x \leq 1$ with the given thermal boundary conditions.

First order in ε. The first order expansion coefficient $\tilde{u}_1(x,s)$ satisfies the following transformed boundary value problem

$$(r^2 s^2 - s^2)\tilde{u}_1(x,s) = 0 \tag{53}$$

$$x = 0, \quad (s-1)\tilde{u}_1(0,s) = \bar{f}(s), \quad x = 1, \quad (s-1)\tilde{u}_1(x,s) = 0.$$

Here u_1 is by choice made to represent the solution of the pure stress problem with the impact stress at $x = 0$. Observing from the first boundary condition in equation (51) it can be said that $u_1(0,t) = f(t)e^t$ which is unstable. This suggests that this perturbation method holds only for small time.

The first order expansion coefficient $\tilde{v}_1(x,s)$ is the solution to the following boundary value problem:

$$(r^2 s^2 - s)\tilde{v}_1(x,s) = 0$$

$$x = 0, \quad [\delta - \bar{\varepsilon}_1]\tilde{v}_1(0,s) = -[\delta - \bar{\varepsilon}_1]\tilde{u}_1(0,s) \tag{54}$$

$$x = 1, \quad [\delta + \bar{\varepsilon}_2]\tilde{v}_1(1,0) = -[\delta + \varepsilon_2]\tilde{u}_1(1,s).$$

Equation (54) shows that the boundary conditions are coupled. $\tilde{v}_1(x.t)$ is the solution of the unsteady Fourier heat equation with boundary conditions that depend on $u_1(x,t)$. so that $v_1(x,t)$ is not purely thermal but depends

on the stress through the boundary conditions.

Second order in . The second order expansion coefficient $u_2(x,s)$ is the solution to the following transformed boundary value problem

$$(r^2 s^2 - s^2) u_2(x,s) = 0$$

$$x = 0, \quad u_2(0,s) \quad \frac{s}{(s-1)^2} v_0(x,s)$$

$$x = 1, \quad u_2(1,s) = \frac{s}{(s-1)^2} v_0(1,s).$$

(55)

Thus, $u_2(x,t)$ is a contribution to the temperature distribution in the form of a mechanical wave due to the zero-order temperature field.

The second order expansion coefficient $v_2(x,s)$ is the solution to the following transformed boundary value problem

$$(s^2 - s) v_2(x,s) = q_2 v_0(x,s), \quad q_2 = -\frac{s}{s-1}$$

$$x = 0, \quad [\delta - \beta_1] v_2(0,s) = -[s - \beta_1] u_2(0,s)$$

$$x = 1, \quad [s + \beta_1] v_2(1,s) = -[s + \beta_2] u_2(1,s)$$

(56)

It is noted that $v_2(x,t)$ is a response to the nonhomogeneous unsteady heat equation with a source term due to $v_0(x,t)$.

REFERENCES

1. Chadwick, P., Thermoelasticity. The Dynamic Theory, Progress in Solid
   Mechanics, Vol. 1, North-Holland Publishing Co., Amsterdam,
   pp. 265-328 (1960).

2. Nowacki, W., Dynamic Problems of Thermoelasticity, Noordhoff
   International Publishing, Layden, The Netherlands (1975).

3. Deresiewicz, H., Plane Waves in a Thermoelastic Solid, J. Acoust. Soc.
   Am., Vol. 29, No. 2, pp. 204-209 (1957).

4. Sternberg, E., Chakravorty, J.G., On Inertia Effects in a Transient
   Thermoelastic Problem, J. Appl. Mech., Vol. 26, pp. 503-509 (1959).

5. Nariboli, G.A., Nyayadhish, V.B., One-Dimensional Thermoelastic Wave,
   Quart. J. Mech. and Appl. Math., Vol. 16, Pt. 4, pp. 473-482 (1963).

6. Achenbach, J.D., Approximate Transient Solutions for the Coupled
   Equations of Thermoelasticity, J. Acoust. Soc. Am., Vol. 36, No.
   1, pp. 10-18 (1964).

7. Dillon, D.W., Jr., Thermoelasticity when the Material Coupling
   Parameter Equals Unity, J. Appl. Mech., Vol. 32, pp. 378-382
   (1965).

8. Soler, A.I., Brull, M.A., On the Solution of Transient Coupled Thermo-
   elastic Problems by Perturbation Techniques, J. Appl. Mech.,
   Vol. 32, pp. 389-399 (1965).

9. Zaker, T.A., Stress Waves Generated by Heat Addition in an Elastic
   Solid, Vol. 32, pp. 143-150 (1965).

10. Dun, H.S., Transient Solutions of a One-Dimensional Thermoelastic
    Wave Propagation Problem, Quart. J. Mech. and Appl. Math., Vol.
    XIX, Pt. 2, pp. 157-165 (1966).

11. Oden, J.T., Finite Element Analysis of Nonlinear Problems in the
    Dynamic Theory of Coupled Thermoelasticity, Nuclear Engineering
    and Design, 10, pp. 465-475 (1969).

12. Nickell, R.E., Sackman, J.L., Approximate Solutions in Linear Coupled
    Thermoelasticity, J. Appl. Mech., Vol. 36, pp. 246-266 (1968).

13. Carlson, D.E., Linear Thermoelasticity, Encyclopedia of Physics,
    Mech. of Solids, II a/2, Springer-Verlag, Berlin, p. 328 (1972).

14. Brun, L., Londe Simple Thermoelastique Lineaire, J. de Mecanique,
    pp. 863-885 (1975).

339

15. Ignaczak, J., Thermoelastic Counterpart to Boggio's Theorem of Linear Elastodynamics, Bull. de L'Academie Poionaise des Sciences, Serie des Sciences Techniques, Vol. 24, No. 2, pp. 129-137 (1976).

16. David, J.L., Chen, Y., Analysis of the Thermoelastic Problem in a Slab, ASME Publication 79-HT-59 (1979).

17. Davis, J.L., Chen, Y., Thermoelastic Stresses in Gun Barrels, ARO Report 80-1, pp. 161-176 (1979).

18. Molinari, A., Réponse dynamique dún demi-espace thermoélastique à un depôt d'e'nergie, J. de Mecanique, pp. 433-468 (1979).

19. Truesdell, C., Encyclopedia of Physics, Vol. VI a/2, Mechanics of Solids II, p. 237 (1972).

SOLUTION TO THE RIEMANN PROBLEM FOR THE EQUATIONS OF GAS DYNAMICS

IN A TUBE WITH VARYING CROSS SECTION

Reza Malek-Madani and Shao-Shiung Lin

Mathematics Research Center, University of Wisconsin

ABSTRACT

The equations of gas dynamics in a tube with varying cross section are an example of a nonhomogeneous system of conservation laws. In this work we study the Riemann problem for this system by viewing it as a perturbation of the classical equations of gas dynamics in a uniform tube. Also, we study the Riemann problem and the formation of singularities for a related, but simpler, problem of nonhomogeneous Berger's equation.

1. Introduction. The equations of gas dynamics in a uniform tube have been studied quite extensively in recent years. It is well known that, as a hyperbolic conservation law, these equations exhibit discontinuous solutions, while the initial value problem is not mathematically well posed in the class of weak solutions [1]. It is not difficult to envisage the mathematical reason for the nonsmoothness of solutions. These equations enjoy a full set of real characteristics and, if the initial values are chosen properly, the information carried by the characteristics will overlap and shocks develop. The problem under study in this paper has one additional property, namely the variation in the tube's cross section, that will persumably contribute even further to the shock producing mechanisms.

Section 2 concerns with the derivation of the equations studied in this work. The arguments of Hughes [2] have been followed and, as it will become apparent, the system under consideration is an example of nonhomogeneous hyperbolic conservation laws. In Section 3 a simpler but related problem is discussed for the purpose of understanding the shock producing mechanisms that do not exist in the homogeneous problem.

Section 4 concerns the solution of the Riemann problem. It is well known [3], [4] that the solution of the Riemann problem played an essential role in developing a numerical scheme in order to solve the initial value problem for the equations of gas dynamics in a tube with uniform cross section. Motivated by this fact T. P. Lui [5] applied a modified Riemann problem for the general n^{th} order nonhomogeneous conservation laws and developed an iterative scheme which converges to the weak solution of the initial value problem. Although the above scheme is quite successful theoretically it is rather difficult to implement it. Since we have in mind a concrete example from the equations of gas dynamics it is our contention to propose a simpler Riemann problem and hope that it would give rise to more manageable computations. We are presently studying this problem.

2. <u>Derivation of the model equation.</u> Consider an inviscid isentropic gas flow

through a two dimensional duct $D = \{(x,y) | A_1(x) \leq y \leq A_2(x), -\infty < x < \infty \}$. The

motion of the gas is governed by the equations of conservation of mass and linear

momentum

$$\rho_t + (\rho u)_x + (\rho v)_y = 0 ,$$

$$(\rho u)_t + (\rho u^2 + P)_x + (\rho u v)_y = 0 , \qquad (2.1)$$

$$(\rho v)_t + (\rho u v)_x + (\rho v^2 + P)_y = 0$$

with $P = f(\rho)$, where $\rho = \rho(x,y,t)$ is the density, $P = P(x,y,t)$ is the pressure

and $\underset{\sim}{u} = (u,v)$ is the velocity vector, together with the Neumann boundary

conditions

$$u(x,A_i(x),t)A_i'(x) = v(x,A_i(x),t), \qquad i = 1,2$$

and the initial conditions

$$\rho(x,y,0) = \rho_0(x,y) ,$$

$$u(x,y,0) = u_0(x,y) ,$$

$$v(x,y,0) = v_0(x,y) .$$

In the remainder of this section we will outline briefly the procedure

discussed in [3] which approximates (2.1) by a one-dimensional nonhomogeneous system

in the variables ρ and u. For a physical quantity $q(x,y,t)$ defined in the

region D we define the average $\langle q \rangle$ of q in the y-direction

$$\langle q \rangle = \frac{1}{A(x)} \int_{A_1(x)}^{A_2(x)} q(x,y,t)dy$$

where $A(x) = A_2(x) - A_1(x)$. Averaging each equation in (2.1) and using the

boundary conditions yield

$$\langle \rho \rangle_t + \langle \rho u \rangle_x = -\frac{A'(x)}{A(x)} \langle \rho u \rangle ,$$

$$\langle \rho u \rangle_t + \langle \rho u^2 \rangle_x + \langle P \rangle_x = -\frac{A'(x)}{A(x)} \langle \rho u^2 \rangle , \qquad (2.2)$$

$$\langle \rho v \rangle_t + \langle \rho u v \rangle_x + \langle P_y \rangle = -\frac{A'(x)}{A(x)} \langle \rho u v \rangle ,$$

$$\langle P \rangle = \langle f(\rho) \rangle .$$

In order to further simplify (2.2) we make the following assumptions:

(A) the total variations of $A_1(x)$ and $A_2(x)$ are small,

(B) the quantity $|\frac{v}{u}| \ll 1$, i.e., the flow is predominantly in the x-direction,

(C) $\langle f(\rho) \rangle = \hat{f}(\langle \rho \rangle)$ for some $\hat{f}$.

Then it is reasonable to assume that

(D)
$$\langle \rho u \rangle = \langle \rho \rangle \langle u \rangle \, ,$$
$$\langle \rho u^2 \rangle = \langle \rho \rangle \langle u \rangle^2$$

etc. An asymptotic analysis with respect to $|\frac{v}{u}|$ adds more plausibility to the equations (2.3). Thus (2.2) becomes

$$\rho_t + (\rho u)_x = - \frac{A'(x)}{A(x)} \rho u \, ,$$

$$(\rho u)_t + (\rho u^2 + P)_x = - \frac{A'(x)}{A(x)} \rho u^2 \, , \qquad (2.3)$$

$$P = f(\rho)$$

where we have made the following identifications

$$\langle \rho \rangle \sim \rho(x,t), \qquad \langle u \rangle \sim u(x,t)$$

etc.

System (2.3) is the one-dimensional approximation of (2.1). It should be pointed out that as far as the authors know there has not been a rigorous analysis of how reasonable the assumptions (A-D) are. Nevertheless, the system (2.3) is a mathematically tractible model of (2.1). It is believed that the study of (2.3) will shed some light to the structure of the solutions of the more difficult, but exact, equations of gas flow in a duct with variable cross section.

3. Formation of singularities for the equation $u_t + f(u)_x = A(x)u$. Before

proceeding with the solution to the Riemann problem for the system (2.3) it is

instructive to study how the spatial dependence of (2.3) enters as an important

feature in producing shocks. Consider the nonhomogeneous Berger's equation

$$u_t + uu_x = a(x)u . \tag{3.1}$$

The following proposition is in the same spirit as the ideas proposed in [6].

Proposition 3.1. If $a(x) > 0$ and $a'(x) < 0$ for all x, then the solution to

$$\begin{cases} u_t + uu_x = a(x)u \\ u(x,0) = u_0 \end{cases} \tag{3.2}$$

will form a shock at finite time for every positive initial value u_0.

The proof of the above proposition follows immediately from the following lemma

and corollary.

Lemma 3.1. Consider the initial value problem (3.2). Let $x(\xi,t)$ be the

characteristics defined by

$$\frac{dx}{dt} = u , \qquad x(0) = \xi$$

$$\frac{du}{dt} = a(x)u , \qquad u(0) = u_0 . \tag{3.3}$$

Then

$$x(\xi,t) = \int_0^t u(\xi,s)\,ds + \xi ,$$

$$u(\xi,t) = \int_\xi^{x(\xi,t)} a(s)\,ds + u_0 , \tag{3.4}$$

$$x_\xi(\xi,t) = G^{-1}(\xi,t)[1 - a(\xi) \int_0^t G(\xi,s)\,ds] , \tag{3.5}$$

where

$$G(\xi,t) = \exp(-\int_0^t a(x(\xi,s))ds) \ ,$$

$$u_\xi(\xi,t) = x_\xi(\xi,t)a(x(\xi,t)) - a(\xi) \ . \tag{3.6}$$

Proof. (3.4a) and (3.4b) follow from (3.3) by integrating with respect to t and x. (3.5) follows from (3.4) by observing that x_ξ satisfies

$$x_{\xi t} = a(x(\xi,t))x_\xi - a(\xi) \ .$$

(3.4a) follows by integrating (3.3a) in the time interval $[0,t]$. Since

$$\frac{du}{dx} = a(x) \ , \tag{3.7}$$

(3.4b) is then easily obtained by integrating (3.7) in the x-direction. (3.5) and (3.6) are subsequently derived from (3.4).

Corollary 3.1. Assume that u_0 is positive.

(1) If $a(x) \leqslant 0$ for all x then (3.2) has a global smooth solution.

(2) If $a(x) > 0$ and $a'(x) > 0$ then

$$a(\xi) \int_0^\infty G(\xi,s)ds < 1 \tag{3.8}$$

and (3.2) has global smooth solutions.

(3) If $a(x) > 0$ for all x and $a'(x) < 0$ for all x then

$$a(\xi) \int_0^\infty G(\xi,s)ds > 1 \ . \tag{3.9}$$

Hence, there exists $\tilde{t}$ such that

$$x_\xi(\xi,\tilde{t}) = 0 \ .$$

If conditions of part 3 of Corollary 3.1 hold, then the solution forms a shock at least beginning from $\tilde{t}$.

Proof. The proof of this corollary relies essentially on (3.5). This equation provides us with a means of measuring at what rate characteristics starting at two

points ξ_1 and ξ_2 on the x-axis will approach each other. (1) Follows immediately from (3.5) since if $a(x) \leqslant 0$ then x_ξ is always positive and the characteristics will be expanding. The proof of (2) is essentially the same. Since $a(x) > 0$ a simple phase plane analysis shows that $x(\xi,t) > \xi$ for $t > 0$. Since a is increasing we have

$$a(x(\xi,t)) > a(\xi)$$

which, in turn, implies that

$$-\int_0^t a(x(\xi,s))ds < -ta(\xi) \ . \tag{3.10}$$

Exponentiating (3.10) and recalling the definition of $G(\xi,t)$ yields

$$\int_0^t G(\xi,s)ds < -\frac{1}{a(\xi)} e^{-ta(\xi)} + \frac{1}{a(\xi)} \ . \tag{3.11}$$

(3.8) then follows by letting t approach infinity. Thus, a global smooth solution to (3.2) exists since, as can be seen from (3.5), (3.8) forces the characteristics to expand. On the other hand, if a is decreasing, the inequality (3.8) is reversed and thus there exists $\tilde{t}$ such that

$$a(\xi) \int_0^{\tilde{t}} P(\xi,s)ds = 1 \ ,$$

or equivalently,

$$x_\xi(\xi,\tilde{t}) = 0 \ . \tag{3.12}$$

This completes the proof of Corollary 3.1.

Proof of Proposition 3.1. We note that $\tilde{t} = \tilde{t}(\xi)$. Now, we claim that there exists ξ_1 and ξ_2 such that

$$x(\xi_1,\tilde{t}(\xi)) = x(\xi_2,\tilde{t}(\xi)) = \tilde{x} \ . \tag{3.13}$$

By way of contradiction, assume for all τ_1 and τ_2, $\tau_1 \neq \tau_2$

$$x(\tau_1, \tilde{t}(\xi)) \neq x(\tau_2, \tilde{t}(\xi)) . \tag{3.14}$$

Let $f(\tau) = x(\tau, \tilde{t}(\xi))$. Then (3.14) implies that f is a monotone function and, therefore, $f'(\tau) = x_\xi(\tau, \tilde{t}(\xi)) \neq 0$ for all τ which contradicts (3.12). Hence, there are two characteristics starting at ξ_1 and ξ_2 which meet at $(\tilde{x}, \tilde{t})$. On the other hand, by the standard uniqueness theorem in ordinary differential equations, the above characteristics viewed in the (x, u) plane reach the line $x = \tilde{x}$ at two different values of u. Therefore, a smooth solution cannot exist in a neighborhood of $(\tilde{x}, \tilde{t})$. This completes the proof of Proposition 3.1.

Next, we turn to the question of the Riemann problem for the related equation

$$u_t + f(u)_x = g(x, u) . \tag{3.15}$$

We assume that f is genuinely nonlinear, i.e., $f'' > 0$. Consider (3.15) with the Riemann initial condition

$$u(x,0) = \begin{cases} u_r , & x > 0 \\ u_\ell , & x < 0 . \end{cases} \tag{3.16}$$

We will give a brief outline of how the local solution to (3.15)-(3.16) is constructed. Our claim is that the initial discontinuity (3.16) is immediately resolved by the corresponding conservation law

$$u_t + f(u)_x = 0 . \tag{3.17}$$

Then the term $g(x, u)$ governs the evolution of the resolved waves. Hence, to solve (3.15)-(3.16) we divide the problem into two cases:

Case A: The solution to (3.16)-(3.17) is a rarefaction. Let

$$u^0(\xi) = \begin{cases} u_\ell & \text{if } \xi < \xi_\ell \\ h(\xi) & \text{if } \xi_\ell < \xi < \xi_r \\ u_r & \text{if } \xi > \xi_r \end{cases} \tag{3.18}$$

be this solution, where

$$\xi_\ell = f'(u_\ell), \quad \xi_r = f'(u_r), \quad :'(b(x)) \quad x, \quad \xi = \frac{x}{t} \; .$$

Consider

$$\frac{dx}{dt} = f'(\tilde{u}), \qquad x(\xi,0) = 0$$

$$\frac{d\tilde{u}}{dt} = q(x,\tilde{u}), \qquad \tilde{u}(\xi,0) = u^0(\xi), \quad \xi_\ell \le \xi \le \xi_r \; . \tag{3.19}$$

Let $(\tilde{u}(\xi,t), x(\xi,t))$ be the solution of (3.19) on $\xi_\ell \le \xi \le \xi_r$. It is not difficult to show that $x_\xi(\xi,t) \ne 0$ for $\xi_\ell \le \xi \le \xi_r$. Thus

$$u(x,t) = \tilde{u}(\xi(x,t),t)$$

is a solution of (3.17)–(3.18) within the region $x_\ell(t) \le x \le x_r(t)$, with $x_i(t) = x(\xi_i,t), i = r,\ell.$

Case B: The solution to (3.16)–(3.17) is a shock. Let

$$u^0(\xi) = \begin{cases} u_\ell & \text{if} \quad \xi > s \\[2mm] u_r & \text{if} \quad \xi > s \end{cases} \tag{3.20}$$

be that solution with

$$s = \frac{f(u_r) - f(u_\ell)}{u_r - u_\ell} \; , \qquad \xi = \frac{x}{t} \; .$$

Then, in a similar manner to Case A we construct the solution to (3.15)–(3.16), namely,

$$u(x,t) = \begin{cases} u_\ell(x,t) & \text{if} \quad x < \tilde{x}(t) \\[2mm] u_r(x,t) & \text{if} \quad x > \tilde{x}(t) \end{cases} \tag{3.21}$$

where

$$
\begin{cases}
\dfrac{\partial}{\partial t}\, u_i + f(u_i)_x = q(x, u_i) \\[2mm]
u_i(x, 0) = u_i
\end{cases}
\qquad i = r, \ell \ , \qquad\qquad (3.22)
$$

and

$$
\frac{d\tilde{x}}{dt} = \frac{f(u_r(\tilde{x}, t)) - f(u_\ell(\tilde{x}, t))}{u_r(\tilde{x}, t) - u_\ell(\tilde{x}, t)} \ , \qquad \tilde{x}(0) = 0 \ . \qquad\qquad (3.23)
$$

4. **Solution to the Riemann problem for the equations (2.3).** In [3] Glimm proposed
an iterative scheme in order to obtain the solution to the initial value problem for
the general conservation law

$$u_t + f(u)_x = 0, \quad u(x,0) = u_0(x) \tag{4.1}$$

where $f : R^n \to R^n$ is smooth and genuinely nonlinear. The building block of this
iterative scheme is the solution to the associated Riemann problems

$$u_c(x,0) = \begin{cases} u_\ell, & x < c \\ \\ u_r, & x > c . \end{cases} \tag{4.2}$$

The set of step functions in (4.2) is chosen as a pointwise approximation of the
initial value. The new feature in the system (2.3) is the nonhomogeneity which is
due to the geometry of the duct. In this section we discuss the solution to the
Riemann problem for (2.3) which arises from the discretization of the initial
condition and the boundary of the duct. First we note that (2.3) can be written in
the form

$$(A\rho)_t + (A\rho u)_x = 0 ,$$
$$(A\rho u)_t + (A\rho u^2 + AP)_x = -A'P \tag{4.3}$$

where $A(x)$ has the form

$$A(x) = \begin{cases} 1 , & x < 0 \\ 1 - \varepsilon, & x > 0 \end{cases} \tag{4.4}$$

and

$$u_0(x) = \begin{cases} (\rho_-, u_-), & x < 0 \\ (\rho_+, u_+) & x > 0 . \end{cases} \tag{4.5}$$

As in [5], we assume that both the initial condition and the boundary of the duct

have small bounded variations. When $\varepsilon = 0$ (4.3)-(4.5) reduces to the classical Riemann problem for the equations of gas dynamics in a uniform tube [7]. For the case ε positive we apply the same ideas as in Section 3, namely, the solution to (4.3)-(4.5) can be viewed as a small perturbation of the solution to the corresponding problem when $\varepsilon = 0$. The implicit function theorem is the main tool in obtaining the exact solution of (4.3)-(4.5). To illustrate the method we choose a particular solution of the $\varepsilon = 0$ case and carry out the necessary calculations. Let

$$
(\rho(x,t), u(x,t)) = \begin{cases} (\rho_-, u_-), & 0 < \dfrac{x}{t} < s \\[2mm] (\rho_m, u_m), & s < \dfrac{x}{t} < (P'(\rho_m))^{1/2} \\[2mm] g(\dfrac{x}{t}), & (P'(\rho_m))^{1/2} < \dfrac{x}{t} < (P'(\rho_+))^{1/2} \end{cases}
$$

be the physically admissible solution to (4.3)-(4.5) with $\varepsilon = 0$, i.e., the solution to the Riemann problem consists of a backward shock $(\rho_-, u_-; \rho_m, u_m; s)$ and a forward rarefaction wave connecting (ρ_m, u_m) to (ρ_+, u_+) (cf. [7]). Then the solution to (4.3)-(4.5) with ε positive consists of a backward shock $(\rho_-, u_-; \rho_1(\varepsilon), u_1(\varepsilon); s(\varepsilon))$, a discontinuity $(\rho_1(\varepsilon), u_1(\varepsilon); \rho_2(\varepsilon), u_2(\varepsilon); 0)$ which is due to the geometry of the duct, and a forward rarefaction connecting $(\rho_2(\varepsilon), u_2(\varepsilon))$ to (ρ_+, u_+). The five formulae relating $s(\varepsilon)$, $\rho_1(\varepsilon)$, $u_1(\varepsilon)$, $\rho_2(\varepsilon)$ and $u_2(\varepsilon)$ are

$$
s(\rho_1 - \rho_-) = \rho_1 u_1 - \rho_- u_- ,
$$
$$
s(\rho_1 u_1 - \rho_- u_-) = \rho_1 u_1^2 + P(\rho_1) - \rho_- u_-^2 - P(\rho_-) ,
$$
$$
(1 - \varepsilon)\rho_2 u_2 = \rho_1 u_1 , \tag{4.7}
$$
$$
(1 - \varepsilon)\{\rho_2 u_2^2 + P(\rho_2)\} = \rho_1 u_1^2 + P(\rho_1) + \frac{1}{2}\varepsilon P(\rho_1) - \frac{1}{2}\varepsilon P(\rho_2) ,
$$

$$
u_2 = u_+ + \int_{\rho_+}^{\rho_2} \frac{\sqrt{P'(\rho)}}{\rho} \, d\rho .
$$

After eliminating s and u_1 from (4.7) we can formulate the above problem in the form

$$\underset{\sim}{F}(\rho_1, \rho_2, u_2, \epsilon) = 0 \tag{4.8}$$

where $\underset{\sim}{F} = (F_1, F_2, F_3)$ is

$$F_2(\rho_1, \rho_2, u_2, \epsilon) = (1-\epsilon)\rho_2 u_2 - \rho_1 u_- + \left\{\frac{\rho_1}{\rho_-}(P(\rho_1) - P(\rho_-))(\rho_1 - \rho_-)\right\}^{1/2},$$

$$F_2(\rho_1, \rho_2, u_2, \epsilon) = u_2 - u_+ - \int_{\rho_+}^{\rho_2} \frac{\sqrt{P'(\rho)}}{\rho}\, d\rho, \tag{4.9}$$

$$F_3(\rho_1, \rho_2, u_2, \epsilon) = \rho_1(1-\epsilon)\{\rho_2 u_2^2 + P(\rho_2)\} - (1-\epsilon)^2 \rho_2^2 u_2^2 - \rho_1 P(\rho_1) - \frac{1}{2}\epsilon\rho_1 P(\rho_1)$$

$$+ \frac{1}{2}\epsilon\rho_1 P(\rho_2).$$

We point out that the sign of the square root in (4.9) is chosen so that the usual entropy condition is satisfied [1]. (4.8)–(4.9) is now set up for applying the implicit function theorem. The problem is solved if we can uniquely determine ρ_1, ρ_2, and u_2 in terms of ϵ. A rather tedious calculation leads to the Jacobian of (4.9):

$$\det \left.\frac{\partial(F_1, F_2, F_3)}{\partial(\rho_1, \rho_2, u_2)}\right|_{\epsilon=0} = 2\rho_m(u_m^2 - P'(\rho_m))[P'(\rho_m)]^{1/2}. \tag{4.10}$$

Thus, if the initial step $(\rho_-, u_-; \rho_+, u_+)$ is such that

$$u_\pm^2 - P'(\rho_\pm) \neq 0 \tag{4.11}$$

i.e., the original flow of the gas is either subsonic or supersonic, we see that the small variation in $|\rho_+ - \rho_-|$ implies that

$$u_m^2 - P'(\rho_m) \neq 0. \tag{4.12}$$

Therefore, the determinant in (4.10) is nonzero and we can uniquely solve for ρ_1, ρ_2, and u_2 in terms of ϵ. This completes the solution to the Riemann problem (4.3)–(4.5).

There are still two interesting problems in connection with (4.3) whose answers would be valuable both to the theory and the application. The first question is whether the above scheme actually converges to the weak solution of the initial value problem. The second question in how easily this scheme can be implemented numerically. We are presently studying these questions.

REFERENCES

[1] Lax, P. D., Hyperbolic Systems of Conservation Laws II, Comm. Pure Appl. Math. 10 (1957), 537-566.

[2] Hughes, T. J. R., Ph.D. thesis, University of California, Berkeley.

[3] Glimm, J., Solutions in the Large for Nonlinear Hyperbolic Systems of Equations, Comm. Pure Appl. Math. 18 (1965), 697-715.

[4] Chorin, A., Random Choice Solutions of Hyperbolic Systems, J. Comp. Phys. 22 (1976), 517.

[5] Lui, T. P., Quasilinear Hyperbolic Systems, preprint (1978).

[6] Lax, P. D., Development of Singularities of Solutions of Nonlinear Hyperbolic Partial Differential Equations, J. Math. Phys. 5 (1964), 611-613.

[7] Courant, R., and Friedrichs, K. O., Supersonic Flow and Shock Waves, Interscience Publishers, Inc., New York, 1948.

BEAM MOTIONS UNDER MOVING LOADS SOLVED BY FINITE ELEMENT
METHOD CONSISTENT IN SPATIAL AND TIME COORDINATES

Julian J. Wu
U. S. Army Armament Research and Development Command
Large Caliber Weapon Systems Laboratory
Benet Weapons Laboratory
Watervliet, NY 12189

ABSTRACT.   A solution formulation and numerical results are presented here for the time-dependent problem of beam deflections under a moving load which can be neither a force or a mass. The basis of this approach is the variational finite element discretization consistent in spatial and time coordinates. The moving load effect results in equivalanet stiffness matrix and force vector which are evaluated along the line of discontinuity in a time-length plane. Numerical results for several problems have been obtained. Some of which are compared with solution obtained by Fourier series expansions.

I.   INTRODUCTION.   A solution formulation and some numerical results are presented for beam motions subjected to moving loads. Most of the work on this problem has been related to rail and bridge design (see, for example, reference [1] and many papers cited there from 1910 to 1971). However, the application of the analysis can obviously be extended to tracks for rocket firing and to gun dynamics [2].

In Section II of this paper, a variational formulation for a moving force problem is described. Also given are the procedures which lead to finite element matrix equation. A detailed description of the treatment of a concentrated moving force is given in Section III. The variational problem associated with a gun tube dynamics is presented in Section IV. This gun-tube problem contains the moving mass problem as a special case. Finite element solution can be derived from this formulation, but the details of this more complicated problem is omitted from the present paper. Some of the numerical results obtained for a moving force problem are reported in the last section and are compared with results obtained from series solutions.

II.   SOLUTION FORMULATION FOR A MOVING FORCE PROBLEM.   In this section, the solution formulation will be described in details for a moving force problem. The moving mass problem will be included as a special case of a more general problem of gun motions analysis given in a later section.

Consider a vertical force P moving on an Euler-Bernoulli beam. The differential equation is given by

$$EIy'''' + \rho A \ddot{y} = P \bar{\delta}(x-\bar{x}) \qquad (1)$$

355

where $y(x,t)$ denotes the beam deflection as a function of spatial coordinate x and time t. E, I, A, ρ denote elastic modulus, second moment of inertia, area and material density respectively. A dirac function is denoted by δ, $x = x(t)$ is the location of I, a prime (') denotes differentiation with respect to x and a dot ($\cdot$), differentiation with respect to t.

Introducing nondimensional quantities

$$\hat{y} = y/\ell \; , \quad \hat{x} = x/\ell \; , \quad \hat{t} = t/T \; , \tag{2}$$

where ℓ is the length of the beam and T is a finite time, within which the problem is of interest, Eq. (1) can be written as

$$y'''' + \gamma^2 \ddot{y} = \bar{Q}\bar{\delta}(x-\bar{x}) \tag{3}$$

The hats ($\hat{}$) have been omitted in Eq. (3) and

$$t = \frac{t}{T}$$

$$Q = \frac{P\ell^2}{EI} \tag{4}$$

with

$$C^2 = \frac{\rho A \ell^4}{EI}$$

Boundary conditions associated with Eqs. (1) or (3) will now be introduced in conjunction of a variational problem. Consider

$$\delta I = 0 \tag{5a}$$

with

$$I = \int_0^1 \int_0^1 [y''y\star'' - \gamma^2 \ddot{y}y\star - \bar{Q}\bar{\delta}(x-\bar{x})]dxdt$$

$$+ \int_0^1 dt\{k_1 y(0,t)y\star(0,t) + k_2 y'(0,t)y\star'(0,t)\}$$

$$+ \gamma^2 \int_0^1 dx_1 k_7 [y(x,0) - Y(x)]y\star(x,1)\} \tag{5b}$$

where $y\star(x,t)$ is the adjoint variable of $y(x,t)$. If one takes the first variation of I considering $y(x,t)$ to be fixed:

$$(\delta I)_{\delta y=0} = 0 \tag{5a'}$$

and consider δy^* to be completely arbitrary, it is easy to see that Eqs. (5) is equivalent to the differential equation (3) and the following boundary and initial conditions.

$$y'''(0,t) + k_1 y(0,t) = 0$$

$$y''(0,t) - k_2 y'(0,t) = 0$$

$$y'''(1,t) - k_3 y(1,t) = 0 \qquad 0 \le t \le 1 \qquad (6a)$$

$$y''(1,t) + k_4 y'(1,t) = 0$$

$$\dot{y}(x,0) = 0$$

and

$$\dot{y}(x,1) - k_5[y(x,0) - Y(x)] = 0 \qquad 0 \le x \le 1 \qquad (6b)$$

Taking appropriate values for k_1, k_2, k_3, and k_4, problems with a wide range of boundary conditions can be realized. The initial conditions in Eqs. (6b) are that the beam has zero initial velocity, and, if one takes k_5 to be ∞ (or larger number compared with unity),

$$y(x,0) = Y(x) \qquad (6b')$$

The meaning for cases where k_5 is not so need not be our concern here.

To derive the finite element matrix equations, one begins with Eq. (5a') and write

$$(\delta I)_{\delta y=0} = 0 \qquad (7a)$$

$$= \int_0^1 \int_0^1 [y'' \delta y^{*''} - \gamma^2 \dot{y} \dot{\delta y}^* - Q\bar{\delta}(x-\bar{x})\delta y''] dx dt$$

$$+ \int_0^1 dt [k_1 y(0,t)\delta y^*(0,t) + k_2 y'(0,t)\delta y^*(0,t)$$

$$+ k_3 y(1,t)\delta y^*(1,t) + k_4 y'(1,t)\delta y^{*'}(1,t)]$$

$$+ \int_0^1 dx [\gamma^2 k_5 y(x,0) - Y(x)]\delta y^*(x,1) \qquad (7b)$$

357

Introducing element local variables

$$\xi = \xi^{(i)} = Kx - i + 1$$
$$\eta = \eta^{(i)} = Lt - j + 1 \tag{8a}$$

or

$$x = \frac{1}{K}(\xi + i - 1)$$

$$t = \frac{1}{L}(\xi + j - 1) \tag{8b}$$

where K is the number of divisions in x and L, in t. (A typical grid scheme is shown in Figure 1). Equation (7b) can now be written as

$$\sum_{i=1}^{K}\sum_{j=1}^{L}\int_0^1\int_0^1 \frac{K^3}{L}y''_{(ij)}\delta y^{*''}_{(ij)} - \frac{\gamma^2 L}{K}y_{(ij)}\delta y^{*}_{(ij)}\,d\xi\,d\eta$$

$$+\sum_{j=1}^{L}\int_0^1 d\eta \left[\frac{k_1}{L}y_{(ij)}(0,\eta)\delta y^{*}_{(ij)}(0,\eta) + k_2\frac{K}{L}y'_{(ij)}(0,\eta)\delta y^{*'}_{(ij)}(0,\eta)\right.$$

$$+\sum_{i=1}^{K}\int_0^1 \frac{d\xi}{K}\left\{\gamma^2 k_5(y_{(ij)}(\xi,0)\delta y^{*}_{(ij)}(\xi,1))\right\}$$

$$=\sum_{i=1}^{K}\sum_{j=1}^{L}\frac{Q}{L}\int_0^1\int_0^1 \bar{\delta}(x-\bar{x})\delta y^{*}_{(ij)}(\xi,\eta)d\xi\,d\eta$$

$$+\sum_{i=1}^{K}\frac{\gamma^2 k_5}{K}\int_0^1 d\xi\,[Y_{(i)}(\xi)\delta y^{*}_{(iL)}(\xi,1)] \tag{9}$$

The shape function vector is now introduced. Let

$$y_{(ij)}(\xi,\eta) = \underset{\sim}{a}^T(\xi,\eta)Y_{(ij)}$$

$$y^{*}_{(ij)}(\xi,\eta) = \underset{\sim}{a}^T(\xi,\eta)Y^{*}_{(ij)} = Y^{*T}_{(ij)}\underset{\sim}{a}(\xi,\eta) \tag{10}$$

Equation (9) then becomes

$$\sum_{i=1}^{K} \sum_{j=1}^{L} \delta Y^{\star T}_{(ij)} \{\frac{k_1}{L} \underline{A} - \frac{k_2}{K} \underline{B}^{\star}\} Y_{(ij)}$$

$$+ \sum_{i=1}^{L} \delta Y^{\star T}_{(ij)} \{\frac{k_1}{L} \underline{B}_1 + \frac{k_2 K^2}{L} \underline{B}_2\} Y_{(ij)}$$

$$+ \sum_{i=1}^{L} \delta Y^{\star T}_{(Kj)} \{\frac{k_3}{L} \underline{B}_3 + \frac{k_4 K^2}{L} \underline{B}_4\} Y_{(ij)}$$

$$+ \sum_{i=1}^{K} \delta Y^{\star T}_{(iL)} \{\frac{Y^{\star} k_5}{K} \underline{B}_5\} Y_{(iL)}$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{L} \delta Y^{\star T}_{(ij)} \frac{Q}{L} \underline{F}_{(ij)} + \sum_{i=1}^{K} \delta Y^{\star T}_{(iL)} \frac{Y^{\star} k_5}{K} \underline{G}_{(i)} \qquad (11)$$

where, as it can be seen easily, that

$$\underline{A} = \int_0^1 \int_0^1 \underline{a}_{,\xi\xi} \, \underline{a}^T_{,\xi\xi} \, d\xi d\eta$$

$$\underline{B} = \int_0^1 \int_0^1 \underline{a}_{,\eta} \, \underline{a}^T_{,\eta} \, d\xi d\eta$$

$$\underline{B}_1 = \int_0^1 \underline{a}(0,\eta)\underline{a}^T(0,\eta)d\eta \quad , \quad \underline{B}_2 = \int_0^1 \underline{a}_{,\xi}(0,\eta)\underline{a}^T_{,\xi}(0,\eta)d\eta$$

$$(12)$$

$$\underline{B}_3 = \int_0^1 \underline{a}(1,\eta)\underline{a}^T(1,\eta)d\eta \quad , \quad \underline{B}_4 = \int_0^1 \underline{a}_{,\xi}(1,\eta)\underline{a}^T_{,\xi}(1,\eta)d\eta$$

$$\underline{B}_5 = \int_0^1 \underline{a}(\xi,1)\underline{a}^T(\xi,0)d\xi$$

and

$$\underline{F}_{(ij)} = \int_0^1 \int_0^1 \underline{a}(\xi,\eta)\bar{\delta}_{(ij)}(\xi-\bar\xi)d\xi d\eta \quad , \quad \underline{G}_{(i)} = \int_0^1 \underline{a}(\xi,1)Y_{(i)}(\xi)d\xi$$

Editorial remark: For definition of $\bar{}_{(ij)}$ see top of page containing equation (18a).

Now Eq. (11) can be assembled in a global matrix equation

$$\delta Y^{*T} \, K \, Y = \delta Y^* \, F \tag{13}$$

By virtue of the fact that δY^* is not subjected to any constrained conditions, one has

$$K \, Y = F \tag{14}$$

which can be solved routinely. Numerical results of several problems in this class will be presented in a later section.

III. FORCE VECTOR DUE TO A MOVING CONCENTRATED LOAD. We shall describe here the procedures involved to arrive at the force vector contributed by a moving concentrated load. This force vector has appeared in Eq. (12) as

$$F_{(ij)} = \int_0^{t_1} \int_0^{1} a(\xi, \tau) \, \bar{X}_{(ij)}(\xi - \tau) \, d\xi \, d\tau \tag{15}$$

The shape function $a(\xi, \tau)$ is a vector of 16 in dimension. In the present formulation we have chosen the form:

$$a_K(\xi, \tau) = \bar{b}_i(\xi)\bar{b}_j(\tau) \quad , \qquad \begin{array}{l} k = 1,2,3,\ldots 16 \\ i,j = 1,2,3,4 \end{array} \tag{16}$$

The relations between k and i,j are given in Table I. These are the consequences of the choice of the shape function such that $\bar{Y}_{(ij)}$, the generalized coordinates of the (ij)th element, represent the displacement, slope, velocity, and angular velocity at the local nodal points. Thus

$$\bar{b}_i(\xi) = \sum_{p=1}^{4} \bar{b}_{ip} \xi^{p-1} \tag{17}$$

The values of $\bar{b}_{ip}$ are given in Table II.

TABLE I. RELATIONSHIP BETWEEN (i,j) AND k IN EQUATION (16)

k	(i,j)	k	(i,j)
1	(1,1)	9	(1,3)
2	(2,1)	10	(2,3)
3	(1,2)	11	(1,4)
4	(2,2)	12	(2,4)
5	(3,1)	13	(3,3)
6	(4,1)	14	(4,3)
7	(3,2)	15	(3,4)
8	(4,2)	16	(4,4)

TABLE II. VALUES OF $\bar{b}_{ip}$ IN EQUATION (17)

p i	1	2	3	4
1	1	0	-3	1
2	0	1	-2	1
3	0	0	3	-2
4	0	0	-1	1

Now, let us consider $\bar{\delta}_{(ij)}(\xi-\bar{\xi})$. This "function" represents the effect of the Dirac delta function $\delta(x-\bar{x})$ on the (ij)th element. If the curve of travel $x = \bar{x}(t)$ does not go through the element (i,j), $\delta_{(ij)}(\xi-\bar{\xi})$. If it passes through that element, one has

$$\bar{\delta}_{(ij)}(\xi-\bar{\xi}) = \bar{\delta}(x-\bar{x}) = K\bar{\delta}(\xi-\bar{\xi}) \tag{18a}$$

with

$$\bar{\xi} = \bar{\xi}(\eta) \tag{18b}$$

The function $\bar{\xi}(\eta)$ is derived from $\bar{x} = \bar{x}(t)$. For example, if the force moves with a constant velocity, one has

$$\bar{x} = \bar{x}(t) = vt \tag{19a}$$

it follows from Eqs. (8) that

$$\bar{\xi} = \bar{\xi}(\eta) = -i+1 + \frac{vK}{L}(\eta+j-1) \tag{19b}$$

With Eqs. (16), (17), (18), and (19), one writes (15) as

$$F_{(ij)k} = K \int_0^1 \int_0^1 a_k(\xi,\eta)\bar{\delta}(\xi-\bar{\xi})d\xi d\eta \tag{20a}$$

$$F_{(ij)k} = K \int_0^1 \int_0^1 \bar{b}_{ip}\bar{b}_{jq}\, \xi^{p-1}\eta^{q-1}\bar{\delta}(\xi-\bar{\xi})d\xi d\eta \tag{20b}$$

Equation (20) can then be evaluated easily once the exact form of $\bar{\xi}$ is written. For example, if $\bar{\xi} = \eta$, Eq. (20) reduces to

$$F_{(ij)k} = \sum_{p=1}^{4}\sum_{q=1}^{4} K\bar{b}_{ip}\bar{b}_{iq}\int_0^1 \xi^{p+q-2}d\xi$$

$$= \sum_{p=1}^{4}\sum_{q=1}^{4}\frac{K\bar{b}_{ip}\bar{b}_{jq}}{p+q-1} \tag{21}$$

IV. A GUN DYNAMICS PROBLEM AND THE MOVING MASS PROBLEM. In this section, the solution formulation of a gun tube can be obtained as a special case to the gun tube motion problem. The differential equation of this problem can be written as [3]:

$$(EIy'')'' + [P(x,t)y']' + \rho A \ddot{y}$$

$$= - \bar{P}(x,t)y''(x,t)H(\bar{x}-x)$$

$$- m_p[\bar{x}^2 y'' + 2\bar{x}y' + y]\bar{\delta}(\bar{x}-x)$$

$$+ (m_p\, g\, \cos\,\alpha)\bar{\delta}(\bar{x}-x) + \rho A\, g\, \cos\,\alpha \qquad (22)$$

The notations are the same as in the previous section if they have already been defined. The "gun tube" is replacing the "beam" whenever appropriate. The new notations are defined here:

$P(x,t) = \pi R^2(x)p(t)$ = axial force in the tube due to internal pressure alone

$R(x)$ = inner radius of tube

$p(t)$ = internal pressure

$$\bar{P}(x,t) = [-\bar{P}(0,t) + g(\sin\,\alpha) \int_0^x \rho A dx]\, \frac{\int_0^x \rho A dx}{\int_0^x \rho A dx} \qquad (23)$$

= recoil force including tube inertia in axial direction.

$H(x)$ = Heaviside step function

$\bar{x} = \bar{x}(t)$ = position of the projectile

m_p = mass of projectile

g = gravitational acceleration

α = angle of elevation

With similar nondimensionalization as before and assuming that the cross-section is uniform, ballistic pressure is not time dependent. Equation can be written in dimensionless form

$$y'''' + [-\bar{P} + g\sin\alpha][(1-x)y']' + \gamma^2\ddot{y}$$

$$= -\bar{P}\,y''H(\bar{x}-x)$$

$$- \gamma^2 m_p[\bar{x}^2y'' + 2\bar{x}y' + y]\bar{\delta}(\bar{x}-x)$$

$$+ m_p\,g(\cos\alpha)\bar{\delta}(\bar{x}-x) + g(\cos\alpha) \tag{24}$$

Here, now, everything is dimensionless and

$$\gamma^2 = \frac{C^2}{T^2} = \frac{1}{T^2}\frac{\mu A\ell^4}{EI} \tag{25}$$

It is also clear that if one drops the second term on the left hand side and the first and the last terms on the right hand side in the above equation, the equation becomes that for a moving mass problem.

A variational problem associated with the differential equation of Eq. (24) can be obtained through integration-by-parts.

$$\delta I = (\delta I)_y = \sum_{i=1}^{12}(\delta I_i)_y - \sum_{j=1}^{3}(\delta J_j) = 0 \tag{26a}$$

with

$$I_1 = \int_0^1\int_0^1 y''y\star''\,dxdt \quad ; \quad I_2 = (\bar{P}-g\sin\alpha)\int_0^1\int_0^1 y'y\star'\,dxdt$$

$$I_3 = -\gamma^2\int_0^1\int_0^1 \ddot{y}y\star\,dxdt \quad ; \quad I_4 = -\bar{P}\int_0^1\int_0^1 y'y\star'H(\bar{x}-x)\,dxdt$$

$$I_5 = -\bar{P}\int_0^1\int_0^1 y'y\star'\bar{\delta}(\bar{x}-x)\,dxdt \quad ; \quad I_6 = -m_p\gamma^2\int_0^1\int_0^1 t^2 y'y\star'\bar{\delta}(\bar{x}-x)\,dxdt$$

$$I_7 = -m_p\gamma^2\int_0^1\int_0^1 t y'y\star'\bar{\delta}'(\bar{x}-x)\,dxdt \quad ; \quad I_8 = 2m_p\gamma^2\int_0^1\int_0^1 t y'y\star\bar{\delta}(\bar{x}-x)\,dxdt$$

$$I_9 = -m_p\gamma^2\int_0^1\int_0^1 \ddot{y}y\star\bar{\delta}(\bar{x}-x)\,dxdt \quad ; \quad I_{10} = -m_p\gamma^2\int_0^1\int_0^1 \dot{y}y\star\bar{\delta}(\bar{x}-x)\,dxdt$$

$$I_{11} = \int_0^1 \{k_1y(0,t)y\star(0,t) + k_2y(0,t)y\star(0,t) +$$

$$k_3y(1,t)y\star(1,t) + k_4y'(1,t)y\star'(1,t)\}dt$$

$$I_{12} = \kappa_7\int_0^1 y(x,0)y\star(x,1)\,dx \tag{26b}$$

and

$$J_1 = -g \cos \alpha \int_0^1 \int_0^1 y^* dx\, dt$$

$$J_2 = -gm \cos \alpha \int_0^1 \int_0^1 y^* \bar\delta(x-\bar x) dx\, dt$$

$$J_3 = \kappa_7 \int_0^1 Y(x) y^*(x,1)\, dx \tag{26c}$$

The variational problem also produces the following initial and boundary value conditions in addition to the differential equation:

$$\dot y(x,0) = 0$$

$$\dot y(x,1)[1 + m\bar\delta(\tfrac{1}{2}\beta-x)] + \kappa_7[y(x,0) - Y(x)] = 0 \tag{27a}$$

and

$$y''(0,t) - k_2 y'(0,t) = 0$$

$$y''(1,t) + k_4 y'(1,t) = 0$$

$$y'''(0,t) + \kappa_1 y(0,t) + (-\bar P + g\cos\alpha)y'(0,t) + \bar P y'(0,t)H(\tfrac{1}{2}\beta t^2)$$

$$+ m\beta^2 y'(0,t)\bar\delta(\tfrac{1}{2}\beta t^2) = 0$$

$$y'''(1,t) - k_3 y(1,t) + \bar P y'(1,t)H(\tfrac{1}{2}\beta t^2-1) + m\beta^2 y'(1,t)\bar\delta(\tfrac{1}{2}\beta t^2-1) = 0$$

$$\tag{27b}$$

Other than the fact that the present problem is much more complicated than the one associated with a moving force, the basic concept of solution used previously does not change and we shall omit the details of solution formulation here.

V. NUMERICAL DEMONSTRATIONS. Some numerical results obtained will now be presented. Let us consider a simply-supported beam subjected to a unit moving force with a constant velocity

$$v = \frac{\ell}{T}$$

As T varies from ∞ to 0, the velocity varies from 0 to ∞.

It will be helpful to compare v with some reference velocity which is a characteristic of the given beam. It is known that for a simply-supported beam, the first mode of vibration has a frequency (see, for example, [4])

$$f_1 = \frac{w}{2\pi} = \frac{1}{2\pi} \frac{\pi^2}{C} = \frac{\pi}{2C} \quad \text{(cycles per seconds)}$$

and the period,

$$T_1 = \frac{2C}{\pi}$$

where

$$C = \frac{\ell^2}{\pi}\sqrt{\frac{m}{EI}}$$

Consider the vibration as traveling waves. They travel at a speed

$$v_1 = 2\ell f_1 = \frac{\pi \ell}{C}$$

Hence, the relative velocity

$$\bar{v} = \frac{v}{v_1} = \frac{C}{\pi T} = \frac{T_1}{2T}$$

We shall take $C = 1.0$ for the moving force problems. Thus, $f_1 = \frac{\pi}{2} = 1.5708$ Hz. $T_1 = 0.6366$ sec. and

$$\bar{v} = \frac{1}{\pi T}$$

Using a grid scheme of 4 x 4, Tables III, IV, and V show the deflec-
tions as the concentrated force Q = 1.0 moves from the left end to the
right end of the beam. Since we have defined that T is the time required
for the load to travel from one end to another, t = 0.5T, for example,
denotes the point when the load is at the midspan of the beam if v is con-
stant. At t = 1.0 T, the force has reached the other end and the deflec-
tion should be zero in the static case.

Solutions by Fourier series [1] are also obtained and they are also
given in these tables (numbers in parentheses) for close comparisons.

Table III shows that for T = 100 sec, $\bar{v}$ = 1/300 or more or T is more
than 300 times the natural frequency T_1, the deflections as P moves across
the beam is nearly the static deflection. The dynamic effect of the load
in the case T = 100, as indicated by the deflection curve at t = 1.0 T is
indiscernible. For v = 1/3 and v = 3.33, the dynamic effect is very much
pronounced as indicated by Table IV and V. The agreement between the
present results compared reasonably well with the series solution in
Tables II and III. It is extremely well in case of nearly static cases as
shown in Table III.

TABLE III. DEFLECTION OF A SIMPLY SUPPORTED BEAM UNDER A MOVING LOAD
(T = 100 sec.)

$y(x,t)/\ell$ [x 10^{-1}]

t/T \\ x/ℓ	0.	0.25	0.50	0.75	1.00
0.	0. (0.)	0. (0.)	0. (0.)	0. (0.)	0. (0.)
0.25	0. (0.)	.1172 (.1167)	.1432 (.1426)	.0911 (.0907)	0. (0.)
0.50	0. (0.)	.1431 (.1433)	.2082 (.2085)	.1431 (.1434)	0. (0.)
0.75	0. (0.)	.0908 (.0916)	.1427 (.1438)	.1168 (.1176)	0. (0.)
1.00	0. (0.)	-.0047 (-.0002)	-.0066 (-.0003)	-.0046 (-.0002)	0. (0.)

367

TABLE IV. DEFLECTION OF A SIMPLY SUPPORTED BEAM UNDER A MOVING FORCE
(T = 1.0 sec)

$y(x,t)/\ell \ [x \ 10^{-4}]$

x/ℓ t/T	0.	0.25	0.50	0.75	1.00
0.	0. (0.)	0. (0.)	0. (0.)	0. (0.)	0. (0.)
0.25	0. (0.)	.09489 (.09795)	.11349 (.11415)	.07108 (.06942)	0. (0.)
0.50	0. (0.)	.20542 (.20602)	.30502 (.30257)	.21491 (.21126)	0. (0.)
0.75	0. (0.)	.03369 (.05829)	.09522 (.09397)	.09641 (.08092)	0. (0.)
1.00	0. (0.)	-.10299 (.01995)	(.05193) (.03153)	(.11574) (.02405)	0. (0.)
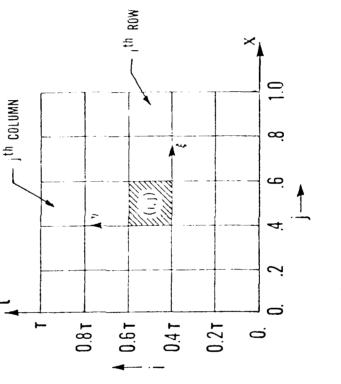
TABLE V. DEFLECTION OF A SIMPLY SUPPORTED BEAM UNDER A MOVING FORCE
(T = 0.1 sec)

$y(x,t)/\ell \ [x \ 10^{-4}]$

x/ℓ t/T	0.	0.25	0.50	0.75	1.00
0.	.0 (.0)	.0 (.0)	.0 (.0)	.0 (.0)	.0 (.0)
0.25	.0 (.0)	.0619 (.0645)	-.0148 (-.0149)	.0043 (.0033)	.0 (.0)
0.50	.0 (.0)	.2002 (.1952)	.1228 (.1262)	-.0594 (-.0479)	.0 (.0)
0.75	.0 (.0)	.3007 (.2929)	.3837 (.3849)	.0770 (.0801)	.0 (.0)
1.00	.0 (.0)	.4601 (.4018)	.4912 (.4880)	.5767 (.5959)	.0 (.0)

REFERENCES

1. L. Fryba, *Vibrations of Solids and Structures Under Moving Loads*, Noordhoff International Publishing Company, Groningen, 1971.

2. J. J. Wu, "The Initial Boundary Value of Gun Dynamics Solved by Finite Element Unconstrained Variational Formulations," Innovative Numerical Analysis For the Applied Engineering Science, R. P. Shaw, et al., Editors, University Press of Virginia, Charlottesville, 1980, pp. 733-741.

3. J. J. Wu, "A Computer Program and Approximate Solution Formulation For Gun Motions Analysis," Technical Report ARLCB-TR-79019, US Army Armament Research and Development Command, Benet Weapons Laboratory, June 1979.

4. K. N. Tong, *Theory of Mechanical Vibration*, John Wiley, New York, 1960, p. 257; p. 308.

FIGURE 1. A Typical Finite Element Grid Scheme Showing the (i,j)th Element and the Global, Local Coordinates.

WAVE PROPAGATION IN PERIODICALLY LAYERED MEDIA

T. C. T. Ting
Department of Materials Engineering
University of Illinois at Chicago Circle
Chicago, Illinois, 60680

ABSTRACT. Wave propagation normal to the layering of a periodically layered medium is studied. The layered medium can be finite or semi-infinite in extent. Each period consists of two layers of linear elastic or viscoelastic materials. The medium is initially at rest and at time $t = 0$ a transient wave is generated by the prescribed boundary conditions. The stress response at a finite x is obtained by the analogy between the exact solution at the centers of odd layers in the layered medium and the solution in a homogeneous viscoelastic medium. In the case of a semi-infinite layered medium, the stress response at a large x is obtained by an asymptotic analysis. For the value of x which is not very large, higher order asymptotic solutions are given. Numerical examples are presented for an elastic composite subjected to a unit step stress in time applied at $x = 0$.

I. INTRODUCTION. Most of the approximate theories for wave propagation in a layered medium focus on the determination of the dispersion relation or the frequency equation due to a harmonic oscillation [1-4], although some of the theories are able to predict the late-time asymptotic solution in a semi-infinite layered medium due to a step load applied at the boundary. For the latter, exact theories may be used to find the asymptotic solution and the wave front solution [5-7].

To predict the transient response at points not necessarily far away from the impact end (where the asymptotic solution does not apply) and to points not necessarily near the wave front, a new theory based on the analogy between the dynamic response of a semi-infinite layered medium and a semi-infinite homogeneous viscoelastic medium has been proposed recently by Ting and Mukunoki [8]. The fundamental idea is to characterize the layered medium by an "equivalent" homogeneous viscoelastic medium such that the dynamic response of the latter is identical to that of the layered medium at the centers of the alternate layers. Although the idea of modeling a composite by a viscoelastic medium is not new [9,10], the "theory of viscoelastic analogy" introduced in [8] succeeds in correlating precisely the analogy between a layered medium and a homogeneous viscoelastic medium. Since wave propagation in a homogeneous linear viscoelastic medium can be solved easily by many known numerical schemes (see [11], for example), one can obtain the

transient wave solution in a layered medium by solving the transient waves in the "equivalent" homogeneous viscoelastic medium.

The theory of viscoelastic analogy presented in [8] is for a semi-infinite medium. In this paper we present a more general theory which applies to a finite layered medium.

Consider a periodic layered medium as shown in Fig. 1 in which each period 2ω consists of two layers of homogeneous, isotropic, linear elastic or viscoelastic materials. The two different materials in the layers will be designated as material 1 and 2, respectively. Thus material 1 occupies layers 1,3,5,... while material 2 occupies layers 2,4,6,... The thicknesses of individual layers are denoted by $2h_i$ $(i=1,2)$ where the subscripts 1 and 2 refer to material 1 and 2, respectively. We will assume that the layered medium is initially at rest and occupies the region $0 \leq x \leq \ell$. We choose the central surface of layer 1 as $x=0$ and the other boundary, $x=\ell$, is assumed to be at the central surface of layer N where N can be an even or odd integer. Hence,

$$\ell = (N-1)\omega \tag{1}$$

We will consider plane wave propagation in the direction x in which the only non-vanishing component of the displacement is in the x direction. We therefore have a one-dimensional wave propagation problem in which the equation of motion and the continuity of the displacement are given by

$$\frac{\partial \sigma_i}{\partial x} = \rho_i \dot{v}_i , \qquad (i = 1,2) \tag{2}$$

$$\frac{\partial v_i}{\partial x} = \dot{\varepsilon}_i , \qquad (i = 1,2) \tag{3}$$

where a dot stands for differentiation with respect to the time t, and $\sigma_i, \varepsilon_i, v_i, \rho_i$ $(i=1,2)$ are the normal stress, normal strain, particle velocity and mass density, respectively. Let $\lambda_i(t)$ and $\mu_i(t)$ be the relaxation functions of the materials. For elastic materials, $\lambda_i(t)$ and $\mu_i(t)$ are independent of t and are identified as Lamé constants. The stress-strain relation can be written in the form of Stieltjes convolution

$$\sigma_i(x,t) = \int_{0^-}^{t} g_i(t-t') d\varepsilon_i(x,t') , \tag{4}$$

$$g_i(t) = \lambda_i(t) + 2\mu_i(t) . \tag{5}$$

where we have assumed that

$$\sigma_i(x,0^-) = v_i(x,0^-) = \varepsilon_i(x,0^-) = 0 . \tag{6}$$

II. GENERAL SOLUTION. The general solution to Eqs. (2-6) can be obtained by the method of Laplace transform and by the use of the Floquet theory. We define the Laplace transform, $\bar{f}(p)$, of a function $f(t)$ by

$$\bar{f}(p) = \int_{0^-}^{\infty} f(t)e^{-pt} \, dt \tag{7}$$

After applying the Laplace transform to Eqs. (2-6), the general solution for the stress and the velocity in layers 1 and 2 can be written as

$$\bar{\sigma}_1(x,p) = \bar{A}_1 \cosh(k_1 x) + \bar{B}_1 \sinh(k_1 x) \tag{8a}$$

$$\bar{v}_1(x,p) = \frac{1}{m_1}\left\{\bar{A}_1 \sinh(k_1 x) + \bar{B}_1 \cosh(k_1 x)\right\} \tag{8b}$$

$$\bar{\sigma}_2(x,p) = \bar{A}_2 \cosh(k_2 x - k_2 \omega) + \bar{B}_2 \sinh(k_2 x - k_2 \omega) \tag{8c}$$

$$\bar{v}_2(x,p) = \frac{1}{m_2}\left\{\bar{A}_2 \sinh(k_2 x - k_2 \omega) + \bar{B}_2 \cosh(k_2 x - k_2 \omega)\right\} \tag{8d}$$

where

$$\left.\begin{array}{l} \omega = h_1 + h_2 \\[2mm] k_i = \sqrt{\bar{\rho}_i p / \bar{g}_i} \\[2mm] m_i = \rho_i p / k_i = \sqrt{\rho_i p \bar{g}_i} \end{array}\right\} \tag{9}$$

$\bar{A}_i$ and $\bar{B}_i$ $(i = 1,2)$ are determined by the continuity condition at $x = h_1$

$$\begin{bmatrix} \bar{\sigma}_1 \\ \bar{v}_1 \end{bmatrix}(h_1,p) = \begin{bmatrix} \bar{\sigma}_2 \\ \bar{v}_2 \end{bmatrix}(h_1,p) \tag{10}$$

and the quasi-periodicity property of the solution together with the continuity condition at $x = 2\omega - h_1$

$$\begin{bmatrix} \bar{\sigma}_2 \\ \bar{v}_2 \end{bmatrix}(2\omega - h_1,p) = \begin{bmatrix} \bar{\sigma}_1 \\ \bar{v}_1 \end{bmatrix}(-h_1,p)e^{-2\omega} \tag{11}$$

373

where κ is the characteristic exponent [12]. Substitution of Eq. (8) into Eqs. (10) and (11) leads to four homogeneous equations for $\bar{A}_i$ and $\bar{B}_i$. The requirement for a non-trivial solution results in the following equation for the characteristic exponent κ:

$$\cosh (2\omega\kappa) = \hat{\theta} \cosh (2k_1 h_1 - 2k_2 h_2)$$

$$- (\hat{\theta}-1) \cosh (2k_1 h_1 - 2k_2 h_2) \tag{12}$$

$$\hat{\theta} = \frac{1}{4}\left(\frac{1}{m_1} + 2 + \frac{1}{m_2}\right) \tag{13}$$

Moreover, $\bar{A}_i$ and $\bar{B}_i$ are related by

$$\left.\begin{array}{ll} \dfrac{\bar{A}_2}{\bar{A}_1} = p\bar{M}e^{-\omega\kappa} \;, & \dfrac{\bar{B}_2}{\bar{A}_1} = -p\bar{R}m_2 e^{-\omega\kappa} \;, \\[3mm] \dfrac{\bar{B}_1}{\bar{A}_1} = -p\bar{L}_1 \;, & \dfrac{\bar{B}_2}{\bar{A}_2} = -p\bar{L}_2 \;, \end{array}\right\} \tag{14}$$

where

$$\left.\begin{array}{l} \bar{L}_1 = m_1 \bar{R} p\bar{M} \;, \qquad \bar{L}_2 = m_2 \bar{R}/(p\bar{M}) \\[3mm] p\bar{M} = \dfrac{m_1 C_1 C_2 + m_2 S_1 S_2}{m_1 \cosh(\omega\kappa)} = \dfrac{m_2 \cosh(\omega\kappa)}{m_2 C_1 C_2 + m_1 S_1 S_2} \\[3mm] p\bar{R} = \dfrac{m_2 C_1 S_2 + m_1 C_2 S_1}{m_1 m_2 \sinh(\omega\kappa)} = \dfrac{\sinh(\omega\kappa)}{m_2 C_1 S_2 + m_1 C_2 S_1} \\[3mm] C_i = \cosh(k_i h_i) \;, \qquad S_i = \sinh(k_i h_i) \end{array}\right\} \tag{15}$$

Notice that if we interchange the subscripts 1 and 2, the expression for $p\bar{R}$ remains unchanged while $p\bar{M}$ becomes $p\bar{M}^{-1}$. Therefore, we can obtain the Stieltjes inversion of $\bar{M}$ by simply interchanging the subscripts 1 and 2 in the expression for $p\bar{M}$ and applying the Laplace inverse transform.

With Eq. (14), the general solution in the layers 1 and 2 is expressed by Eq. (8) can now be reduced to a solution containing only one coefficient, say $\bar{A}_1$. The solutions in other layers are obtained by the quasi-periodicity relation:

$$\begin{bmatrix} \bar{u}_i \\ \bar{v}_i \end{bmatrix} (2n\omega + x, p) = \begin{bmatrix} \bar{u}_i \\ \bar{v}_i \end{bmatrix} (x, p) e^{-2n\omega\kappa} \tag{16}$$

where n is an integer. Moreover, we see from Eq. (12) that if κ is a characteristic exponent, so is $-\kappa$. Therefore, in addition to the general solution with $\bar{A}_1$ as the coefficient, we obtain the second general solution by changing the sign of κ. The coefficient of this second solution will be denoted by $\bar{A}_1'$. Consequently, the general solution for the stress and velocity at any point x in the layered medium can be written as, using Eqs. (8,14,16),

$$\bar{\sigma}_1(2n\omega + x_1, p) = \bar{A}_1\left\{\cosh(k_1 x_1) - p\bar{L}_1 \sinh(k_1 x_1)\right\}e^{-2n\omega\kappa}$$

$$+ \bar{A}_1'\left\{\cosh(k_1 x_1) + p\bar{L}_1 \sinh(k_1 x_1)\right\}e^{2n\omega\kappa} \qquad (17a)$$

$$\bar{v}_1(2n\omega + x_1, p) = \frac{\bar{A}_1}{m_1}\left\{\sinh(k_1 x_1) - p\bar{L}_1 \cosh(k_1 x_1)\right\}e^{-2n\omega\kappa}$$

$$+ \frac{\bar{A}_1'}{m_1}\left\{\sinh(k_1 x_1) + p\bar{L}_1 \cosh(k_1 x_1)\right\}e^{2n\omega\kappa} \qquad (17b)$$

$$\bar{\sigma}_2(2n\omega + \omega + x_2, p) = \bar{A}_1 p\bar{M}\left\{\cosh(k_2 x_2) - p\bar{L}_2 \sinh(k_2 x_2)\right\}e^{-(2n+1)\omega\kappa}$$

$$+ \bar{A}_1' p\bar{M}\left\{\cosh(k_2 x_2) + p\bar{L}_2 \sinh(k_2 x_2)\right\}e^{(2n+1)\omega\kappa} \qquad (17c)$$

$$\bar{v}_2(2n\omega + \omega + x_2, p) = \frac{\bar{A}_1}{m_2} p\bar{M}\left\{\sinh(k_2 x_2) - p\bar{L}_2 \cosh(k_2 x_2)\right\}e^{-(2n+1)\omega\kappa}$$

$$+ \frac{\bar{A}_1'}{m_2} p\bar{M}\left\{\sinh(k_2 x_2) + p\bar{L}_2 \cosh(k_2 x_2)\right\}e^{(2n+1)\omega\kappa} \qquad (17d)$$

where

$$-h_i \leq x_i \leq h_i , \qquad (i = 1, 2) \qquad (18)$$

When proper values for n and x_1 (or x_2) are chosen, Eqs. (17) can be used to determine solution at any point in the layered medium. The two coefficients $\bar{A}_1$ and $\bar{A}_1'$ are determined from the boundary conditions at $x = 0$ and $x = \ell$.

In the next section we will show the analogy between the solution at the centers of the layers and the solution in a homogeneous viscoelastic medium.

III. VISCOELASTIC ANALOGY. The stress and velocity at the centers of the layers have specially simple forms. By letting $x_1 = x_2 = 0$ in Eqs. (17), we have

$$\bar{\sigma}_1(2n\omega, p) = \bar{A}_1 e^{-2n\omega\kappa} + \bar{A}_1' e^{-2n\omega\kappa}$$

$$\bar{v}_1(2n\omega, p) = \frac{p\bar{L}_1}{m_1}\left(-\bar{A}_1 e^{-2n\omega\kappa} + \bar{A}_1' e^{-n\omega\kappa}\right)$$

$$\bar{\sigma}_2(2n\omega + \omega, p) = p\bar{M}\left(\bar{A}_1 e^{-(2n+1)\omega\kappa} + \bar{A}_1' e^{(2n+1)\omega\kappa}\right) \qquad (19)$$

$$\bar{v}_2(2n\omega + \omega, p) = \frac{p\bar{L}_2}{m_2} p\bar{M}\left(-\bar{A}_1 e^{-(2n+1)\omega\kappa} + \bar{A}_1' e^{(2n+1)\omega\kappa}\right)$$

We now consider a homogeneous, isotropic, linear viscoelastic medium which occupies $0 \le x \le \ell$ and which is at rest at $t = 0^-$ and is subjected to certain prescribed boundary conditions at $x = 0$ and $x = \ell$. Let Φ, η and V be the normal stress, normal strain and particle velocity, respectively. Also, let ρ and G be the "equivalent" mass density and the "equivalent" relaxation function of this homogeneous viscoelastic material. The equation of motion, the continuity condition, the stress-strain relation and the initial conditions are

$$\frac{\partial\Phi}{\partial x} = \rho\dot{V}$$

$$\frac{\partial V}{\partial x} = \dot{\eta}$$

$$\Phi(x, t) = \int_{0^-}^{t} G(t - t')\, d\eta(x, t') \qquad (20)$$

$$\Phi(x, 0^-) = V(x, 0^-) = \eta(x, 0^-) = 0$$

By applying the Laplace transform to Eqs. (20), the general solution for the stress and velocity will contain the exponential term

$$\exp\left(\pm \sqrt{\rho p/G}\ x\right) \qquad (21)$$

In view of the exponential terms in Eqs. (19), we will define the "equivalent" relaxation function $G(t)$ by the relation

$$\kappa = \sqrt{\rho p/G} \qquad (22)$$

We will also define the "equivalent" mass density ρ by the average mass density in the layered medium [4,8]:

$$\rho = (\rho_1 h_1 + \rho_2 h_2)/(h_1 + h_2) \qquad (23)$$

With Eq. (22), the general solution to Eq. (20) can be written as

$$\bar{\Phi}(x,p) = \bar{a}\, e^{-\kappa x} + \bar{a}'\, e^{\kappa x} \tag{24a}$$

$$\bar{V}(x,p) = \frac{\kappa}{cp}\left(-\bar{a}\, e^{-\kappa x} + \bar{a}'\, e^{\kappa x}\right) \tag{24b}$$

where $\bar{a}$ and $\bar{a}'$ are arbitrary functions of p.

There are several ways to identify the analogy between Eqs. (19) and (24). If the stress in material 1 is of main interest, we may set

$$\bar{A}_1 = \bar{a}, \qquad \bar{A}_1' = \bar{a}' \tag{25}$$

we then have

$$\left.\begin{aligned} \bar{\sigma}_1(x,p) &= \bar{\Phi}(x,p) \\[2mm] \bar{v}_1(x,p) &= p\bar{J}_1\bar{V}(x,p) \end{aligned}\right\} \qquad \text{for} \quad x = 2n\omega \tag{26a}$$

and

$$\left.\begin{aligned} \bar{\sigma}_2(x,p) &= p\bar{M}\bar{\Phi}(x,p) \\[2mm] \bar{v}_2(x,p) &= p\bar{M}\left\{p\bar{J}_2\bar{V}(x,p)\right\} \end{aligned}\right\} \qquad \text{for} \quad x = (2n+1)\omega \tag{26b}$$

where

$$\bar{J}_i = \frac{cp\bar{L}_i}{\kappa m_i}, \qquad (i = 1,2) \tag{27}$$

It should be pointed out that while $\bar{\Phi}$ and $\bar{V}$ as given by Eqs. (24) are defined for all x, Eqs. (26a) and (26b) apply only to $x = 2n\omega$ and $x = (2n+1)\omega$, respectively. By using the identity,

$$\bar{J}_1/\bar{J}_2 = (p\bar{M})^2 = m_2\bar{L}_1/(m_1\bar{L}_2) \tag{28}$$

the last of Eq. (26b) can be written as

$$\bar{v}_2(x,p) = \frac{1}{p\bar{M}}\left\{p\bar{J}_1\bar{V}(x,p)\right\}, \qquad x = (2n+1)\omega \tag{29}$$

With Eq. (29), we rewrite Eqs. (26) in the following form:

$$\sigma_1(2n\omega,t) = \Phi(2n\omega,t) \tag{30a}$$

$$v_1(2n\omega,t) = V^*(2n\omega,t) \tag{30b}$$

$$\tau_2(2n_2 + \omega, t) = \int_0^t M(t - t') \, dV_2(2n_2 + \omega, t') \qquad (30c)$$

$$v_2(2n_2 + \omega, t) = \int_0^t M^{-1}(t - t') \, dV^*(2n_2 + \omega, t') \qquad (30d)$$

where

$$V^*(x,t) = \int_0^t J_1(t - t') \, dV(x, t') \qquad (30e)$$

and M^{-1} is the Stieltjes inverse of M. See the discussion following Eq. (15) regarding the Stieltjes inverse of M. Thus the stress and velocity at the centers of the layers are related to the stress τ and velocity V in the "equivalent" homogeneous medium. In particular, the stress at the centers of the odd layers, $\sigma_1(2n_2, t)$, is identical to the stress τ in the "equivalent" homogeneous viscoelastic medium.

IV. SOLUTION AT FINITE n. In view of the link between the solution at the centers of the layered medium and that of a homogeneous viscoelastic medium derived above, one can solve the wave propagation problem in the layered medium by solving the wave propagation in the equivalent homogeneous viscoelastic medium. For this purpose, we have to find the relaxation function $G(t)$ of the equivalent viscoelastic medium from Eqs. (22,12). Analytical inversion of the Laplace transform $\bar{G}(p)$ does not appear feasible. Therefore, a method of numerical inversion such as the one used in [13] is employed.

To illustrate the theory of viscoelastic analogy, we consider a layered medium in which both materials 1 and 2 are elastic so that

$$g_1(t) = c_i H(t) \qquad (31)$$

where $H(t)$ is the Heaviside step function. An example of the equivalent viscoelastic relaxation function $G(t)$ obtained by numerically inverting the Laplace transform $\bar{G}(p)$ of Eq. (22) is shown in Fig. 2 along with physical parameters of the elastic layered medium used in the calculation. The physical parameters are taken from [4]. Unlike most real viscoelastic materials, the equivalent viscoelastic relaxation function $G(t)$ is not a monotonically decreasing function of t [16].

We can now replace the elastic layered medium by the homogeneous viscoelastic medium whose relaxation function is given by Fig. 2 to solve transient wave propagation problems. For simplicity, we consider the case in which the stress applied at $x = 0$ is a unit step function in t and $x = 0$. Using the method of characteristics,

we find the stress at the center of the 11th layer is also shown in Fig. 3. Fortunately for this example, the exact solution of wave propagation in the elastic layered medium due to a step load applied at $x = 0$ can be obtained numerically by using the ray theory and by keeping track of each transmitted and reflected amplitude at the interfaces of the layers. This exact solution by the ray theory is also shown in Fig. 3. It is seen that the agreement is excellent despite the fact that $G(t)$ obtained in Fig. 2 is quite crude.

In Fig. 4 we present the stress history at the center of the 3rd layer due to a step load applied at the center of the first layer. This represents a more severe test of the present theory. Notice that while the solution in Fig. 3 shows near asymptotic behavior of modulated oscillations, the solution in Fig. 4 shows no such asymptotic behavior. This is expected because the asymptotic solution does not apply to points near the impact end.

For a finite layered medium, we show in Fig. 5 the response at the center of the 5th layer subject to a unit step stress at the center of the first layer while the center of the 14th layer is fixed. Again, the agreement between the solution obtained by the theory of viscoelastic analogy and the exact solution by the ray theory is excellent.

Using Eqs. (5), one can find the velocity at the centers of the odd layers and the stress and velocity at the centers of the even layers. One can also find the stress and velocity at points which are not at the centers of the layers by using the method of characteristics for elastic and viscoelastic medium [14].
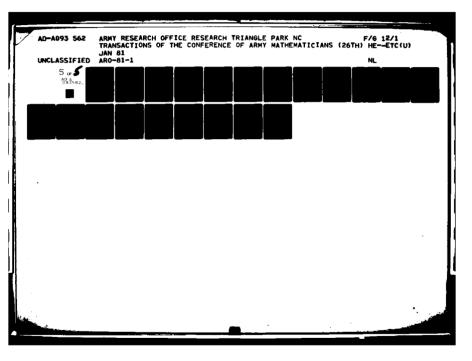
V. SOLUTION AT LAYER n. In this section we assume that each layer of the layered medium is elastic and consider the semi-infinite case so that $\gamma = 0$. The transient wave due to a unit step stress applied at $x = 0$ is obtained from the first of Eqs. (19):

$$\tau_1(2n\omega, p) = \frac{1}{p} e^{-2n\alpha k}$$

The inverse of this equation is

$$\tau(2n\omega, t) = \frac{1}{2\pi i} \int_{Br} \frac{1}{p} e^{pt - 2n\alpha k} dp \qquad (52)$$

where the subscript 1 of τ is omitted here and in the sequel.

Investigation of Eq. (52) in which k is given by Eq. (12) shows that $p = 0$ is a singular point as well as a saddle point. By expanding k in Taylor series in p, Eq. (52) can be written as

$$\sigma(2n\omega,t) = \frac{1}{2\pi i} \int_{Br} \frac{1}{z} e^{\tau z + (1/3)z^3 - \zeta_2 z^5 + \zeta_3 z^7 - \cdots} \, dz \qquad (33)$$

where

$$\tau = \left(\frac{c_\infty}{\omega} t - 2n\right) \Big/ \left(n\beta_1\right)^{1/3}$$

$$\zeta_2 = \beta_2 / \left(n\beta_1\right)^{2/3} \qquad\qquad\qquad (34)$$

$$\zeta_3 = \beta_3 / \left(n\beta_1\right)^{4/3}$$

c_∞ is the group velocity of the elastic layered medium and $\beta_1, \beta_2, \beta_3$ are constants depending on the geometry and material constants of the layered medium [15].

The solution which ignores the ζ_2 and ζ_3 terms is called the one-term asymptotic solution and can be expressed in terms of an integral of an Airy function [6,7]. If we retain the ζ_2 term but ignore the ζ_3 term, we have the two-term asymptotic solution. Finally, if both ζ_2 and ζ_3 terms are retained, we have the three-term asymptotic solution.

In [15], a particular Bromwich contour was selected and Eq. (33) was integrated numerically. A numerical example of the asymptotic solution for $n = 5$ is given in Fig. 6 for the elastic layered medium considered in Figs. 2-4. Comparison with the exact solution by the ray theory shows that the three-term asymptotic solution is satisfactory for this small value of $n = 5$.

A detailed discussion of when one-term, two-term and three-term asymptotic solutions may be considered a good approximation can be found in [15].
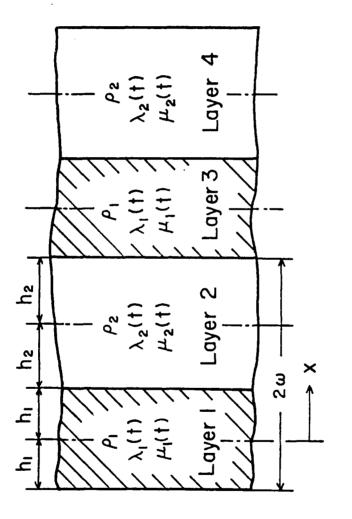
Although the layered medium is assumed to be elastic in this section, the analyses can be extended to viscoelastic layered medium. It is shown in [16] that when the layered medium is viscoelastic, the distance traveled by the wave should appear in the asymptotic analysis to provide a meaningful interpretation of the interaction between the dissipation and dispersion of the viscoelastic layered medium.
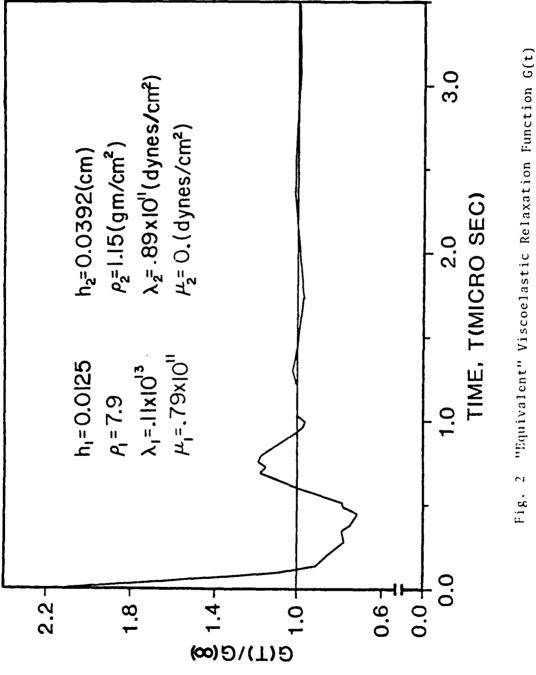
REFERENCES

[1] Sun, C. T., Achenbach, J. D. and Herımann, G., "Continuum Theory for a Laminated Medium," J. Appl. Mech., Vol. 35, 1968, 467-475.

[2] Stern, M., Bedford, A. and Yew, C. H., "Wave Propagation in Viscoelastic Laminates," J. Appl. Mech., Vol. 38, 1971, 448-454.

[3] Drumheller, D. S. and Bedford, A., "On a Continuum Theory for a Laminated Medium," J. Appl. Mech., Vol. 40, 1973, 527-532.

[4] Hegemier, G. A. and Nayfeh, A. H., "A Continuum Theory for Wave Propagation in Laminated Composites, Case 1: Propagation Normal to the Laminates," J. Appl. Mech., Vol. 40, 1973, 503-510.

[5] Peck, J. C. and Gurtman, G. A., "Dispersive Pulse Propagation Parallel to the Interfaces of a Laminated Composite," J. Appl. Mech., Vol. 36, 1969, 479-484.

[6] Sve, C., "Stress Wave Attenuation in Composite Materials," J. Appl. Mech., Vol. 39, 1972, 1151-1153.

[7] Chen, C. C. and Clifton, R. J., "Asymptotic Solutions for Wave Propagation in Elastic and Viscoelastic Bilaminates," Proc. 14th Midwestern Mech. Conf., Univ. of Oklahoma, 1975, 399-417.

[8] Ting, T. C. T. and Mukunoki, I., "A Theory of Viscoelastic Analogy for Wave Propagation Normal to the Layering of a Layered Medium," J. Appl. Mech., Vol. 46, 1979, 329-336.

[9] Barker, L. M., "A Model for Stress Wave Propagation in Composite Materials," J. Composite Materials, Vol. 5, April 1971, 140-162.

[10] Christensen, R. M., "Wave Propagation in Layered Elastic Media," J. Appl. Mech., Vol. 42, 1975, 153-158.

[11] Glauz, R. D. and Lee, E. H., "Transient Wave Analysis in a Linear Time-Dependent Material," J. Appl. Phys., Vol. 25, Aug. 1954, 947-953.

[12] Ince, E. L., "Ordinary Differential Equations," Dover Pub., 1956, 381-384.

[13] Bellman, R., Kalaba, R. E. and Lockett, Jo Ann, "Numerical Inversion of the Laplace Transform: Applications to Biology, Economics, Engineering and Physics," American Elsevier Pub., 1966, 22-47.

[14] Mukunoki, I. and Ting, T. C. T., "Transient Wave Propagation Normal to the Layering of a Finite Layered Medium," Int. J. Solids Structures, Vol. 16, 1980, 239-251.
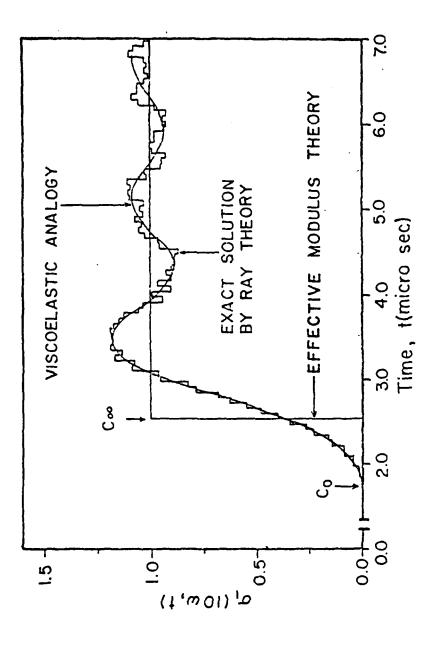
[15]   Ting, T. C. T., "Higher Order Asymptotic Solution for Wave Propagation in Elastic Layered Composites," to appear in Proc. 3rd Int. Conf. on Composite Materials.

[16]   Ting, T. C. T., "The Effects of Dispersion and Dissipation on Wave Propagation in Viscoelastic Layered Composites," to appear in Int. J. Solids Structures.

Fig. 1 Geometry of the layered Medium

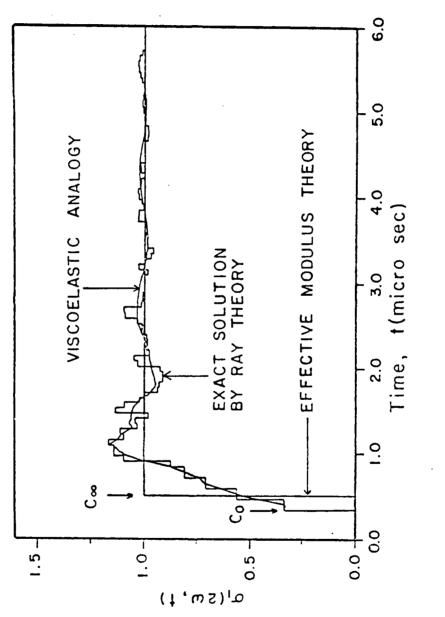Fig. 2 "Equivalent" Viscoelastic Relaxation Function G(t)

Fig. 3 Stress at the center of the 11th layer of the semi-infinite layered medium due to a unit step stress applied at the center of the first layer.
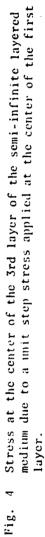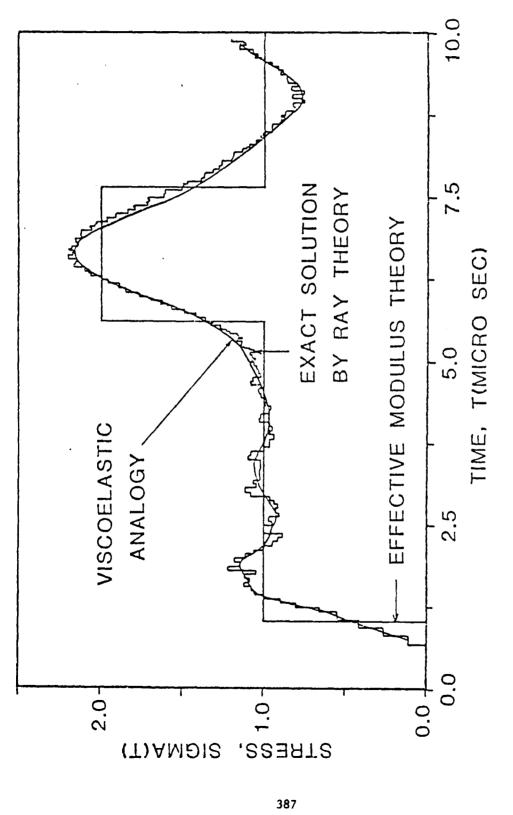
Fig. 4 Stress at the center of the 3rd layer of the semi-infinite layered medium due to a unit step stress applied at the center of the first layer.

Fig. 5   Stress at the center of the 5th layer due to a unit step stress applied at the center of the first layer while the center of the 14th layer is fixed.
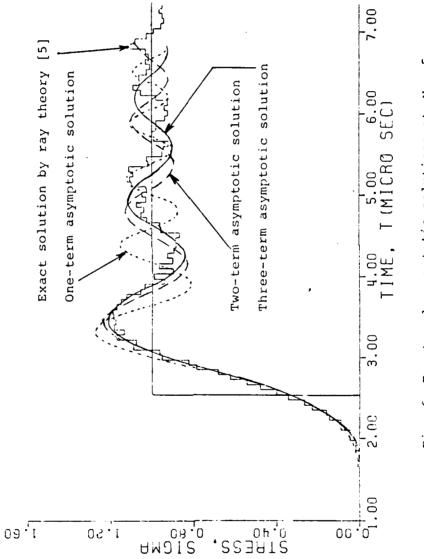
Fig. 6 Exact and asymptotic solutions at N = 5

THEORY OF ACOUSTIC EMISSION

Yih-Hsing Pao
Department of Theoretical and Applied Mechanics
Cornell University
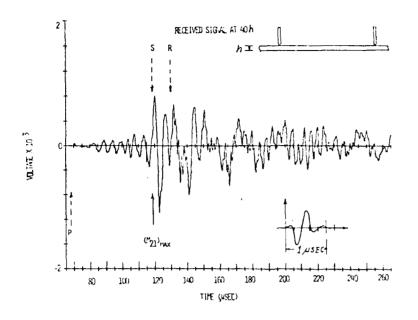Ithaca, New York 14853

ABSTRACT

Acoustic emission is the sudden release of mechanical energy which propagates in the form of elastic waves from a localized region in a material. The technology of locating and characterizing the source of emission for the purpose of detecting failure of the material is also called acoustic emission. The underlying theory of mechanical radiation and dispersion of elastic waves pertaining to the technology of acoustic emission is discussed in this report.

INTRODUCTION

The term of acoustic emission is currently used to describe a physical phenomenon as well as a technology. As a phenomenon, it describes the propagation of transient elastic waves which are radiated from localized regions in a material or structure due to rapid release of strain energy in these regions. By recording and analyzing the transient waves, it is possible to locate the sources of radiation and, in some cases, even to characterize the nature of the sources. The technology that has been developed over the past decade to locate and to characterize the sources is also called acoustic emission [1].

When a material is plastically deformed, micro-cracks and voids are developed. The dynamic processes that generate elastic waves are very complex and a general theory within the framework of materials science is still lacking. For the purpose of detecting the zones of microplastic deformation, the sources of emission may be represented by nuclei of strains of the dynamic theory of elasticity [2]. The emission of waves by these macroscopic sources and the propagation of radiated signals in a wave guide can then be analyzed and compared with experimental observations. This constitutes the solution of the "direct problems" of acoustic emission.

To accomplish the objective of acoustic emission, it requires a solution for the "inverse problem", that is, to determine the locations and characteristics of the assumed macroscopic sources from the signals recorded at various stations in the wave guide. We are, however, still far from accomplishing this objective, both in theory and in practice. The difficulties that one encounters can be illustrated by the following example.

Shown in Fig. 1 is the time record of a simulated acoustic emission in a glass plate of thickness $h = 0.9525$ cm. The signal is generated by a wide-band piezo-electric transducer placed vertically at the surface of the plate and the time record of the input signal is shown at the lower right corner of the figure. A similar transducer is placed at a distance $40 h$ from the source, and it records the output signal as shown in the figure.

Theoretically, the source of emission can be represented by a vertical force acting at the surface of the plate. Even for such an idealized system, however, we are still not able to calculate completely the output signals from the given input. The reason is that the surface response of the plate is expressed mathematically in terms of a double infinite integral which is difficult to evaluate.

On the other hand, when one tries to compare the theoretical results, if obtained, with the experimental data, he will encounter the difficulty of not knowing what the piezoelectric transducer does measure. A displacement, velocity, or stress? For the same reason, the exact nature of the input by the transducer is not certain. Although progress has been made recently in the calibration of transducers [3], precise measurements of the input and output signals of acoustic emission have not been widely reported.

In this report, we present briefly the theoretical basis of acoustic emission based on the dynamic theory of elasticity. We discuss first a mathematical representation of the sources of emission and then the dispersion of the emitted signals in a plate wave guide. Finally, we show how the results of the analysis of dispersion can be applied to locate the source of emission.

1. MACROSCOPIC SOURCE FUNCTIONS

The displacement field, $u(x,t)$, of elastic waves at the spatial coordinate x and time t is governed by the Navier-Cauchy equation [2,4]

$$(\lambda+\mu)\nabla\nabla\cdot u + \mu\nabla^2 u - \rho\ddot{u} = -\rho f(x,t) \qquad (1)$$

where ρ is the mass density and λ, μ are the Lamé constants of the medium. The body force f per unit mass is a source that generates the wave. The corresponding stress field, $\sigma(x,t)$, a second rank tensor, is given by

$$\sigma = \lambda I \nabla\cdot u + \mu(\nabla u + u\nabla) \qquad (2)$$

where I is the idemfactor (isotropic tensor).

When three mutually perpendicular body forces acting at the point x_0, the displacement fields are described by a dyadic function $G(x-x_0,t)$ which satisfies the following dyadic equation,

$$(\lambda+\mu)\nabla\nabla\cdot G + \mu\nabla^2 G - \rho\ddot{G} = -\rho I \delta(x-x_0) f(t) \qquad (3)$$

where $f(t)$ is an arbitrary scalar function in time. The solution for G was found by G.G. Stokes in 1849 [2] and is called Green's dyadics for elastic waves in an infinite medium.

From the Green's dyadic, one can derive the displacement field generated by other types of point sources, known as nuclei of strains in the theory of elasticity. Let a, b and c be three mutually perpendicular unit vectors and $c = a \times b$, and $u(x,a)$ be the displacement field due to a concentrated force in the direction of a at x_0. The displacements due to various nuclei of strains of unit strength are then derivable from G:

Single force along a	$u(x,a) = a\cdot G(x-x_0)$
Double force along a	$-a\cdot\nabla u(x,a)$
Center of dilatation	Sum of three double forces
Single couple about c	$b\cdot\nabla u(x,a)$
Center of rotation about c	$b\cdot\nabla u(x,a) + a\cdot\nabla u(x,-b)$
Double couple without moment	$b\cdot\nabla u(x,a) + a\cdot\nabla u(x,b)$

The double force and double couple without moment have been used to represent the opening of a tensile crack and the sliding of shear crack respectively. The center of dilatation can be used to represent the creation or collapse of a void. The wave fields generated by various point sources are depicted, not in scale, in Fig. 2, where the solid lines indicate the P-wave front (Pressure, Longitudinal wave), and dashed lines the S-wave front (Shear, Transversal wave) the arrows indicating the direction of displacement.
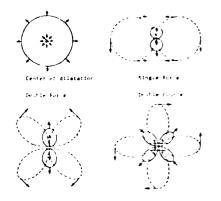
Fig. 2. Dynamic Nuclei
of Strains.

By applying the representation theorem of dynamic elasticity, waves generated by a crack is expressed in terms of an integration of these nuclei of strains over the surface of discontinuity. Let the stress field corresponding to the displacement dyadic G be Σ, a third rank tensor, and, as in Eq. (2),

$$\Sigma = \lambda I \nabla \cdot G + \mu (\nabla G + G \nabla)$$ (4)

The representation theorem states that [5,p.39]

$$u(x,t) = \int_0^t dt_o \iint_S \{u(x_o,t_o) \cdot n_o \cdot \Sigma(x-x_o,t-t_o)$$

$$- T(x_o,t_o) \cdot G(x-x_o,t-t_o)\} dS_o$$ (5)

where $T = n \cdot \sigma$ is the traction across the surface. So with outer normal n_o. Let S surround a surface of discontinuity A, and assume that the traction is continuous, or vanishes, across A. The above integral then reduces to

$$u(x,t) = \int_0^t dt_o \int_A [\cdot u(x_o,t_o)] \cdot n_o \cdot \Sigma(x-x_o,t-t_o) dS_o$$ (6)

where $[u]$ denotes the jump of displacement u across the surface A.

The above integral can further be reduced to [5,p.52]

$$u_j(x,t) = \int_0^t dt_o \iint_A m_{pq} \partial_q G_{jp}(x-x_o,t-t_o) dS_o$$ (7)

where

$$m_{pq} = \lambda n_k [' u_k'] \delta_{pq} + \mu \{ n_p ['u_q'] + n_q ['u_p'] \} \tag{8}$$

and

$$\partial_q \equiv \partial / \partial x_q \quad .$$

The m_{pq} is called the moment density tensor and it is related to the strength and orientation of a crack.

2. DISPERSION IN A WAVE GUIDE

Once a wave is generated by a source in a bounded medium such as a plate or a cylindrical shell, the wave is then multiply reflected by the bounding surfaces which form a wave guide. As a result of interference of the reflected waves, the disturbance observed at some distance from the source is quite different from that originated from the source. This is known as the *geometric dispersion of waves* in a wave-guide.

We note that the Green's functions and general solutions discussed in the previous section are for a source in an infinite elastic solid. To analyze waves in a bounded medium, new Green's functions, one for each type of wave guide, must be found. This amounts to the solving of a new boundary value problem of elastodynamics. So far, only wave guides with simple geometry like a plate, a circular cylinder, and a sphere have been analyzed in detail [4].

Consider an axially symmetric point source in a plate which is bounded by two parallel surfaces $z = \pm h/2$. The normal displacement of a propagating pulse in the plate is given by [4,p.468]

$$u(x,t) = \int_{-\infty}^{\infty} \overline{f}(\omega) U(x,\omega) e^{-i\omega t} d\omega$$

$$U(r,z,\omega) = \int_{-\infty}^{\infty} S(\xi,x_o) \frac{N(\xi,z,\omega)}{D(\xi,\omega)} J_o(\xi r) d\xi \tag{9}$$

In these integrals, ω is the circular frequency and ξ is the radial wave number, J_o is the Bessel function of zeroth order. The source function at x_o is separated into a temporal part and spatial part. The Fourier transform of the temporal part is $\overline{f}(\omega)$, and the spatial part is represented by the function S in terms of the radial wave number. The N and D are complicated transcendental functions of ξz and ξh. When the source is a crack, additional integration over the surface of crack is required and the $S(\xi)$ function should be replaced by the moment density tensor of Eq. (8). Other parts of the integrand should also be modified to reflect the angular dependence and tensorial property.

The zeros of the denominator, or the roots of the equation

$$D(\xi,\omega) = \frac{\tan\beta h/2}{\tan\alpha h/2} - [-\frac{-4\alpha\beta\xi^2}{(\xi^2-\beta^2)^2}]^{\pm 1} = 0 \tag{10}$$

$$\alpha^2 = \omega^2/c_p^2 - \xi^2 \quad , \quad \beta^2 = \omega^2/c_s^2 - \xi^2$$

is the characteristic equations of free waves in an elastic plate. It is known as the Rayleigh-Lamb equation, and has been a subject of extensive study [6]. By finding the roots of this equation, one can determine the phase and group velocities and modal shapes of various modes of waves propagating in the plate.

Despite its simple appearance, evaluation of the double integral of Eq. (9) is not easy. Two methods of evaluation are available. One is the method of normal mode, the other is the method of generalized ray. In the method of normal mode, the integral in ξ for U is evaluated by the calculus of residues, the poles being the roots of the Rayleigh-Lamb equation, and each term of the residual series is the contribution from one normal mode. The integral in ω must then be evaluated numerically. Details of the analysis are discussed in Ref. 4.

In the method of generalized ray, the integrand of U in Eq. (9) is expanded into a series, each term of the series represents the wave propagating along a generalized ray path. The double integral, one in ω and one in ξ, of each term of the series is then evaluated by applying the Cagniard method [7]. Numerical results for various types of point sources listed in the previous section are shown in a new article soon to appear [8].

From these preliminary calculations, we can conclude that the method of normal mode is effective for waves at far field over a long time $(r > 10\ h)$ and the method of generalized ray is for near field and short duration $(r < 10\ h)$. Without going into details of these results, we can say that the wave pattern at the surface of a plate is very sensitive to the location of the receiver, relative to the source, and to the temporal function of the assumed point source in a plate.

3. LOCATING AND CHARACTERIZING SOURCES

The analysis as described in the previous section provides information that can be used to locate the source of emission in a plate. In a dispersive wave guide, a wave packet with strong magnitude propagates with the group velocity (c_g) which is, in general, different from the speed of P-wave (c_p), S-wave (c_s) or Rayleigh surface wave (c_R). Let the strong wave packet be radiated from the source at (x_o,y_o) in a two dimensional space and at the time t_o. If the same wave packet is detected by three or more transducers at stations $(x_n,y_n), n = 1,2,3,$ the distance traveled by the wave from the unknown source to each station is then given by

$$r_n = [(x_n-x_o)^2 + (y_n-y_o)]^{1/2} \qquad n = 1,2,3, \tag{11}$$

Since the travel time for the same wave packet which propagates with speed c_g is $t_n - t_o$, we have the following three equations for three unknowns, x_o, y_o, and t_o,

$$r_n = (t_n - t_o)c_g \qquad (n = 1,2,3) \tag{12}$$

By solving this system of equations, we can determine the location of the source. This is known as the method of triangulation.

It should be noted that to apply the method of triangulation, we need to know the c_g of the dominant signal that is detectable by all three receivers. For waves in a plate, Eq. (10) shows that the maximum group velocity of the lowest antisymmetric mode (the flexural mode) is equal to $c_g = 1.01 c_s$. The arrival time for this group of waves in a glass plate is marked in Fig. 1 by an arrow and $(M_{o1})_{max}$ under a very strong signal. Additional analysis of Eq. (9) revealed that the magnitude of this group of waves was indeed the strongest of the early arrived groups at far field. Hence by identifying the arrival times of this group at various receiving stations, we should be able to locate the source by solving three equations of (12).

For wave field near the source, the wave packet has not yet been fully developed. The solutions of generalized ray are then used to determine the travelling speed of a predominanat signal.

In the current practice of non-destructive testing of materials, once a source of "weakness" is located by the acoustic emission, other means of testing are used to inspect closely the source region. Eventually, one hopes not only to locate the source but also to find out what type of source emission it is. Mathematically, this is equivalent to the determination of both $f(\omega)$ and $C(\xi)$ in Eq. (9) when $u(x,t)$ is given. Some progress has been made to solve this inverse problem when either $f(\omega)$ or $C(\xi)$ is known (see Vol. 2 of Ref. 2). The progress, however, has not yet reached the stage of application to acoustic emission.

References

[1] J.C. Spanner, Acoustic Emission, Techniques and Applications, Index Publications, Evanston, Il. 1974.

[2] A.E.H. Love, The Mathematical Theory of Elasticity, Dover Publications, New York, 1944.

[3] W. Sachse and N.N. Hsu, "Ultrasonic Transducers for Materials Testing and Their Characterization" in Physical Acoustics, v. 14, edited by W.P. Mason and R.N. Thurston, Academic Press, New York, 1979.

[4] J. Miklowitz, The Theory of Elastic Waves and Wave Guides, North-Holland Publishing Co., New York, 1978.

[1] K. Aki and P.G. Richards, <u>Quantitative Seismology—Theory and Methods</u>, Vol. ... and 2, W.H. Freeman and Co., San Francisco, 197_.

[2] Y.H. Pao and R.R. Kaul, "Waves and Vibrations in ... Isotropic and Anisotropic Plates", in <u>R.D. Mindlin and Applied Mechanics</u>, edited by G. Herrmann, Pergamon Press, New York, 197_.

[3] Y.H. Pao and R.K. Chichester?, "The Generalized Ray Theory and ... Propagation of Layered Elastic Solids", in <u>Physical Acoustics</u>, v. 1_, edited by W.P. Mason and R.N. Thurston, Academic Press, New York, 197_.

[4] A.N. Ceranoglu and Y.H. Pao, "Propagation of Elastic Pulses and Acoustic Emission in a Plate, Part I, II, III", to appear in Journal of Applied Mechanics, 19__.

26th Conference of Army Mathematicians

10-12 June 1980

Achenbach, Professor Jan, Department of Civil Engineering, Northwestern
    University, Evanston, IL 60201.

Chandra, Dr. Jagdish, Director, Mathematics Division, U. S. Army Research
    Office, Box 12211, Research Triangle Park, NC 27709.

Chen, Dr. Peter C.T., Research Mathematician, Benet Weapons Laboratory, LCWSL,
    ARRADCOM,Watervliet Arsenal, Watervliet, NY 12189.

Dafermos, Professor C. M., Divsion of Applied Mathematics, Brown University,
    Box F, Providence, RI 02912.

Davis, Julian L., ATTN: SCA-T, Bldg. 3410, U. S. Army Armament R&D Command,
    Dover, NJ 07801.

Devereaux, COL Alfred B. Jr., Commander and Director, U. S. Army Cold Regions
    Research and Engineering Laboratory, Box 282, Hanover, NH 03755.

Drew, Professor Donald A., Mathematical Sciences, Rensselaer Polytechnic
    Institute, Troy, NY 12181.

Elder, Alexander S., Mechanical Engineer, USARRADCOM, Ballistic Research
    Laboratory, Aberdeen Proving Ground, MD 21005.

Flaherty, Professor Joseph E., Applied Math & Mechanics Section, Research Branch,
    Benet Weapons Laboratory, LCWSL, ARRADCOM, Watervliet Arsenal, Watervliet,
    NY 12189.

Fleishman, Professor B., Department of Math Sciences, Rensselaer Polytechnic
    Institute, Troy, NY 12181.

Freitag, Dr. D.R., Technical Director, U. S. Army Cold Regions Research and
    Engineering Laboratory, Box 282, Hanover, NH 03755.

Haug, Professor Edward J., Jr., Materials Division, College of Engineering,
    University of Iowa, Iowa City, IA.

Kapila, A. L., Assistant Professor, Math Research Center, University of Wisconsin,
    610 Walnut Street, Madison, WI 53706.

Lin, Professor S. S., Mathematics Research Center, University of Wisconsin,
    610 Walnut Street, Madison, WI 53706.

Ludford, Professor Geoffrey S. S., Applied Mathematics, Theoretical & Applied
    Mechanics, Thurston Hall, Cornell University, Ithaca, NY 14850.

Malek-Madani, Professor R., MRC-University of Wisconsin, 610 Walnut Street, Madison, WI 53706.

Masaitis, Ceslovas, Mathematician, USA Ballistic Research Laboratory/ARRADCOM, ATTN: DRDAR-BLB, Aberdeen Proving Ground, MD 21005.

Meyer, Dr. R.E., Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, WI 53706.

Nakano, Dr. Yoshisuke, U. S. Army Cold Regions Research and Engineering Laboratory, Box 282, Hanover, NH 03755.

Noble, Professor Ben, Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, WI 53706.

Nohel, John A., Director and Professor of Mathematics, Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, WI 53706.

O'Hara, G. Peter, Mechanical Engineer, USA-ARRADCOM, Watervliet Arsenal, Watervliet, NY 12189.

Pao, Professor Yih-Hsing, Chairman, Department of Theoretical & Applied Mechanics, Cornell University, Thurston Hall, Ithaca, NY 14853.

Polk, John, Terminal Ballistics Division, Ballistics Research Center, Aberdeen Proving Ground, MD 21005.

Poore, Aubrey B., Associate Professor, Mathematics Research Center and Colorado State University, Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, WI 53706.

Powell, John D., Research Physicist, Ballistic Research Laboratory, ATTN: DRDAR-BLB, Aberdeen Proving Ground, MD 21005.

Robinson, Richard, Army Concepts Analysis Agency, 8120 Woodmont Avenue, Bathesda, MD 20014.

Ross, Edward W., Jr., Staff Mathematician, U. S. Army Natick R&D Command, Kansas Street, Natick, MA 01776.

Saibel, Dr. Edward, Chief, Solid Mechanics Branch, U. S. Army Research Office, Box 12211, Research Triangle Park, NC 27709.

Srivastav, Professor Ram P., Applied Mathematics & Statistics, Suny at Stony Brook, Stony Brook Campus, Stony Brook, NY 11794.

Sterrett, Dr. K. F., Chief, Research Division, U. S. Army Cold Regions Research and Engineering Laboratory, Box 282, Hanover, NH 03755.

Takagi, Dr. Shunsuke, U. S. Army Cold Regions Research and Engineering Laboratory, Box 282, Hanover, NH 03755.

Tallingston, Arnold, ARADCOM U. S. Army Armament R&D Command, Picatinny
    Arsenal, Bldg. 3310, Dover, NJ 07801.

Tasi, Professor James, State University of New York, Department of Mechanical
    Engineering, Stony Brook, NY 11794.

Thompson, Dr. James L., Act C, Survival Technology Function, U. S. Army Tank
    Automotive R&D Command, ATTN: DRDTA-ZSS, Warren, MI 48090.

Ting, Professor, T. C. T., University of Illinois at Chicago Circle,
    Department of Materials Engineering, Chicago, IL 60680.

Tracey, Dr. Dennis M., Mechanical Engineer, Army Materials and Mechanics
    Research Center, Arsenal Street, Watertown, MA 02172.

Vasilakis, John D., Mechanical Engineer, USA ARRADCOM, Benet Weapons Laboratory,
    Bldg. 115, Watervliet Arsenal, Watervliet, NY 12189.

Weeks, Dr. Wilford, U. S. Army Cold Regions Research and Engineering Laboratory,
    Box 282, Hanover, NH 03755.

Wu, Dr. Julian, Benet Weapons Laboratory, Watervliet Arsenal, Watervliet, NY
    12189.

<center>SUPPLEMENTAL REGISTRATION LIST</center>   <div align="right">10 June 1980</div>

Alexander, Assistant Professor Roger K., Rensselaer Polytechnic Institute,
    Department of Mathematical Sciences, Troy, NY 12181.

Atkinson, Col. John C., United States Army Reserve, St. Louis, MO.

Chow, Professor Pao L., Wayne State University, Department of Mathematics,
    Detriot, MI 48202.

Coleman, Dr. Norman P., ATTN: DRDAR-SCF-CC, U.S. Army Armament R&D Command,
    Dover, NJ 07801.

Norman, Dr. Paul D., Burroughs Corporation, 33 Williams Way, Caln Township,
    PA 19335.

Veraden, Assistant Professor Vasundara V., The Ohio State University, 155
    W. Woodruff Avenue, Columbus, OH 43210.

<center>ADDITIONS</center>   <div align="right">12 June 1980</div>

Lenoe, Dr. Edward, U. S. Army Mechanics and Materiels Division, Watertown
    MA 02172.

Ludford, Proffessor Geoffrey S. S., Thurston Hall, Cornell University, Ithaca,
    NY 14850.

Stewart, D. Scott, Thurston Hall, Cornell University, Ithaca, NY 14850.

<center>399</center>

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
ARO Report No. 81-1	AD AC93502	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
TRANSACTIONS OF THE TWENTY-SIXTH CONFERENCE OF ARMY MATHEMATICIANS	
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Army Mathematics Steering Committee on Behalf of the Chief of Research, Development and Acquisition	January 1981
	13. NUMBER OF PAGES
	399

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
US Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. The findings in this report are not to be construed as official Department of the Army position unless so designated by other authorized documents.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This is a technical report resulting from the Twenty-Sixth Conference of Army Mathematicians. It contains most of the papers in the agenda of this meeting. These treat various Army applied mathematical problems.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

crack problems	floating elastic plates
solitary waves	stress-strain problems
flame propagation	screw threads
combustion	fracture mechanics
stochastic equations	bifurcation
intermittent motion	distributed systems
control systems	wave propagation
maneuvering targets	gas dynamics
volterra integral equations	finite element method
cubic splines	layered media
waves	