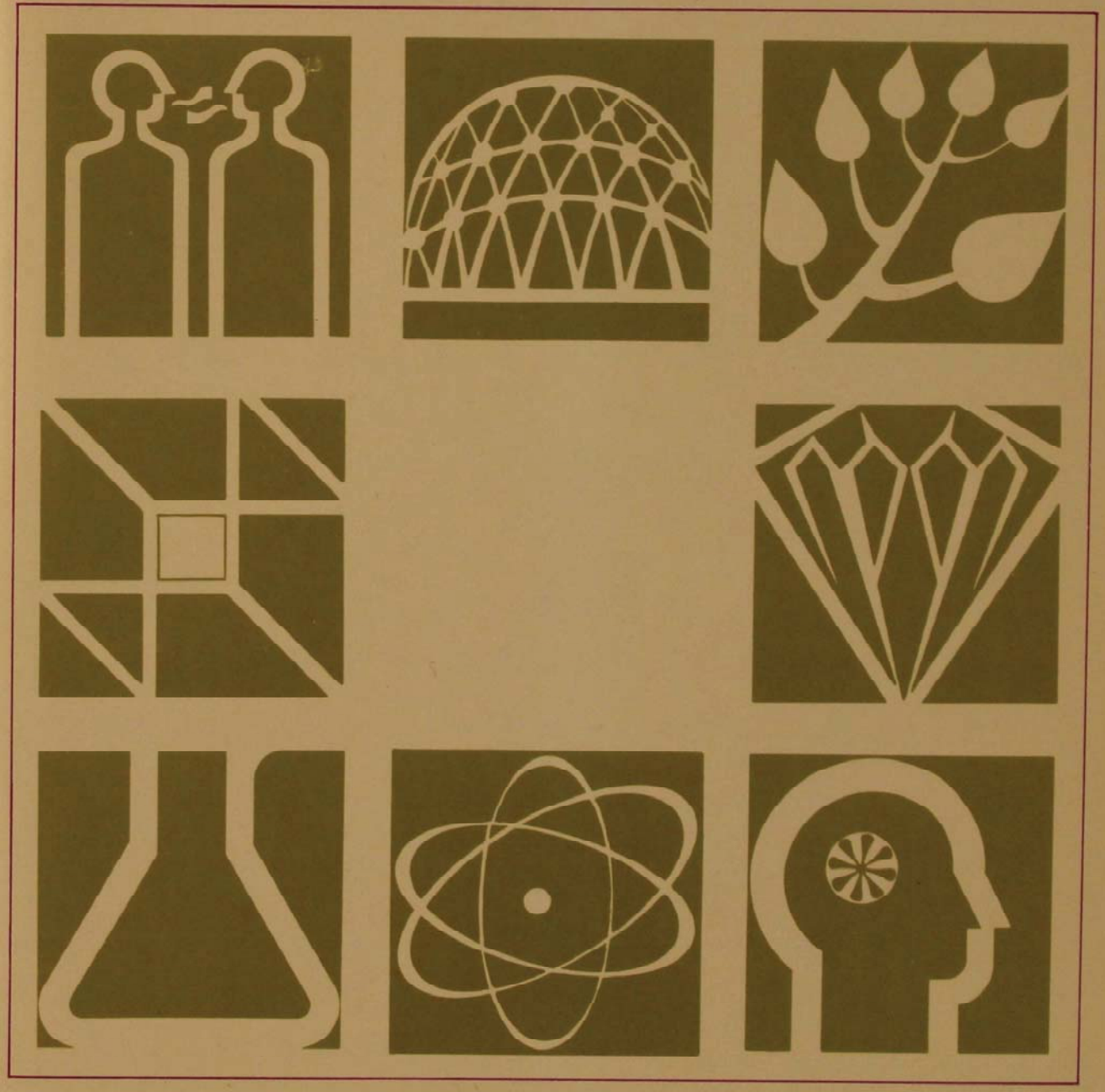
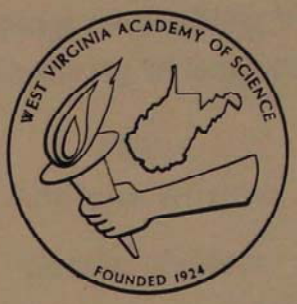
Volume 45, Number 4

Proceedings of the West Virginia Academy of Science 1973





Proceedings of the West Virginia Academy of Science 1973

Vol. 45—No. 4

THE FORTY-EIGHTH ANNUAL SESSION

Fairmont State College

Fairmont, West Virginia

APRIL 5, 6, 7, 1973

Printed by
McClain Printing Company
Parsons, West Virginia

Cover Design
Meredith Pearce
West Virginia University
Office of Publications
Morgantown

West Virginia Academy of Science, Inc.

The West Virginia Academy of Science was founded at Morgantown, November 28, 1924. The Academy was incorporated under the Code of West Virginia on May 9, 1959 as "a nonstock corporation, which is not organized for profit but for the advancement of learning and scientific knowledge."

Proceedings of the West Virginia Academy of Science

Editor
Anthony Winston

Section Editors

Virgil G. Lilly
Biology

Robert E. Adams
Biology

William G. Martin
Biology

W. Newman Bradshaw
Biology

John Grunninger
Chemistry

Harold V. Fairbanks
Engineering Sciences

John T. Sears
Engineering Sciences

Milton T. Heald
Geology

John J. Renton
Geology

Henry W. Gould
Mathematics

Arnold J. Levine
Physics

Robert L. Decker
Psychology and Education

Richard S. Little
Social Science

Elizabeth A. Bartholomew
Newsletter

Acknowledgment. The West Virginia Academy of Science gratefully acknowledges the financial support of the following for the publication of the 1973 proceedings.

Claud Worthington Benedum Foundation
West Virginia University Foundation—Biology Fund.
The West Virginia University Chapter of the Society of the Sigma Xi.

Contents

MATHEMATICS SECTION

H. W. Gould, *Improved Evaluation of the Finite Hypergeometric Series F(−n, ½;j+1;4)* 317

David Ryan and George Trapp, *Partitioning Methods for Accelerating Gauss-Seidel Interations* 323

Jin Bai Kim, *A Generalized Homomorphism of a Group* 326

W. C. Sisarcick, *Linear Functionals on the Space of Entire Functions* 328

William Simons and George Trapp, *Excludable and Includable Points* 332

David C. Rine, *A Correspondence Between Control Logic of Associative Memories and Markov Algorithms* 335

W. Thurmon Whitley, *A Note on f-Ideals in Rings of Continuous Functions* 345

David Cusick, *Torsion Submodule and Injective Hull* 347

Cynthia Deitz and Stanley Wearden, *An Evaluation of Normalizing Transformations for Skewed Data* 351

Michael Mays, John Schleusner, and William Simons, *Open Sets From Disjoint Closed Intervals* 358

James Miller, *On the Maximum Modulus for Meromorphic Univalent Functions* 360

ENGINEERING SECTION

M. R. Pamidi and S. H. Advani, *Constitutive Properties of Brain Tissue* 366

Harold V. Fairbanks and John F. Dakan, *Some Observations on the Effect of Ultrasonics on the Solidification of Nodular Cast Iron* 375

Ronald L. Huston and Jeffrey C. Huston, *A New Experimental Method of Stress Measurement in the Interior of an Elastic Body* 382

William Squire, E. E. Barth, H. F. Chou, and L. S. Fan, *Solution of an Eigenvalue Problem by Several Weighted Residual Methods* 389

Alvin M. Strauss, *Cancer-Lymphocyte Interaction Analysis as a Problem in Differential Game Theory* 392

Robert O. Weedfall and George M. Farrell, *Agricultural Weather in West Virginia: The Unusual Growing Season of 1972* 400

R. L. Huston and A. M. Strauss, *International Conflict—A Basis for Prediction and Systems Analysis—* 405

Pradip Khaladkar, John Sears, and Harold Fairbanks, *Influence of Ultrasonics on the Flow of Polymer Melt* 412

Ronald L. Gorrell and R. A. Bajura, *Loss Coefficients in Closely spaced Sharp-Edged Area Contractions* 421

A. H. Mace, Jr., and R. A. Bajura, *Flow in a One Dimensional Channel with an Oscillating Boundary* 429

PSYCHOLOGY AND EDUCATION SECTION

Don E. McLaughlin and Herbert P. Kagen, *The Effects of Sulfur Dioxide on Learning and Activity in the Rat* 439

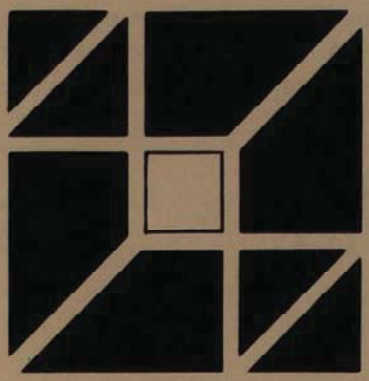
Robert E. Anderson, *Environmental Education: A Broad Spectrum Approach With Emphasis on Role-Playing and Local, Student-Directed Projects* 445

Robert B. Cameron, *On the Nature of the Basic Binocular Disparity Cue in Stereoscopic Vision* 452

SOCIAL SCIENCE SECTION

John R. Warner, Jr., Nancy M. Wolfe, Albin R. Gilbert, and Dana G. Cable, *Testing Women's Attitudes Toward Abortion by Latency-Weighted Responses* 459

Chester E. Zimolzak, *Changing Locational Trends in Coal Production And Their Relationship to Settlement Patterns* 463



Mathematics

Improved Evaluation of the Finite Hypergeometric Series $F(-n, \frac{1}{2}; j+1; 4)$

H. W. Gould
Department of Mathematics
West Virginia University, Morgantown, West Virginia 26506

Abstract

Recurrence relations, depending on the parity of n, are given for the binomial sum

$$F(-n, \tfrac{1}{2}; j+1; 4) = \sum_{k=0}^{n} (-1)^k \binom{n}{k}\binom{2k}{k}\binom{j+k}{k}^{-1}$$

from which the sum may be calculated for any integer j. A short table for $-1 \leqslant j \leqslant 6$ is given. Various other results appear, and the bracket function is used to simplify the expressions. For example, it is shown that

$$F(-n, \tfrac{1}{2}; [\tfrac{n}{2}]; 4) = 3 - \left(8 + \frac{4}{n-1}\right)\frac{1-(-1)^n}{2},$$

for $n \geqslant 2$. A closed form for the general case is not found, but some information about the general sum is found. The sum is found to be a quotient of polynomials, the denominator polynomials having a very simple form. The results of this paper appear without proof in the author's book *Combinatorial Identities* (revised edition, published by the author, Morgantown, West Virginia 1972).

Introduction and Summary

An interesting binomial sum

$$(1.1)\qquad F(-n, \tfrac{1}{2}; j+1; 4z) = \sum_{k=0}^{n} (-1)^k \binom{n}{k}\binom{2k}{k}\binom{j+k}{k}^{-1} z^k$$

arises from the definition of the hypergeometric series. The series does not seem to have been evaluated in closed form in general. A few special cases have appeared in the literature. Because of the well-known formula of Gauss for the special case F(a, b; c; 1), we know that

$$F(-n, \tfrac{1}{2}; j+1; 1) = 2^{-2n} \binom{2n+2j}{n+j} \binom{2j}{j}^{-1} .$$

Sandham [4] posed as an easy problem to show that

$$F(-2n, \tfrac{1}{2}; n+1; 4) = 1, \tag{1.2}$$

and Spiegel [5] proved this using Euler's integral. The simplicity of Sandham's formula suggested to me that progress might be made with (1.1) in the case z = 1. The results given below represent a condensation of work I first did in 1954. The results were tabulated in my book [3] without proof. As a contribution to the study of combinatorial identities the proofs may be of interest, though the general pattern of the coefficients has still not been determined.

It is convenient to consider (1.1) according to the parity of n, and we accordingly define

$$S_j^n = F(-2n, \tfrac{1}{2}; n+j+1; 4) \tag{1.3}$$

and

$$R_j^n = F(-2n-1, \tfrac{1}{2}; n+j+1; 4). \tag{1.4}$$

We shall obtain the recurrence relations

$$2\,\frac{2n+2j+1}{n+j+1}\,S_{j+1}^n = 3S_j^n + R_j^n \tag{1.5}$$

and

$$2\,\frac{2n+2j+1}{n+j+1}\,R_{j+1}^n = 3R_j^n + S_{j-1}^{n+1} , \tag{1.6}$$

which, together with some initial values, are sufficient to generate the following table of special values. The formulas are valid for $n \geqslant 0$ except as indicated.

TABLE OF VALUES

j	S_j^n	R_j^n
-1	$3 \quad (n \geq 1)$	$-\frac{5n+2}{n} \quad (n \geq 1)$
0	1	-1
1	$\frac{n+1}{2n+1}$	0
2	$\frac{3(n+1)(n+2)}{2(2n+1)(2n+3)}$	$\frac{n+2}{2(2n+3)}$
3	$\frac{(n+2)(n+3)(11n+10)}{4(2n+1)(2n+3)(2n+5)}$	$\frac{5(n+2)(n+3)}{4(2n+3)(2n+5)}$
4	$\frac{(n+2)(n+3)(n+4)(43n+35)}{8(2n+1)(2n+3)(2n+5)(2n+7)}$	$\frac{21(n+2)(n+3)(n+4)}{8(2n+3)(2n+5)(2n+7)}$
5	$\frac{9(n+2)(n+3)(n+4)(n+5)(19n+14)}{16(2n+1)(2n+3)(2n+5)(2n+7)(2n+9)}$	$\frac{(n+3)(n+4)(n+5)(85n+168)}{16(2n+3)(2n+5)(2n+7)(2n+9)}$
6	$\frac{(n+3)(n+4)(n+5)(n+6)(683n^2+1825n+924)}{32(2n+1)(2n+3)(2n+5)(2n+7)(2n+9)(2n+11)}$	$\frac{11(n+3)(n+4)(n+5)(n+6)(31n+60)}{32(2n+3)(2n+5)(2n+7)(2n+9)(2n+11)}$

For the general case we find that

$$S_j^n = 2^{1-j} g_j(n) \prod_{i=1}^{j} (2n + 2i - 1)^{-1}, \quad j \geq 1, \tag{1.7}$$

and

$$R_j^n = 2^{1-j} f_j(n) \prod_{i=2}^{j} (2n + 2i - 1)^{-1}, \quad j \geq 2. \tag{1.8}$$

As a matter of fact these follow by repeated use of (1.5) and (1.6). Here $g_j(n)$ and $f_j(n)$ are some polynomials in n whose precise form we have not determined, however g is of degree j and f is of degree j - 1.

Finally, we obtain the finite series

$$F(-n, \tfrac{1}{2}; j + 1; 4z) \tag{1.9}$$

$$= 2^{2j} \binom{2j}{j}^{-1} \frac{1}{t} \sum_{k=0}^{t-1} \left(\cos \frac{2\pi k}{t}\right)^{2j} \left\{1 - 4z \left(\sin \frac{2\pi k}{t}\right)^2\right\}^n$$

which is valid for integers $t > 2n + 2j$, and is of the type considered by Good [2], Carlitz [1] and others, for Legendre polynomials and other classical polynomials.

Proofs of the Formulas

We apply the well-known integral formula

$$(2.1) \qquad \binom{2k}{k}\binom{j+k}{k}^{-1} = 2^{2j+2k+1}\binom{2j}{j}^{-1}\frac{1}{\pi}\int_0^{\pi/2} \sin^{2k}x \cos^{2j}x\, dx$$

to (1.1) and this yields

$$(2.2) \qquad F(-n, \tfrac{1}{2}; j+1; 4z) = 2^{2j+1}\binom{2j}{j}^{-1}\frac{1}{\pi}\int_0^{\pi/2} \cos^{2j}x(1 - 4z\sin^2 x)^n\, dx.$$

This integral is symmetrical in sin and cos because of the symmetry of (2.1) in k and j. It is possible to get (2.2) from Euler's integral. Since $\sin^2 x = 1 - \cos^2 x$, clearly (2.2) gives our binomial sum in terms of an integral of a polynomial in cos x. In general, however, it is easy to prove that for any polynomial f,

$$(2.3) \qquad \frac{2}{\pi}\int_0^{\pi/2} f(\cos x)\, dx = \frac{1}{2\pi}\int_0^{2\pi} f(\cos x)\, dx = \int_0^1 f(\cos 2\pi u)\, du,$$

and for f(z) a polynomial of degree n in z,

$$(2.4) \qquad \frac{1}{t}\sum_{k=0}^{t-1} f(\cos\frac{2\pi k}{t}) = \int_0^1 f(\cos 2\pi u)\, du, \quad \text{for all } t > n.$$

Relation (2.4) is the remarkable formula that shows that certain Riemann approximating sums actually equal their corresponding integrals when a sufficient number of terms are taken in the sum.

Relations (2.3) and (2.4) yield from (2.2)

$$F(-n, \tfrac{1}{2}; j+1; 4z) = 2^{2j}\binom{2j}{j}^{-1}\int_0^1 (\cos 2\pi x)^{2j}(1 - 4z\sin^2 2\pi x)^n\, dx$$

and prove (1.9).

Because of the trigonometric identity $\cos x\,(1 - 4\sin^2 x) = \cos 3x$ we obtain from (2.2) and (1.3)-(1.4)

$$(2.5)\qquad S_j^n = 2^{2n+2j+1}\binom{2n+2j}{n+j}^{-1}\frac{1}{\pi}\int_0^{\pi/2}\cos^{2j}x\cos^{2n}3x\,dx,$$

and

$$(2.6)\qquad R_j^n = 2^{2n+2j+1}\binom{2n+2j}{n+j}^{-1}\frac{1}{\pi}\int_0^{\pi/2}\cos^{2j-1}x\cos^{2n+1}3x\,dx.$$

Combining these we obtain

$$S_j^n - R_j^n = 2^{2n+2j+1}\binom{2n+2j}{n+j}^{-1}\frac{1}{\pi}\int_0^{\pi/2}\cos^{2j-1}x\cos^{2n}3x(\cos x - \cos 3x)\,dx$$

which simplifies (using $\cos x - \cos 3x = 4\cos x - 4\cos^3 x$) to

$$S_j^n - R_j^n = 4S_j^n - 2\frac{2n+2j+1}{n+j+1}S_{j+1}^n,$$

which is precisely (1.5).

To obtain (1.6) we must only simplify the difference

$$2\frac{2n+2j+1}{n+j+1}R_{j+1}^n - 3R_j^n$$

in a similar way.

From (2.5) we have Sandham's formula $S_0^n = 1$ at once. To show that $R_1^n = 0$ it is sufficient to note that

$$\int_0^{\pi/2}\cos x\cos^{2n+1}3x\,dx = 0$$

since the integrand can be written as a sum of cosines of multiples of $2x$.

The value of R_2^n now follows from (1.6) using S_0^n and R_1^n. The value of R_0^n is found as follows. When $j = 0$ the integrand in (2.6) may be written (using identities) in the form $(-1 + 2\cos 2x)\cos^{2n}3x$, so that

$$R_0^n = -2^{2n+1}\binom{2n}{n}^{-1}\frac{1}{\pi}\int_0^{\pi/2}\cos^{2n}3x\,dx + 2^{2n+2}\binom{2n}{n}^{-1}\frac{1}{\pi}\int_0^{\pi/2}\cos^{2n}3x\cos 2x\,dx$$

$$= -1$$

since the second integral is identically zero (since its integrand is a sum of cosines of multiples of $2x$).

The value of S_1^n now follows from (1.5) using S_0^n and R_0^n. The remainder of the table of values is now easily found using (1.5) and (1.6).

While it is true that recurrence relations (1.5)-(1.6) could be combined into a single relation in various ways, it is convenient for calculation to have them in the form given here.

If we next compare (1.7) with (2.5) and (1.8) with (2.6) we find

$$(2.7)\qquad f_j(n) = j!2^{2n+2j}\binom{n+j}{n}\binom{2n}{n}^{-1}\frac{1}{(2n+1)\pi}\int_0^{\pi/2}\cos^{2j-1}x\cos^{2n+1}3x\,dx$$

and

$$(2.8)\qquad g_j(n) = j!2^{2n+2j}\binom{n+j}{n}\binom{2n}{n}^{-1}\frac{1}{\pi}\int_0^{\pi/2}\cos^{2j}x\cos^{2n}3x\,dx.$$

These relations reveal in a way why it is relatively easy to express the original series in terms of n but not so easy in terms of j. For we may convert an integral of the form

$$\int_0^{\pi/2}\cos^{p}x\cos^{q}3x\,dx$$

into a series of terms by using $\cos 3x = \cos x(1 - 4\sin^2 x)$, and then use the binomial theorem and integrate term by term. This gives a series the number of whose terms depends on q. But it is more difficult to replace cos x by something in terms of cos 3x. This would give a series the number of whose terms depends on p. Expansions of this sort would allow us to sum other similar series.

Now it is possible to use the Euler transformation

$$F(a, b; c; z) = (1 - z)^{-a}F\left(a, c - b; c; \frac{z}{z - 1}\right)$$

to convert our series into another form, and the result is

$$(2.9)\qquad F\left(-n, \tfrac{1}{2}; j + 1; 4z\right) = \binom{2j}{j}^{-1}\sum_{k=0}^{n}\binom{n}{k}\binom{2j + 2k}{j + k}2^{-2k}(4z)^k(1 - 4z)^{n-k}.$$

This series might shed other light on the nature of the sum.

Condensation of Results

It is interesting to note that the values tabulated above for the several special cases may be expressed in a condensed form by use of the bracket function. Recalling that [x] denotes the greatest integer $\leqslant x$ it is not difficult to see that some of the results listed above may be summarized in the following formulas:

$$(3.1)\qquad F\left(-n, \tfrac{1}{2}; \left[\tfrac{n}{2}\right]; 4\right) = 3 - \left(8 + \frac{4}{n - 1}\right)\frac{1 - (-1)^n}{2}, \quad n \geqslant 2,$$

which combines S_{-1}^n and R_{-1}^n;

$$(3.2)\qquad F\left(-n, \tfrac{1}{2}; \left[\frac{n + 1}{2}\right]; 4\right) = 2(-1)^n + 1, \quad n \geqslant 1,$$

which combines S_{-1}^n and R_0^n;

$$(3.3) \qquad F\left(-n, \frac{1}{2}; \left[\frac{n+2}{2}\right]; 4\right) = (-1)^n, \quad n \geqslant 0,$$

which combines S_0^n and R_0^n;

$$(3.4) \qquad F\left(-n, \frac{1}{2}; \left[\frac{n+3}{2}\right]; 4\right) = \frac{(-1)^n + 1}{2}, \quad n \geqslant 0,$$

which combines S_0^n and R_1^n;

$$(3.5) \qquad F\left(-n, \frac{1}{2}; \left[\frac{n+4}{2}\right]; 4\right) = \frac{n+2}{2(n+1)} \frac{(-1)^n + 1}{2}, \quad n \geqslant 0,$$

which combines S_1^n and R_1^n.

Other more complicated combinations can be given, but we have not determined the general pattern.

I wish to take this occasion to thank my friend and colleague Dr. A. M. Chak for useful discussions during a revision of this paper.

Literature Cited

1. Carlitz, L. 1957. A formula for the product of two Hermite polynomials. *J. London Math. Soc.* 32:94-97.
2. Good, I. J. 1955. A new finite series for Legendre polynomials. *Proc. Cambridge Phil. Soc.* 51:385-388.
3. Gould, H. W. 1972. Combinatorial Identities, A Standardized Set of Tables Listing 500 Binomial Coefficient Summations. Revised edition, published by the author, Morgantown, W. Va.
4. Sandham, H. F. 1951. Problem 4438. *Amer. Math. Monthly* 58:343.
5. Spiegel, M. R. 1952. Solution of Problem 4438. *Amer. Math. Monthly* 59:641-642.

Partitioning Methods for Accelerating Gauss-Seidel Interations

David Ryan and George Trapp
Department of Statistics and Computer Science
West Virginia University, Morgantown, West Virginia 26506

Abstract

The rebalancing method of accelerating the convergence of the Gauss-Seidel iteration scheme is discussed. This method is shown to be related to a least squares procedure. A numerical example is given which shows that this method can reduce the number of iterations to convergence to below the number required using optimal successive over-relaxation.

When approximating solutions to partial differential equations, one frequently is required to solve a large system of linear equations. In the case of elliptic equations, these large linear systems are often solved using iterative techniques, see reference [6] or [7] for a detailed discussion.

In this note we describe a method for accelerating the rate of convergence of such iterative procedures. For simplicity we will only consider the problem of accelerating the Gauss-Seidel iteration scheme. However, the approach is clearly applicable to other schemes as well.

For completeness we will summarize the Gauss-Seidel method and its standard acceleration procedure—over-relaxation. Let $Ax = y$, where A is an n x n matrix, and x and y are vectors. To solve for x, a sequence of vectors $x^0, x^1, x^2, \ldots$ is formed, where x^0 is a guess and x^{i+1} is given by $x^{i+1} = Gx^i + b$, where $G = -(L+D)^{-1}U$ and $b = -(L+D)^{-1}y$. Here we have partitioned A as $A = L + D + U$, where D is the diagonal of A, L is the lower triangular part of A, and U is the upper triangular part of A.

Various conditions are known which guarantee the convergence of this process, see [6], for example, if A is symmetric positive definite. We will assume that we have a matrix A so that the Gauss-Seidel method converges.

This procedure may be written component wise as follows:

$$x_j^{i+1} = \frac{1}{a_{jj}} \left(b_j - \sum_{k=1}^{j-1} a_{jk} x_k^{i+1} - \sum_{k=j+1}^{n} a_{jk} x_k^{i} \right),$$

for $j = 1, \ldots, n$. By linearly extrapolating the Gauss-Seidel scheme, one obtains what is termed relaxation. Again component wise, it may be written as:

$$x_j^{i+1} = (1-w)x_j^i + wz_j^i \text{, where } z_j^i \text{ is the standard}$$

Gauss-Seidel computed value given above. It is well known [6] that for any w in (0,2), the relaxation scheme converges. Note $w = 1$ corresponds to Gauss-Seidel. Moreover, there exists a unique w^* which gives the greatest rate of convergence. Various methods are known for estimating w^*, see reference [6]. Our concern, however, is with a alternate method of accelerating convergence.

Wachspress [7] first introduced the concept of rebalancing. His basic idea was to partition the components of the approximation solution, then seek a normalizing coefficient for each partition so that when each component of a partition is multiplied by the corresponding coefficient, the new vector of normalized components is a better approximation. This idea has been extended and studied further by Nakamura [3], [4]. In the next section, we reformulate the coarse mesh acceleration procedure in terms of least squares theory. This approach allows one to estimate errors. Prior to this work, the only mathematical approach to this procedure was given by Froehlich [2]. His work is primarily concerned with showing that the reduced set of equations has the same fundamental properties as the original set.

In the final section, we present a numerical example which illustrates the utility of this method and reinforces the results reported in [1], [4].

Partitioning

Given an approximation x to the true solution, we partition x as $x = (x_1, x_2, \ldots, x_k)$ where each x_i is a subset of the components of x. Let P_i denote the projection into the subspace $(0, \ldots, x_i, \ldots, 0)$. Then clearly $P_iP_j = 0$ for $i \neq$

j, and the sum of the P_i's is the identity. Normally, k will be much smaller than n.

We seek a vector z of the form $z = \sum_{i=1}^{k} c_i P_i x$ which is a better approximation than x.

The problem then is to determine the c_i. Replacing x in Ax = y by z, we obtain the equation

$$(1) \qquad \sum_{i=1}^{k} c_i u_i = y$$

where $u_i = AP_i x$. This is a system of n equations in the k unknowns $c_1, \ldots c_k$.

Many options are now available for finding the c_i. A standard least square solution will guarantee that $\| Az - y\| \leqslant \| Ax - y\|$ because $x = \Sigma P_i x$ and $c_1 = \ldots = c_k = 1$ is an allowable set of coefficients (here $\|a\|^2 = \Sigma a_i^2$).. Another approach is a modified least squares solution. Letting $e_j = P_j e$ where $e = (1,1,1, \ldots ,1)$ and taking the inner product of equation (1) with each e_j, we obtain

$$(2) \qquad \sum_{i=1}^{k} c_i(u_i,e_j) = (y,e_j) \text{ , } j = 1, \ldots ,k.$$

This system is solved for the c_i and then z is computed. No norm estimates are known to determine if this z is better than x. However empirical studies indicate that a better approximation is obtained.

The basic advantage of the modified scheme over the standard least squares procedure is that equations (2) may be formed using far less computing time than is necessary to form the standard least squares system.

Numerical Example

We now discuss the results of the modifies method for one particular problem. Let A be the tridiagonal matrix with $a_{i,i+1} = a_{i,i-1} = -1$ and $a_{ii} = 2$ except $a_{11} = 3$ and $a_{nn} = 1$. This matrix is known to be ill conditioned. Let $y_i = 2(-1)^i$ for $i = 2, \ldots ,n-1$ and $y_1 = y_n = 1$. The components of the true solution to Ax = y are $x_{2i} = 2$ and $x_{2i+1} = 1$. We will only report our results for n = 100. We started with an initial guess of $x_i = 3$, and continued each problem until the norm of the approximation minus the true solution was less than .01.

The case w = 1 (Gauss-Seidel) took over 5000 iterations. We determined the true w for this problem to be w = 1.93. Using this true w from the beginning, the number of iterations was reduced to 720. This number however is not representative of a real problem because one normally uses a sequence of estimates to the true w as the scheme progresses and hence 720 is the least possible iteration number using over-relaxation.

When applying the partitioning acceleration scheme, we computed seven over-relaxation iterates using w = 1.8, then we formed equations (2) with k = 8 and used the resulting z as an initial guess for seven more w = 1.8 iterations, and continued this process. The number of iterations needed for convergence was then reduced to 540. Our w = 1.8 was somewhat arbitrary, we only wanted a value less than the true w. Other test results indicate that w = 1.85 would decrease the iteration count. Choosing seven iterations as the time between applying the partitioning procedure was also arbitrary. A study of optimizing the

time between accelerating the iterates would seem quite useful. We partitioned the components of x, into eight subvectors with x_1 being the first 12 components of x, x_2 the next 13 components, etc. Again no attempt was made to optimize the selection of the number of components or the size or distribution of the components.

It seems clear to us that this partitioning procedure should be investigated further, because even when applied in an non-optimal manner it improves the convergence rate of the widely used optimal over-relaxation scheme.

Literature Cited

1. Esposito, V., and S. Nakamura. 1973. Course mesh rebalancing applied to hydraulic analysis. Presented at May meeting of Amer. Nuclear Soc., Ann Arbor, Mich.
2. Froehlich, R. 1967. A theoretical foundation for coarse mesh variational techniques. Gulf Atomics Technical Report No. GA-7870.
3. Nakamura, S. 1971. Course mesh acceleration of iterative solution of neutron diffusion equation. *Nuclear Sci. and Engineering* 43:116-120.
4. Nakamura, S. 1970. A variational rebalancing method for linear iterative convergence schemes of neutron diffusion and transport equations. *Nuclear Sci. and Engineering* 39:278-283.
5. Nakamura, S., and T. Ohnishi. 1971. A study of iterative solutions for the finite element method. Trans. Amer. Nuclear Soc. 14, No. 2:662.
6. Varga, R. S. 1962. Matrix Iterative Analysis. Prentice-Hall. Englewood Cliffs, N. J.
7. Wachspress, E. L. 1966. Iterative Solution of Elliptic Systems and Applications to the Neutron Diffusion Equations of Reactor Physics. Prentice-Hall, Englewood Cliffs, N. J.

A Generalized Homomorphism of a Group

Jin Bai Kim

Department of Mathematics

West Virginia University, Morgantown, West Virginia 26506

Yong M. Lee

Mathematics Department

Trenton State College, Trenton, New Jersey 08625

Abstract

A mapping φ from a group G into a semigroup S is called a partial homomorphism of G into S if there is a subgroup G_o of G such that $\varphi(gx)=\varphi(g)\varphi(x)$ for all g in G_o and for all x in G. In this note we study partial homomorphisms of G into S. A necessary condition for $\varphi(G)$ to be a regular semigroup is obtained (in Theorem 3).

The concept of a homomorphism of groups plays a fundamentally important role in modern algebra. There is some inherent appeal in generalizing the concept to general algebraic systems. For example: By a partial homomorphism φ of a partial groupoid S into a partial groupoid S′ such that if a and b are elements of S such that ab is defined in S, then the product $(a\varphi)(b\varphi)$ is defined in S′ and equal to $(ab)\varphi$[1, p. 93].

This definition is a very weak (or a greatly generalized) version of the homo-

morphism of a group into a group, regarding the domain and the range of the mapping. We define a partial homomorphism which is stronger than the partial homomorphism of groupoids but is weaker than the homomorphism of groups.

Definition

Let G be a group and e denote the identity of G. Let S be a semigroup. A mapping π of G into S is called a partial homomorphism of G if there exists a subgroup G_o of G such that $\pi(gx)=\pi(g)\pi(x)$ for all g in G_o and for all x in G. A one to one partial homomorphism is called a partial isomorphism. We shall say that π is a partial homomorphism of G into S with an associated subgroup G_o.

We study a partial homomorphism of a group into a semigroup which is defined as above and give three theorems which indicate how the work relates to groups and semigroups.

THEOREM 1. Let π be a partial homomorphism of a group G into a semigroup S with an associated subgroup G_o. If π' is also a partial homomorphism of G into S with $\pi(G)=\pi'(G)$ and $\pi(e)\neq\pi'(e)$. Then there exists a subgroup S_1 of S such that S_1 and $\pi(G_o)$ are isomorphic and disjoint.

Proof

It is clear that $\pi(G_o)$ is a subgroup of S. Let π' be a partial homomorphism of G into S with $\pi(G)=\pi'(G)$ and $\pi(e)\neq\pi'(e)$, where e is the identity of G. We show that $\pi(e)\pi'(e)=\pi'(e)$. To do this, let $\pi'(e)=\pi(b)$ and $\pi(e)=\pi'(a)$, for some a, b in G. We can see that $\pi(e)\pi'(e)=\pi(e)\ \pi(b)=\pi(eb)=\pi(b)=\pi'(e)$. Similarly we can prove that $\pi'(c)\pi(c)=\pi(c)$. Now we show that $\pi(G_o)\pi'(c)=\{\pi(g)\pi'(c): g\in G_o\}=S_1$ is a group. Let $x\in S_1$. Then x takes the form $x=\pi(g)\pi'(e)$ for some g in G_o. We can see that $x\pi(g^{-1})\pi'(e)=\pi(g)\ \pi'(e)\pi(g^{-1})\ \pi'(e)=\pi(g)\ \pi'(e)\ \pi(eg^{-1})\pi'(e)=\pi(g)\pi'(e)\ \pi(e)\ \pi(g^{-1})\ \pi'(e)=\pi(g)\ \pi(e)\ \pi(g^{-1})\ \pi'(c)=\pi(c)\ \pi'(c)$, $\pi(c)\pi'(c)=\pi(c)\ \pi'(c)\pi(c)\ \pi'(e)$, $\pi(g^{-1})\ \pi'(e)x=\pi(e)\ \pi'(e)$, and hence S_1 is a group. Define a mapping f: $\pi(G_o)\to\pi(G_o)\pi'(e)=S_1$ by $f(x)=x\pi'(e)$ for $x\in\pi(G_o)$. Letting $g_1, g_2\in G_o$, we can see that $f(\pi(g_1)\pi(g_2))=f\ (\pi(g_1g_2))=\pi(g_1g_2)\ \pi'(e)=\pi(g_1eg_2)\ \pi'(e)=\pi(g_1)\pi(e)\ \pi(g_2)\ \pi'(e)=\pi(g_1)\ \pi'(e)\pi(e)\ \pi(g_2)\ \pi'(e)=\pi(g_1)\ \pi'(e)\pi(g_2)\pi'(e)=f\ (\pi\ (g_1))f(\pi(g_2))$. Now assume that $f(\pi(g_1))=f(\pi(g_2))$. Then by post multiplication by $\pi(e)$, we obtain $\pi(g_1)=\pi(g_2)$. This proves that S_1 and $\pi(G_o)$ are isomorphic. It is easy to see that $\pi(G_o)\cap S_1=\phi$, the empty set.

THEOREM 2. Let π_1 be a partial homomorphism of a group G into a semigroup S with an associated subgroup G_o. The following statements are equivalent.

(a) $\pi_1(G)$ contains n distinct idempotents $\{e_i: i=1,2,..,n\}$ and $e_ie_j=e_j$.

(b) There exist n distinct partial homomorphisms $\{\pi_i: i=1,2,..,n\}$ with $\pi_i(e)\neq\pi_j(e)$ for $i\neq j$ and $\pi_1(G)=\pi_i(G)$ $(i=1,2,..,n)$.

Proof

(a) implies (b). Let $\pi_1(g_i)=e_i$ be an idempotent for some g_i in G. Define π_i by $\pi_i(x)=\pi_1(xg_i)$ $(x\in G)$. Letting $g\in G_o$, observe that $\pi_i(gx)=\pi_1(gxg_i)=\pi_1\ (g)\pi_1\ (xg_i)=\pi_1(g)\ \pi_i(x)=\pi_1(ge)\ \pi_i\ (x)=\pi_1(g)e_ie_1\ \pi_i\ (x)=\pi_1(g)\ \pi_1(g_i)e_1\pi_i\ (x)=\pi_1(gg_i)e_1\pi_i(x)\ =\pi_i(g)\pi_1(e)\pi_1(xg_i)\ =\pi_i(g)\ \pi_1(xg_i)=\pi_i(g)\ \pi_i(x)$ and $\pi_i(e)=\pi_1(eg_i)=\pi_1(g_i)=e_i$; which shows that π_i is a partial homomorphism of G into S and G_o is its associated subgroup.

(b) implies (a). We assume the conditions of (b). Clearly $\pi_1(G)$ contains n idempotents e_i $(i=1,2,..,n)$. For each $\pi_i(e)=e_i$, there is g such that $\pi_i(g)=e_j$.

Observe that $e_ie_j=\pi_i(e)\pi_i(g)=\pi_i(eg)=\pi_i(g)=e_j$.

THEOREM 3. Let π_1 be a partial homomorphism of G into a semigroup S with an associated subgroup G_o. Let $[G:G_o]=m$. If there exist m-1 partial homomorphisms $\{\pi_i:i=1,2,..,m\}$ of G into S such that $\pi_1(G)=\pi_i(G)$ and $\{\pi_i(e):i=1,2,..,m\}$ are m distinct idempotents in S. Then $\pi_1(G)$ is a regular semigroup.

Proof

It is not difficult to show that $\pi_1(G)=\bigcup_{i=1}^{m}\pi_1(G_o)\pi_i(e)$. We shall show that $\pi_1(G)$ is a semigroup. Let $h,k \in \pi_1(G)$ and let $\pi_1(g_1)\pi_i(e)$ and $k=\pi_1(g_2)\pi_j(e)$. Then $hk=\pi_1(g_1)\pi_1(g_2)\pi_j(e)=\pi_1(g_1g_2)\pi_j(e) \in \pi_1(G)$. Hence $\pi_1(g)$ is a semigroup. We see that $h(\pi_1(g_1^{-1})\pi_j(e)h=\pi_1(g_1)\pi_1(g_1^{-1})\pi_1(g_1)\pi_i(e)=h$, and hence $\pi_1(G)$ is regular. This proves the theorem.

The author is indebted to the referee for remarks leading to a smoother presentation in this paper.

Literature Cited

1. Clifford, A. H., and G. B. Preston. 1961. The algebraic theory of semigroups, Math. Surveys of the Amer. Math. Society, No. 7, Vol. 1, Providence, R.I.
2. Zassenhaus, Hans J. 1949. The theory of groups, Chelsea Publishing Co., New York.

Linear Functionals on the Space of Entire Functions

W. C. Sisarcick
Department of Mathematics
Marshall University, Huntington, West Virginia 25701

Abstract

A metric d is defined on the linear space Γ of entire functions of one variable. The classes of scalar homomorphisms and metrically bounded linear functionals on (Γ,d) are characterized. Natural definitions on the class of metrically bounded linear functionals on (Γ,d) give rise to a Banach space isometrically isomorphic to the Banach space $\{\mathbb{C} \times \mathbb{C}, +, \cdot, \|\cdot\|\}$ over $\mathbb{C}$, where + and $\cdot$ are the usual definitions of addition and scalar multiplication, and $\|(a,b)\| = |a| + |b|$.

Introduction

In 1948, V. G. Iyer [1] considered the linear space Γ of entire functions of one variable, together with the metric d defined by

$d(f,g)=\sup\{|a_0 - b_0|, |a_n - b_n|^{1/n}, n \geq 1\}$ where

$f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma$ and $g(z) = \sum_{n=0}^{\infty} b_n z^n \in \Gamma$. He proves the following.

Theorem A. The space (Γ,d) is a complete, separable linear metric space.

Theorem B. Every continuous linear functional F on (Γ,d) is of the form

$$F(f) = \sum_{n=0}^{\infty} b_n a_n$$

where $f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma$ and $\{b_n\}_{n=0}^{\infty}$ is a fixed sequence of complex numbers for which $\{|b_n|^{1/n}\}_{n=0}^{\infty}$ is bounded, and for every such $\{b_n\}_{n=0}^{\infty}$ the corresponding F, defined as above, is a continuous linear functional on (Γ,d).

Following Wilansky [3], a scalar homomorphism F on (Γ,d) is a linear functional on (Γ,d) such that $F(f \cdot g) = F(f) \cdot F(g)$ for all $f, g \in \Gamma$. Here $f \cdot g$ is the entire function defined by $(f \cdot g)(z) = f(z) \cdot g(z)$.

Theorem 1. Let F be a function from (Γ,d) into $\mathbb{C}$, $F \neq 0$. Then F is a continuous scalar homomorphism if and only if there exists a unique $b \in \mathbb{C}$ such that for all

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma,$$

$$F(f) = f(b).$$

Following Iyer [2], a linear functional F on (Γ,d) is said to be metrically bounded if and only if there exists a non-negative number M such that for all $f \in \Gamma$, $|F(f)| \leqslant M\, d(f,f_0)$. Here f_0 is the element in Γ such that $f_0(z) = 0$ for all $z \in \mathbb{C}$.

It is well known that linear functionals on normed linear spaces are continuous if and only if they are bounded. Iyer [1] proves that no norm can be defined on the linear space Γ that produces a topology equivalent to the topology generated by d. Because of this we have Theorem 2.

Theorem 2. Let F be a linear functional on (Γ,d). If F is metrically bounded, F is continuous. Also, there exist continuous linear functionals on (Γ,d) that are not metrically bounded.

Let B denote the class of metrically bounded linear functionals on (Γ,d). Then $\{B, +, \cdot,\}$ is a linear space over $\mathbb{C}$, where + and $\cdot$ are defined as usual. For $F \in B$, define $\|F\| = \inf\{M \geqslant 0 \mid \text{for all } f \in \Gamma,\ |F(f)| \leqslant M\, d(f,f_0)\}$. It is clear that $\{B, +, \cdot, \|\cdot\|\}$ is a normed linear space.

Theorem 3. Let F be a function from Γ into $\mathbb{C}$. Then $F \in B$ if and only if there exists a unique $(a,b) \in \mathbb{C} \times \mathbb{C}$ such that for all $f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma$,

$$F(f) = a_0 a + a_1 b.$$

Also, $\|F\| = |a| + |b|$.

Proofs

Proof of Theorem 1. Let F be a function from Γ into $\mathbb{C}$, $F \neq 0$. If F is a continuous scalar homomorphism, by Theorem B there is a unique sequence $\{b_n\}_{n=0}^{\infty}$ of complex numbers such that for all $f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma$, $F(f) = \sum_{n=0}^{\infty} a_n b_n$. Note that $b_0 = 1$. For if $g_n(z) = z^n$, $n \geqslant 0$, each $g_n \in \Gamma$, $b_0 = F(g_0) = F(g_0 \cdot g_0) = F(g_0) \cdot F(g_0) = b_0^2$. If $b_0 = 0$, then for all $n \geqslant 1$, $b_n = F(g_n) = F(g_0 \cdot g_n) =$

$F(g_0) \cdot F(g_n) = 0$ and $F \equiv 0$, a contradiction. If $b_1 = 0$, then for any $n \geq 1$, $b_n = F(g_n) = (F(g_1))^n = b_1^n = 0$ and $F(f) = a_0 = f(b_1)$. If $b_1 \neq 0$, then for any $n \geq 0$, $b_n = F(g_n) = (F(g_1))^n = b_1^n$ and $F(f) = \sum_{n=0}^{\infty} a_n b_1^n = f(b_1)$.

Conversely, given $b \in \mathbb{C}$, define $F:\Gamma \to \mathbb{C}$ by $F(f) = f(b)$, $f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma$. It is clear that F is a $\neq 0$ scalar homomorphism on (Γ,d). Let $b_0 = 1$ and $b_n = b^n$, $n \geq 1$. Then $\{|b_n|^{1/n}\}_{n=1}^{\infty}$ is bounded, $F(f) = f(b) = a_0 + \sum_{n=1}^{\infty} a_n b^n = \sum_{n=0}^{\infty} a_n b_n$. Hence, by Theorem B, F is continuous.

Proof of Theorem 2. Let F be a metrically bounded linear functional on (Γ,d), $f \in \Gamma$ and $\epsilon > 0$ be given. Let $M \geq 0$ be such that for all $g \in \Gamma$, $|F(g)| \leq M\, d(g,f_0)$. Choose $\delta = \epsilon/M+1$. Then if $g \in \Gamma$ and $d(f,g) < \delta$, $|F(f) - F(g)| = |F(f-g)| \leq M\, d(f-g,f_0) = M\, d(f,g) \leq (M+1)d(f,g) < (M+1)\delta = \epsilon$. Hence, F is continuous at f and, therefore, continuous on (Γ,d).

For an example of a continuous linear functional on (Γ,d) that is not metrically bounded, let $b_n = n$, $n = 0,1,2,3,\cdots$. Then $\{|b_n|^{1/n}\}_{n=1}^{\infty}$ is a bounded sequence, and by Theorem B, $F: \Gamma \to \mathbb{C}$ defined by $F(\sum_{n=0}^{\infty} a_n z^n) = \sum_{n=0}^{\infty} n a_n$, $f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma$, is a continuous linear functional. If F is metrically bounded, there exists an $M \geq 0$ such that $|\sum_{n=0}^{\infty} n a_n| \leq M \sup\{|a_0|, |a_n|^{1/n}, n \geq 1\}$ for all $\{a_n\}_{n=0}^{\infty}$ such that $\lim_{n\to\infty} |a_n|^{1/n} = 0$. Let k be a positive integer, $k > M + 1$. Choose $a_n = 0$ of $n \neq k$, $a_k = k$. Then $\lim_{n\to\infty} |a_n|^{1/n} = 0$, but $|\sum_{n=0}^{\infty} n a_n| = k^2$ and $M \sup\{|a_0|, |a_n|^{1/n}, n \geq 1\} = M\, k^{1/k} < k \cdot k^{1/k} < k \cdot k = k^2$. Hence, F is not bounded.

Proof of Theorem 3. Let $F \in B$. By Theorem B and Theorem 2, there exists a unique sequence of complex numbers $\{b_n\}_{n=0}^{\infty}$ such that $F(\sum_{n=0}^{\infty} a_n z^n) = \sum_{n=0}^{\infty} a_n b_n$ and $|\sum_{n=0}^{\infty} a_n b_n| \leq \|F\| \sup\{|a_0|, |a_n|^{1/n}, n \geq 1\}$ for all $\sum_{n=0}^{\infty} a_n z^n \in \Gamma$. Assume $b_k \neq 0$ for some $k \geq 2$. Choose $a_n = 0$ if $n \neq k$ and $a_k \neq 0$ such that $\|F\| < |a_k|^{1-1/k} \cdot |b_k|$. Then $\sum_{n=0}^{\infty} a_n z^n \in \Gamma$ and $|\sum_{n=0}^{\infty} a_n b_n| = |a_k b_k| > \|F\| \cdot |a_k|^{1/k} = \|F\| \sup\{|a_0|, |a_n|^{1/n}, n \geq 1\}$, a contradiction. Hence, $F(\sum_{n=0}^{\infty} a_n z^n) = a_0 b_0 + a_1 b_1$. For $f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma$, we now have $|F(f)| = |a_0 b_0 + a_1 b_1| \leq |a_0||b_0| +$

$|a_1||b_1| \leqslant (|b_0| + |b_1|)d(f,f_0)$. Therefore, $\|F\| \leqslant |b_0| + |b_1|$. To show equality here, it sufficies to show there exists a $g_0 \in \Gamma$ such that $|F(g_0)| = (|b_0| + |b_1|)d(g_0,f_0)$. For $j = 1,2$, let $a_j = \bar{b_j} \cdot |b_j|^{-1}$ if $b_j \neq 0$, $a_j = 0$ if $b_j = 0$. Choose $g_0(z) = a_0 + a_1 z \in \Gamma$. Then $|F(g_0)| = |a_0 b_0 + a_1 b_1| = |b_0| + |b_1| = (|b_0| + |b_1|)d(g_0,f_0)$.

Conversely, given $(a,b) \in \mathbb{C} \times \mathbb{C}$, define $F: \Gamma \to \mathbb{C}$ by $F(f) = a_0 a + a_1 b$, $f(z) = \sum_{n=0}^{\infty} a_n z^n \in \Gamma$. By Theorem B, F is a continuous linear functional. As in the preceeding, $F \in B$ with $\|F\| = |a| + |b|$.

Remarks

Defining addition and scalar multiplication on $\mathbb{C} \times \mathbb{C}$ as usual and defing a norm by $\|(a,b)\| = |a| + |b|$, it is clear that $\{\mathbb{C} \times \mathbb{C}, +, \cdot, \|\cdot\|\}$ is Banach space. From Theorem 3, it is clear that this space is isometrically isomorphic to the Banach space of metrically bounded linear functionals on (Γ,d).

Analogous results for entire functions of n variables are derived in a similar manner. Here, the Banach space of metrically bounded linear functionals will have dimension $n + 1$.

The author acknowledges useful suggestions of the referee.

Literature Cited

1. Iyer, V. G. 1948. On the Space of Integral Functions–I. *Indian Math Soc. J.* 12:13-30.
2. Iyer, V. G. 1956. On the Space of Integral Functions–IV. *Proc. Amer. Math. Soc.* 7:644-49.
3. Wilansky, A. 1964. Functional Analysis. Blaisdell Publishing Company, New York. 291 pp.

Excludable and Includable Points

William Simons
Department of Mathematics
George Trapp
Department of Statistics and Computer Science
West Virginia University, Morgantown, West Virginia 26506

Abstract

Excludable and includable points are defined, and a study of these points is made for a convex collection of N points in the plane. Also, a criterion is presented for determining whether any given point is excludable or includable, and a method for finding an excluding (including) circle is given.

In previous work [2], we considered a collection of $M + n + 1$ points in R^n and found a condition which ensured an affirmative answer to this separability question: Given k, $0 \leqslant k \leqslant M$, is there a sphere through n preselected points and one other which contains k points in its interior and M - k in its exterior?

This question has in an informal sense a dual, namely: Can k preselected points be separated from the remaining points by a sphere? Kirchberger's Theorem [1], [4], which deals with separation by hyperplanes, can be generalized to provide an answer to this dual question [4, page 177, problem 6.4], [3]. However, if k = 1 Kirchberger's Theorem is not needed because the dual question is then trivial—*unless* the separating spherical surface is additionally required to contain n + 1 of the remaining points; whereupon, the problem becomes quite involved, and Kirchberger's Theorem is of no help. This paper is concerned with this final problem, in the plane, of characterizing which of those points can be separated from the others by a circle. Stated more precisely, let X be a collection of N points, $N \geqslant 4$, in *general circle-free position* in the plane; that is, no three points lie on a line and no four points lie on a circle. Then we say the point $x \in X$ is an *excludable point* if there is a circle through three points of X which contains x in its exterior and all the remaining points in its interior. Similarly, $x \in X$ is an *includable point* if there is a circle through three points of X which contains x in its interior and all the remaining points in its exterior. Let

$$E = \{x \in X \mid x \text{ is excludable but not includable}\},$$
$$I = \{x \in X \mid x \text{ is includable but not excludable}\},$$

and

$$D = \{x \in X \mid x \text{ is both excludable and includable}\}.$$

In addition, we will use the notation conv(S) for the convex hull of the set S; C(x,y,z) is the circle through the points x, y, and z; C*(x,y,z) is the circle C(x,y,z) together with its interior; C(x,y) is the circle through the points x and y having the line segment connecting x and y as a diameter; and C*(x,y) is C(x,y) together with its interior.

The set X is convex if $\text{conv}(X - \{x\}) \subsetneq \text{conv}(X)$ for every $x \in X$. If X is not convex then those points x for which $\text{conv}(X - \{x\}) = \text{conv}(X)$ are called interior

points, and it is easy to show that every interior point belongs to I. Indeed, if x_0 is an interior point then for every circle C(u,v,w) such that $X - \{x_0\} \subset C^*(u,v,w)$ it follows, since $C^*(u,v,w)$ is convex, that $C^*(u,v,w) \supset \text{conv}(X - \{x_0\}) = \text{conv}(X) \ni x_0$ and hence x_0 is not excludable. To show x_0 is includable we can construct an including circle by first picking $u \in X$ so that $C(x_0,u)$ contains no points of X in its interior. Letting p be a point on the line through x_0 and u and moving p from x_0 away from u, we obtain an expanding family of discs, $C^*(p,u)$, each containing x_0. For some p, say p', C(p',u) will pass through two points of X, u and v, and $C^*(p',u) \cap X = \{x_0,u,v\}$. Since x_0 is an interior point there is some point of X, different from x_0 on the same side of the line through u and v as x_0. Therefore, by Lemma 4 below, we can obtain a circle through u, v, and a third point of X, which contains only x_0 in its interior. Hence, x_0 is includable.

The nature of non-interior points requires a more careful analysis. Thus, from here on, we assume X is convex.

We label the points of X sequentially from 1 to N by going clockwise around the convex figure. Each x_i has x_{i-1} and x_{i+1} as its adjacent points, if we agree to take $x_{N+1} = x_1$ and $x_0 = x_N$. In [2], we showed that given two adjacent points x_i, x_{i+1} there is a circle $C(x_i,x_{i+1},y)$ for some $y \in X$ which excludes exactly one point of X, and a circle $C(x_i,x_{i+1},z)$ for some $z \in X$ which includes exactly one point of X. Suppose the excluded point is x_j. Then by considering x_j and x_{j+1} we are assured, as above, of finding another excludable point. Thus X has at least two excludable points.

Similarly, we can show that X has at least two includable points. Much more, however, is true.

Theorem 1: Let X *be a convex set of* N *points in general circle-free position in the plane, then:*

i) *every point in* X *is either excludable or includable,* i.e. $E \cup I \cup D = X$,

ii) *at least half of the points in* X *are excludable (includable),*

iii) *at least two points of* X *are not excludable (includable), consequently,* $I \neq \phi$ *and* $E \neq \phi$.

The definitions of excludable and includable points give no indication of how to find an excluding (or including) circle for a particular excludable (or includable) point, nor do they provide a constructive means of determining whether a point is excludable or includable. These shortcomings are completely overcome by the next two theorems.

Theorem 2: Let $x_i \in X$ *be an excludable point. Then there is a circle* $C(x_{i-1},x_{i+1},x_j)$, *through some* x_j *and the two points adjacent to* x_i, *which is an excluding circle for* x_i.

If x_i *is an includable point then similarly, for some non-adjacent* x_k, *the circle* $C(x_{i-1},x_{i+1},x_k)$ *is an including circle for* x_i.

Theorem 3: Let C_i *be those points of* X *inside and on the circle through* x_i *and its two adjacent points;* i.e., $C_i = C^*(x_{i-1},x_i,x_{i+1}) \cap X$.

Then

$$x_i \in I \text{ if } C_i = X,$$

$$x_i \in E \text{ if } C_i = \{x_{i-1},x_i,x_{i+1}\},$$

and

$$x_i \in D \text{ if } \{x_{i-1},x_i,x_{i+1}\} \subsetneq C_i \subsetneq X.$$

We will use the following fundamental result. Its proof, which depends on nothing more than the fact that two circles intersect at most in two points, is omitted.

Lemma 4. Given a circle $C(x,y,z)$ *and a point* p, *let* H *be the half-plane containing* p *determined by the line through* x *and* y. *Then* p *is in the interior of* $C(x,y,z)$ *if and only if*

$$C^*(x,y,p) \cap H \subsetneq C^*(x,y,z) \cap H,$$

and p *is in the exterior of* $C(x,y,z)$ *if and only if*

$$C^*(x,y,z) \cap H \subsetneq C^*(x,y,p) \cap H.$$

In view of Lemma 4, Theorem 3 follows immediately from Theorem 2. To prove Theorem 2, let x_i be an excludable point. Thus, there exists a disc $C^*(x_j,x_k,x_m)$ which excludes x_i and contains the remaining points. In particular, x_{i-1} and x_{i+1} are in $C^*(x_j,x_k,x_m)$. Therefore, at least one of x_j, x_k, x_m is not in $C^*(x_{i-1},x_i,x_{i+1})$, and by Lemma 4, x_i can be excluded by a circle through x_{i-1},x_{i+1} and some third point. The part of Theorem 2 concerning includable points is proved in an entirely similar manner.

We turn now to the proof of Theorem 1. Since C_i in Theorem 3 must satisfy one of the three conditions listed there, part i) is proved.

Consider two adjacent points x_i,x_{i+1}. As before $C(x_i,x_{i+1},y)$ for some $y \in X$ excludes exactly one point of X, say x_j. If x_j is not adjacent to either x_i or x_{i+1} then both $C(x_{i-1},x_i,x_{i+1})$ and $C(x_i,x_{i+1},x_{i+2})$ contain x_j in their exteriors. Thus, x_i and x_{i+1} belong to $E \cup D$. If $x_j = x_{i-1}$, then $C^*(x_j,x_i,x_{i+1})$ contains X and by Theorem 3, $x_i \in I$. Similarly, if x_j is x_{i+2}, then $x_{i+1} \in I$. Since x_j cannot be both, one of x_i,x_{i+1} is in $E \cup D$. Therefore in every pair of adjacent points, one point is excludable and the other may or may not be. Hence, the excludable part of ii) is proved; analogous reasoning proves the includable part of ii).

Suppose now every point in X is excludable. Then by Theorem 2 $C(x_N,x_2,x_j)$ is an excluding circle for x_1, where $4 \leqslant j < N$. The index $j \neq 3$, for if it were $C^*(x_1,x_2,x_3)$ would contain X, and hence x_2 would not be excludable. Again by Theorem 2 $C(x_1,x_3,x_k)$ is an excluding circle for x_2, with $k \leqslant j$.

This construction can be continued only to $C(x_{j-3},x_{j-1},x_m)$, the excluding circle for x_{j-2}. Now we must have $x_m = x_j$, and hence x_{j-1} is not excludable. Thus, there is at least one point which is not excludable. Assume there is only one, and renumber the points so that that point is x_N. Then proceeding exactly as before, we deduce that x_{j-1} can not be excluded. Since $j < N$, this is a second point in I. The includable statement can be proved in this same way using including circles. This completes the proof of Theorem 1.

Let N be odd. Then if $M = (N+1)/2$ we have from Theorem 1, part ii), card $(E \cup D) \geqslant M$ and card $(I \cup D) \geqslant M$. By construction E, I, and D are disjoint, therefore $D \neq \phi$. Thus we have shown the following.

Corollary 5. Given the hypothesis of Theorem 1, if N *is odd then there exists a point which is both excludable and includable.*

For N = 4, D is always empty, and for N = 6 there is a configuration of X for which D is empty. Whether or not D must be non-empty for large even N is an open question.

Literature Cited

1. Eggleston, H. G. 1966. Convexity. Cambridge Univ. Press. London.
2. Simons, W., and G. Trapp. 1972. Separating points by spheres. Technical Report TR 72-5. Dept. of Statistics and Computer Science. West Virginia Univ. 4 pp.
3. Lay, S. R. 1971. On separation by spherical surfaces. *Amer. Math. Monthly.* 78:1112-13.
4. Valentine, F. A. 1964. Convex Sets. McGraw-Hill, New York.

A Correspondence Between Control Logic of Associative Memories and Markov Algorithms

David C. Rine
Department of Statistics and Computer Science
West Virginia University, Morgantown, West Virginia 26506

Abstract

The basic structure underlying serial associative memory addressing schemes is currently a conditional logic evaluation. This evaluation is a special case of a Markov algorithm. Decision algebras are used in order to model multiple-valued selector networks and these networks realize serial associative memory addressing schemes. The purpose of this paper is to draw attention to the not well-known nor understood equivalence between the control logic for associative memory addressing schemes and Markov algorithms.

Key words and phrases. Associative memory, content addressable memory, conditional expressions, Post algebras, Markov algorithms, functional-distributed logic.

Introduction

Associative Memories

Associative memories (AM's) are just now becoming popular, and their applications are not completely known (Bell, Bhandarkar, Feucht, Rege, Siewiorek, 1972-1973); one use of AM's is as cache memories; it is felt that these memories may have applications in emulation of existing systems; they can, also, be used to perform a certain class of logic functions (the question of logic functions will be discussed later on in this paper). Another application of associative memories is the Pattern Articulation Unit (PAU) of Illiac III (1971), a large pattern recognition computer; the PAU employs a two-dimensional iterative array of 1024 (32 x 32) identical processing modules locally connected to execute Boolean functions, threshold logic, and signal path building; it is augmented by its own control unit and by an unconventional core memory, called the transfer memory, which in conjunction with the iterative array, operates as an associative memory. The AM of Illiac III processes images in parallel; some AM's associate in parallel, rather than sequencially, in order to speed up processing. Our modeling in this paper, of association, will use sequencial association rather than parallel association. When sequencial association is used the underlying logic is called, by the author, *conditional* (or functional); and, when parallel association is used the underlying logic is called, by the author, *distributed* (or not functional).

Both the IBM 360/67 and IBM 360/85 use eight and 1,000-word associative (content-addressable) memories, respectively, to increase performance, the CDC 6600 instruction buffer is in effect a small content-addressable memory (CAM). In the IBM 360/67, eight associative storage registers are in its associative array; each stores an individual page address. Bits 36 and 37 of each register reflect the validity and the usage of that entry, respectively. Both bits are set to 1 when the register is loaded, and both are reset to 0 when the segment table register (control register 0) contents are changed. If, within a user's time slice, eight

separate pages are referenced (i.e. all associative registers have been loaded and each bit 37 is a 1) all bit 37's are reset to 0. Subsequent references to a register will cause bit 37 for that register to be set. During subsequent translations, the result will be loaded into the *lowest-numbered* register that has a 0 in a bit 37 (the roles of 1 and 0 could have been reversed). Bit 37 is set to a 1 as a result of this load. With this technique, the associative array always contains the most recent and/or most frequently used page addresses. Since instructions frequently occur in sequence, an "IC register" is implemented to hold and provide the translated page address for use in serial instruction sequences; this avoids repeated accesses of the page address from the associative array.

Minsky (1972) has treated rotating storage devices as partially associative memories by addressing conventional cylinders "by content"; the cylinder storage is divided into two parts: data space and control space. Suppose that the items (n, d) are stored by recording *n* on a control cell and d on its *associated* data cell. Given a *predicate* p we can retrieve an item *satisfying* p(n) (e.g. p(n): n.AND.m=b, b a bits pattern and m a mask) by reading continuously the control track, computing p on each control cell. If a control cell containing an n which satisfies p(n) (*true*) is found, then switch to its associated data cell to read the d part of the item.

The purpose of this paper is to draw attention to the not well-known nor understood equivalence between the control logic for associative memory addressing schemes and Markov algorithms.

Methods

Addressing Schemes

Earlier in the paper we mentioned the construction of small prototype associative addressable memories. Let us look at a prototype that incorporates many of the fundamental concepts of AM's or CAM's that we have thus far observed.

Foster (1970) has described the structure for a small simulated associative memory (SAM).

From a user's point of view SAM is a word organized machine with eight bytes per word. These may be 8 bit or 6 bit bytes depending on the implementation. There is a scratch pad memory of 64 words that are directly addressable by number 0-77 octal); there are three high speed registers:

(1) The *Instruction* Register (IR) used to hold the current instruction;
(2) The *Comparand* Register (CR) to hold the pattern of bytes being searched for; and,
(3) The *Mask* Register (MR) which controls which portions of the comparand must be matched in a search operation.

Finally, there is the Associative Memory itself which consists of N eight byte words, where N is as large as is convenient for the installation. Connected with each word of the associative memory there are four tagbits: T_0, T_1, T_2, T_3, and an instruction bit I. The words of the AM are ordered from "beginning" to "end" but have no address. Each word, therefore, has a "predecessor" and a "successor". Words whose $T_0, \ldots, T_3$, I bits are set to 1 are called "responders on $T_0, \ldots, T_3$, I".

The *basic instruction cycle* has three phases:

(1) Find the first (nearest the beginning) responder on I and copy its contents into the IR;

(2) Move the activity on I forward one position; i.e. reset the I bit of the word just read and set the I bit of its successor;
(3) Execute the instruction in the IR.

If we consider all the I bits collectively as the "I-store", then phase two is equivalent to right shift of one of the I-store. The AM is used to hold both instructions and data; and, as is obvious from the above, instructions must be stored in the order in which they are to be executed. The scratch pad memory has two "read only" locations, 00 which contains all zeros and 77 which contains all ones. The remaining 62 locations may be filled with any desired patterns by use of different instructions.

The tag-stores T_0, T_1, T_2, T_3 are used for sequencing (and retrieval) of instructions in the AM; this is accomplished by certain skipping and looping instructions responding to the various logic states that T_0, T_1, T_2 or T_3 (some-

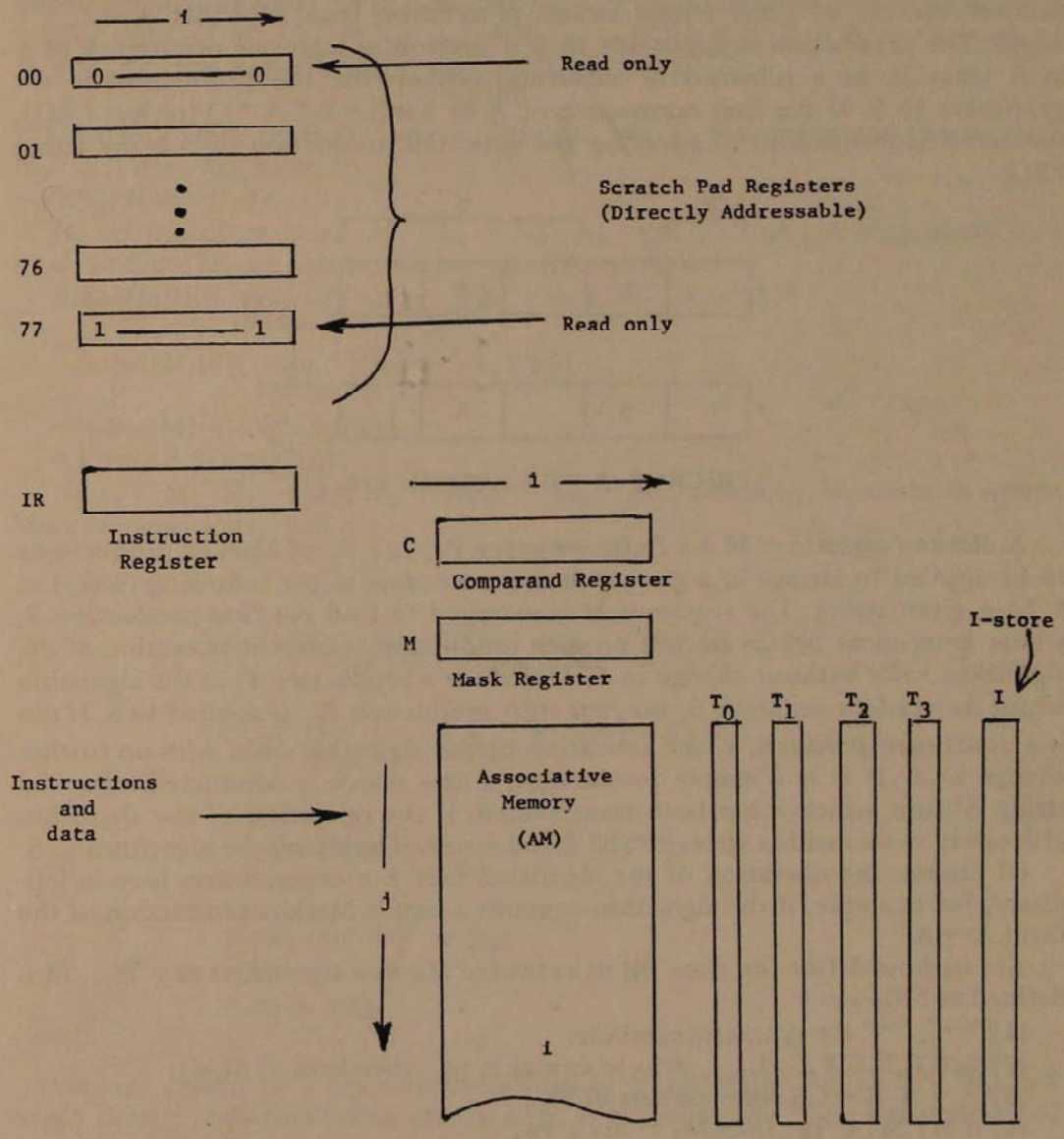


FIGURE 1. Components of SAM.

times in combinations) are set or reset to by another set of logic instructions for T_0, T_1, T_2, T_3; for example, the logic instruction 'TOR t, u, v', for each bit position j in tag-stores t and u, takes the logic result of t_j V u_j and assigns it to the j^{th} bit position of tag-store v, i.e. t_j V $u_j \rightarrow v_j \forall j$.

Markov algorithms have been defined in Curry (1963).

Consider strings of symbols from a given finite symbol set, called the *alphabet.* Assume the alphabet does not contain the (syntactic) symbols "→" and "•" (alphabet with "→" and "•" form a larger symbol set). A *simple (Markov) production* is a string A → B, where A and B are strings in the alphabet. A *conclusive (Markov) production* is a string A→• B, where A and B are strings in the alphabet. In the production A → B (A →• B) the *antecedent* is A and the *consequent* B.

Let A → B (or A→ • B) be a simple Markov production, where A and B are strings in an alphabet $\mathcal{A}$. Let S be a string of symbols in where $S(\mathcal{A})$ denotes the set of finite length strings of symbols from alphabet A, i.e. S ϵ $S(\mathcal{A})$. The production is applicable to S if there is at least one occurrence of A in S (that is, as a subword or substring). Otherwise, the production is *not applicable* to S. If the first occurrence of A in S is S = P * A * Q (or just PAQ), the *result* (substitution) of applying the indicated production of S is the string PBQ.

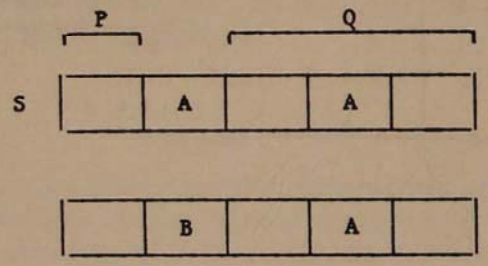


FIGURE 2. A →B is applicable to S.

A *Markov algorithm* M is a finite sequence $P_1, \ldots, P_n$ of Markov productions to be applied to strings in a given alphabet according to the following rules. Let S be a given string. The sequence M is searched to find the *first* production P_i whose antecedent occurs in S. If no such production exists, the operation of the algorithm *halts* without change in S. If there is a production P_j in the algorithm whose antecedent occurs in S, the *first* such production P_{j_1} is applied to S. If this is a *conclusive production,* the operation of the algorithm *halts* with no further change in S. If it is a simple production, a new search is conducted using the string S' into which S has been transformed. If the operation of the algorithm ultimately ceases with a string S*, S* is the result of applying the algorithm to S.

Of course, the operation of the algorithm may not cease; it may loop indefinitely, for example, if the algorithm contains a simple Markov production of the form A → A.

Let us now define the class $\mathcal{M}$ of *extended Markov algorithms* M ϵ $\mathcal{M}$. $\mathcal{M}$ is defined as follows:

(1) "→", "•" are syntactic symbols;
(2) A,B,C,D,E,F,G,H, . . . denote strings in $\mathcal{A}$, members of $S(\mathcal{A})$;
(3) A → B, A →• B are members of $\mathcal{M}$;
(4) if $M_1, M_2 \epsilon \mathcal{M}$, then $M_1 \rightarrow M_2 \epsilon \mathcal{M}$;

(5) if $M_1, M_2, M_3, \ldots, M_n \in \mathfrak{M}$, then the *sequence* $\langle M_i \rangle_{i=1}^{n} = (M_1, M_2, \ldots, M_n) \in \mathfrak{M}$;

(6) nothing is an extended Markov algorithm, in $\mathfrak{M}$, unless it follows from (1)–(5).

Let M be an extended Markov algorithm, $M \in \mathfrak{M}$, and let S be a string in $\mathcal{A}$, $S \in S(\mathcal{A})$. The notation

"Subst(M)(S)(min.{consequence | antecedent $\in$S})" denotes that we are to search for the *minimal*, relative to the natural number ordering, of all consequences in M whose antecedents are in S and *substitute* that minimal consequence for its antecedent which is in S, the result being a new string S', if the minimal consequence exists.

For example, if *Halt* is a symbol and if M: $A \rightarrow B$ and S: PAQ, then Subst(M)(S)(B or *Halt*) = PBQ, *Halt*. If M: $A \rightarrow\cdot B$, then Subst(M)(S)($\cdot$B or *Halt*) = PBQ, *Halt*. If M: $(A_1 \rightarrow X_1, A_2 \rightarrow X_2, \ldots, A_n \rightarrow X_n)$, $X_i = B_i$ or $\cdot B_i$, and A_i, $X_i \in S(\mathcal{A})$, then Subst(M)(S) $\left(\min_{1 \leqslant i \leqslant n} \{X_i \mid A_i \in S\} \right)$

= Subst(M)(S)($X_i\hat{}$ or *Halt*), where $X_i\hat{}$ is the min.X_i if it exists and the result is *Halt* if it does not exist,

= $PX_i\hat{}Q$, *Halt*.

If, on the other hand, M: $(A_1 \rightarrow M_1, A_2 \rightarrow M_2, \ldots, A_n \rightarrow M_n)$ where $A_i \in S(\mathcal{A})$ and the M_i are simple Markov algorithms, then

$$\text{Subst(M)(S)}\left(\min_{1 \leqslant i \leqslant n} \{ (\min_{1 \leqslant j \leqslant n_i} \{X_{i_j} \mid A_{i_j} \in S\}) \mid A_i \in S\} \right)$$

$$= \text{Subst(M)(S)}\left(\min_{1 \leqslant i \leqslant n} \{S(X_{i_j}\hat{}) \mid A_i \in S\} \right)$$

= Subst(M)(S)($X_{ij}\hat{}\hat{}$ or *Halt*)

= $PX_{ij}\hat{}\hat{}Q = S(X_{ij}\hat{}\hat{})$, *Halt*.

Now, if M: $(N_1 \rightarrow M_1, N_2 \rightarrow M_2, \ldots, N_a \rightarrow M_a)$ where N_i, M_i are both simple Markov algorithms, then

$$\text{Subst(M)(S)}\left(\min_{1 \leq i \leq a} \{ (\min_{1 \leq j \leq b_i} \{X_{i_j} \mid A_{i_j} \in S\}) \right.$$

$$\left. (\min_{1 \leq k \leq a_i} \{Y_{i_k} \mid A_{i_k} \in S\})_i \in S\} \right)$$

$$= \text{Subst(M)(S)}\left(\min_{1 \leq i \leq a} \{ X_i \mid Y_i \in S\} \right)$$

$$= \text{Subst(M)(S)}(X_i\hat{} \text{ or } \underline{\text{Halt}})$$

$$= PX_i\hat{}Q, \underline{\text{Halt}}.$$

We will return to a rigorous definition of Subst(M)(S) for all $M \in \mathfrak{M}$, $S \in S(\mathcal{A})$ shortly, but first let us clarify what is meant by the "first antecedent" of M in S.

We read S from left to right and read one letter of S at a time. When we have

completed any occurrence of the antecedent A of A → X, then we store it in the antecedent address.

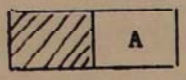


FIGURE 3. Antecedent address.

For example, let S: 11211211211, A: 11211, and B: 33; then, we have S': 33211211. Notice that the occurrences of A, in this case three of them, can be overlapping and can be such that more than just one A (in this case two) is removed by the substitution.

The possibilities can be even worse, and the number of A's removed from S by substituting for one A can become unbounded. Letting S: 111111111111111 consist of fifteen 1's and letting A: 11111111 consist of eight 1's, we see that removing the left-most A, S': B1111111, at the same time removes eight additional A's from S, these A's not being available for use in future substitutions. Moreover, if $2n - 1$ 1's occur, where A: 11 . . . 1 is n 1's, then removing one A removes n A's of length n. Letting S: 10101010101010101010 consist of eleven 10's and A: 101010101010, then by removing the left-most A we remove six A's; and if S: 10 . . . 10 has $2n - 1$ 10's, n the number of 10's in A, then removing the left-most A removes n A's which will not be available for use in future substitutions.

Let us now formally define the mapping Subst(S): $\mathcal{M} \rightarrow S(\mathcal{A})$, for every string S, where $\mathcal{M}$ is the class of extended Markov algorithms and $S(\mathcal{A})$ is the class of strings over $\mathcal{A}$ including a symbol denoted "*Halt*".

First, let us define what is meant by the *rank* of members of $\mathcal{M}$. Inductively, define the *rank* of $M \in \mathcal{M}$ as follows:

(i) A,B,C,D,E,F,G,H, . . . strings in $\mathcal{A}$ are of rank 1;

(ii) Markov algorithms are of rank 2;

(iii) $M \in \mathcal{M}$ is of rank α ($\alpha \geq 2$, α positive integer) if it *contains* at least one $M' \in \mathcal{M}$ (or $S(\mathcal{A})$) of rank $\alpha - 1$, as either an antecedent or consequence, but none of rank greater than $\alpha - 1$, i.e. $(M_1 \rightarrow N_1, M_2 \rightarrow N_2, \ldots, M_k \rightarrow N_k)$ is of rank α if there exists an M_i or N_i $(i=1, \ldots, k)$ such that the rank of M_i or N_i is $\alpha - 1$ (but no M_i or N_i has rank greater than $\alpha - 1$).

Define Subst(S): $\mathcal{M} \longrightarrow S(\mathcal{A})$ inductively as follows:

(I) Let $A \in \mathcal{A}$ be of rank 1; then, since $A \in \mathcal{M}$ we define

Subst(A)(S)(A or Halt) = PAQ, Halt.

Let $M = (A_1 \longrightarrow X_1, A_2 \longrightarrow X_2, \ldots, A_n \longrightarrow X_n) \in \mathcal{M}$ be of rank 2; then, we define $\text{Subst}(M)(S)(\min_{1 \leq k \leq n} \{X_k \mid A_k \in S\}) = \text{Subst}(M)(S)(X_k$ or Halt) $= PX_k Q$, Halt, where $S = PA_k Q$ and min. has a value of Halt if the minimum does not exist; but, Subst(s) does exist then with a value of Halt.

(II) Assume that arbitrary $M' \in \mathcal{M}$ is of rank α and that

$$\text{Subst}(M')(S)(\min_{1 \leq j(a) \leq n(a)} \{\ldots(\min_{1 \leq k(b) \leq n_j} \{X_{j_{k(b)}} \mid A_{j_{k(b)}} \in S\}) \blacksquare$$

$$\text{antecedent}\Big(\min_{1\le k(b)\le n(b)_j}\{X_{j_{k(b)}} \mid A_{j_{k(b)}} \in S\} \in S \cdots\}\Big)$$

$$= \text{Subst}(M')(S)\Big(\min_{1\le j(a)\le n(a)}\{\cdots (X_{j_{k(b)}}\hat{\ }) \blacksquare \text{antecedent}(X_{j_{k(b)}}\hat{\ }) \in S \cdots\}\Big)$$

$$= \text{Subst}(M')(S)(X_{j_{k(b)}}\hat{\ } \text{ or } \underline{\text{Halt}}) = PX_{j_{k(b)}}\hat{\ }\, Q, \underline{\text{Halt}},$$

is defined where $S = PA_{j_{k(b)}}\hat{\ }\, Q$; this is the induction hypothesis. Then,

we can define for every arbitrary $M \in \mathcal{M}$ of rank $\alpha + 1$

$$\text{Subst}(M)(S)\Big(\min_{1\le i\le n}\{\Big(\min_{1\le j(a)\le n(a)_i}\{\ldots\Big(\min_{1\le k(b)\le n(b)_{i_j}}\{X_{(i,j,k(b))} \mid A_{(i,j,k(b))} \in S\}\Big)$$

$$\text{Antecedent}\Big(\min_{1\le k(b)\le n(b)_{i_j}}\{X_{(i,j,k(b))} \mid A_{(i,j,k(b))} \in S\}\Big) \in S \ldots\}\Big)$$

$$\Big(\min_{1\le t(c)\le m(c)\hat{\ }_i}\{\ldots\Big(\min_{1\le u(d)\le m(d)\hat{\ }_{i_r}}\{Y_{(\hat{i},t,u(d))} \mid A_{(\hat{i},t,u(d))} \in S\}\Big)$$

$$\text{Antecedent}\Big(\min_{1\le u(d)\le m(d)\hat{\ }_{i_r}}\{Y_{(\hat{i},t,u(d))} \mid A_{(\hat{i},t,u(d))} \in S\}\Big) \in S \ldots\}\Big) \in S\}\Big)$$

$$= \text{Subst}(M)(S)\Big(\min_{1\le i\le n}\{X_{(i,\hat{j},k(b))} \quad Y_{(\hat{i},\hat{t},u(d))} \in S\}\Big)$$

$$= \text{Subst}(M)(S)(X_{(\hat{i},\hat{j},\hat{k}(b))} \text{ or } \underline{\text{Halt}}) = PX_{(\hat{i},\hat{j},\hat{k}(b))}\, Q, \underline{\text{Halt}}.$$

for every i before and after the conditional symbol "■" by the induction hypothesis for arbitrary M' $\in \mathcal{M}$ of rank α, and then by the initial induction step for rank 2. (Above we have used "(i,j,k)" instead of triple subscripts.)

This completes the definition of the mapping Subst(S) from $\mathcal{M}$ into $S(\mathcal{A})$.

Now, in order to work with heavy notation more easily let us define two special mappings D_s, $\overline{D}_s$: $\mathcal{M} \to S(\mathcal{A})$, $S(\mathcal{A})$ includes *Halt*, for every $S \in S(\mathcal{A})$ as follows:

For every $P \in \mathcal{M}$

$$D_s(P) = \begin{cases} \text{Subst}(P)(S), & \text{antecedent}(P) \in S \\ \underline{\text{Halt}}, & \text{antecedent}(P) \notin S; \end{cases}$$

$$\overline{D}_s(P) = \begin{cases} \text{Subst}(P)(S), & \text{antecedent}(P) \notin S \\ \underline{\text{Halt}}, & \text{antecedent}(P) \in S. \end{cases}$$

Let K be a complete lattice (eventually it will be N, the counting numbers, or a hierarchical graph complete lattice) and let $\hat{s}$: K → $\mathcal{M}$ be a mapping.

Results

Building Schemes and Selector Networks

We now have the following theorem where the equivalence is determined by evaluating the left and right hand expressions.

Theorem. Using Subst(S), we have the equivalence $M = \underset{j \in K}{\text{lub}}$ (glb(Conseq(M(j)), $\overline{D}_{Halt}$ [$\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))}], D_s (antec(M(j))))), where M is an extended Markov algorithm over (K, $\mathcal{M}$), the lub and glb occurring in the induced lattice structure of K in $\mathcal{M}$.

Proof: Let M ∈ M be an extended Markov algorithm and S ≠ *Halt*. Subst(M)(S) is as defined in Definition I-II. Let m ∈ K. Suppose that ∀ k ∈ K ∋ k < m we have antec(M(k)) ∈ S; but, antec(M(m)) ∈ S. Now, V k< m Subst(D_s (antec(M(k))))(S) = *Halt* but Subst(D_s (antec(M(m))))(S) ≠ *Halt*.

Thus, if m < j, then Subst($\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))})(S) ≠ *Halt*, Subst($\overline{D}_{Halt}$ [$\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))}])(S) = *Halt*, and Subst(glb(Conseq(M(j)); $\overline{D}_{Halt}$ [$\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))}], D_s (antec(M(j))))(S) = *Halt*. If m ⩾ j, then Subst($\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))})(S) = *Halt*, and Subst($\overline{D}_{Halt}$[$\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))}])(S) ≠ *Halt*. If antec(M(k)) ∈ S, then the Subst(glb(Conseq(M(j)), $\overline{D}_{Halt}$[$\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))}], D_s (antec(M(j))))(S) is determined by Conseq (M(j)); otherwise, it is *Halt. Now, a greatest lower bound k, say* k_0, *such that antec(M(*k_0*))* ∈S. For each j∈ K such that j < k_0 Subst(glb(Conseq(M(j)), $\overline{D}_{Halt}$[$\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))}], D_s (antec(M(j))))(S) = *Halt*. For each j ∈ K such that j > k_0, then antec(M(j)) ∈ S and Subst(D_s (antec(M(j))))(S) ≠ *Halt* or antec(M(j)) ∈ S and Subst(D_s (antec(M(j))))(S) = *Halt*. Whenever antec(M(j)) ∈ S Subst(glb-(Conseq(M(j)), $\overline{D}_{Halt}$[$\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))}], D_s (antec(M(j))))(S) would be determined by Conseq(M(j)) except for the fact that for j > k_0 Subst($\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))})(S) ≠ *Halt* and Subst($\overline{D}_{Halt}$[$\underset{k<j}{\text{lub}}$ {D_s (antec(m(k)))}])(S) = *Halt;* so, Subst(glb(Conseq(M(j)), $\overline{D}_{Halt}$[$\underset{k<j}{\text{lub}}$ {D_s (antec(M(k)))}], D_s (antec(M(j))))(S) = *Halt*. Now let j = k_0. Subst($\underset{k<k_0}{\text{lub}}$ {D_s (antec(M(k)))})(S) = *Halt*, and Subst($\overline{D}_{Halt}$[$\underset{k<k_0}{\text{lub}}$ {D_s (antec(M(k)))}])(S) ≠ *Halt;* moreover, Subst(glb(Conseq(M(k_0)), $\overline{D}_{Halt}$[$\underset{k<k_0}{\text{lub}}$ {D_s (antec(M(k)))}], D_s

$(\text{antec}(M(k_0)))))(S) = \text{Subst}(\text{Conseq}(M(k_0)) \wedge \bar{D}_{Halt}[\underset{k<k_0}{\text{lub}} \{D_s(\text{antec}(M(k)))\}]$

$\wedge D_s(\text{antec}(M(k_0))))(S)\ \text{Subst}(\text{Conseq}(M(k_0)))(S) = \text{Subst}(M)(S)(\underset{k}{\text{glb}}\{\text{Conseq}$

$(M(k)) \blacksquare \text{antec}(M(k)) \in S\})$, which is the substitutional value of M with respect to S. Thus, $\text{Subst}(M)(S) = \text{Subst}(\underset{k\in k}{\text{lub}}(\text{glb}(\text{Conseq}(M(j)), \bar{D}_{Halt}[\underset{k<j}{\text{lub}}\{D_s(\text{antec}$-

$(M(k)))\}], D_s(\text{antec}(M(j))))))(S)$. This proves the theorem. ▲

The previous theorem allows one to derive quite a large number of addressing schemes for serial associative memory computers. For example, let $x = (a_1 \to b_1, \ldots, a_m \to b_m)$ be a Markov algorithm. Let the a_k's come from the set $\{0, 1\}$, let the b_k's come from the set $\{0, 1\}$, and let S = 1. Then, we have that the result of $\text{Subst}(x)(1)(\underset{1<k<m}{\text{min.}}\{b_k \mid a_k = 1\})$ gives us the first b_k such that $a_k = 1$.

Next, if we take b_k to be a string in a Boolean algebra, then it has a logic value of 0 or 1. Under these conditions we first apply Subst(x)(1) and then take the value, *val* for short, of b_k. Hence, we can define the logical value of x to be $\text{val}(x) = \text{val}(\underset{1\leqslant k\leqslant m}{\text{min.}}\{b_k \mid a_k = 1\})$. If instead of just bits a_k we replace the a_k's by Boolean logic conditions s_k, then we first apply val to s_k's in order to get the bits. Then, we first do $\text{Subst}(x)(1)(\underset{1\leqslant k\leqslant m}{\text{min.}}\{b_k \mid \text{val}(s_k) = 1\})$ and then take

$\text{val}(x) = \text{val}(\underset{1\leqslant k\leqslant m}{\text{min.}}\{b_k \mid \text{val}(s_k) = 1\})$.

When one specializes Markov algorithms in this way to strings of logical expressions (e.g. Boolean expressions), then a Markov algorithm becomes what is known as a "Conditional Expression". We have the following easy logical equivalences for propositions p,q:

(conjunction) $p \wedge q = (p \to q, 1 \to 0) = q \to p, 1 \to 0)$,
(disjunction) $p \vee q = (p \to 1, 1 \to q) = (q \to 1, 1 \to p)$,
(negation) $\neg p = (p \to 0, 1 \to 1)$, and
(implication) $p \Rightarrow q = (p \to q, 1 \to 1)$.

It is much easier to think of associative addressing schemes in terms of "selector networks". These networks can be constructed in multiple-valued logic as well as Boolean logic. The following conceptualization was first noticed by Rosenbloom (1942), but in a much less concise formulation. An n-valued *Decision* Algebra (Post Algebra) is defined by an n, fixed integer, $n \geqslant 2$, and a distributive lattice L with zero and unit u such that:

axiom 1. For every element $x \in L$ there exist n elements $C_0(x)$, $C_1(x), \ldots, C_{n-1}(x)$ which are pairwise disjoint and whose supremum is u; that is, $C_i(x) \wedge C_j(x) = 0$ for $i \neq j$ and $\overset{n-1}{\underset{i=0}{\vee}} C_i(x) = u$.

axiom 2. There exist n fixed elements of L, denoted $0 = f_0, f_1, \ldots, f_{n-2}, f_{n-1} = u$ with the properties:
2a. The elements form a *chain*, with $f_{i-1} \leqslant f_i$ for $1 \leqslant i \leqslant n-1$.
2b. If $x \in L$ and $x \wedge f_1 = 0$, then $x = 0$.
2c. If $x \in L$ and, for some i, $x \vee f_{i-1} = f_i$, then $x = f_i$.

axiom 3. For every $x \in L$, $x = \overset{n-1}{\underset{i=0}{\vee}} (f_i \wedge C_i(x))$.

The following corollaries state that any serial addressing scheme taking on multiple values (a multiple-valued conditional expression) can be built up from operators (gates) of an n-valued decision algebra L in three distinct ways, and such a scheme is a selector network. The algebra "inclusion" suggested by the corollaries is proper (e.g. if i=1, then $C_0(p)$, $C_2(p)$ are not equivalent to any conditional expression). Note that the Post Algebra abbreviation for $\theta = f_0 \leqslant f_1 \leqslant f_2 \ldots \leqslant f_{n-2} \leqslant f_{n-1} = u$ or $\{f_0, f_1, f_2, \ldots, f_{n-2}, f_{n-1}\}$ is $\{0, 1, 2, \ldots, n-2, n-2\}$.

Corollary 1. Let $x \equiv (s_1 \rightarrow r_1, s_2 \rightarrow r_2, \ldots, s_m \rightarrow r_m)$ be a conditional expression such that for i fixed $(0 \leqslant i \leqslant n-1)$ $\mathrm{val}(x) = \mathrm{Subst}(x)(i)\ \mathrm{val}(\min_{1 \leqslant k \leqslant m} \{r_k \mid \mathrm{val}(s_k) = i\})$. Then, $x = \bigvee_{j=1}^{m} (r_j \wedge \{C_0[\bigvee_{k=1}^{j-1} C_i(s_k)]\} \wedge C_i(s_j))$, where C_0 C_1, $C_2, \ldots, C_m, \ldots, C_{n-1}$ are the unary operators of the decision algebra L.

Corollary 2. Let $x \equiv (s_1 \rightarrow r_1, s_2 \rightarrow r_2, \ldots, s_m \rightarrow r_m)$ be a conditional expression such that for i fixed $(0 \leqslant i \leqslant n-1)$ $\mathrm{val}(x) = \mathrm{Subst}(x)(i)\ \mathrm{val}(\min_{1 \leqslant k \leqslant m} \{r_k \mid \mathrm{val}(s_k) \geqslant i\})$. Then, $x = \bigvee_{j=1}^{m} (r_j \wedge \{C_0[\bigvee_{j=1}^{j-1} D_i(s_k)]\} \wedge D_i(s_j))$ where D_i is defined by

$$D_i(y) = \bigvee_{t=1}^{n-1} C_t(y).$$ (High-modified valuation.)

Corollary 3. Let $x \equiv (s_1 \rightarrow r_1, s_2 \rightarrow r_2, \ldots, s_m \rightarrow r_m)$ be a conditional expression such that for i fixed $(0 \leqslant i \leqslant n-1)$ $\mathrm{val}(x) = \mathrm{Subst}(x)(i)\ \mathrm{val}(\min_{1 \leqslant k \leqslant m} \{r_k \mid \mathrm{val}(s_k) \leqslant i\})$. Then, $x = \bigvee_{j=1}^{m} (r_j \wedge \{C_0[\bigvee_{k=1}^{j-1} \bar{D}_i(s_k)]\} \wedge \bar{D}_i(s_j))$ where D_j is defined by

$$\bar{D}_i(y) = \bigvee_{j=0}^{i} C_j(y).$$ (Low-modified valuation.)

The author wishes to thank the referees for suggestions leading to improved presentation.

Literature Cited

1. Bell, C. G., D. P. Bhandarkar, D. L. Feucht, S. L. Rege, and D. L. Siewiorek. 1972. Large scale integration—a designer's viewpoint. Technical Report, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pa.
2. Curry, H. B. 1963. Foundations of mathematical logic. McGraw-Hill, New York. See Chapter 2, Sec. E.
3. Foster, C. C. 1970. A simulated associative memory. Technical Report, Computer Science Department, University of Massachusetts, Amherst, Mass.
4. IBM System/360 Model 67. Functional Characteristics. IBM Systems Reference Library.
5. Illiac III Reference Manual, Volume I: The Computer System. 1971. Pattern articulation unit. Contribution by R. T. Borovec, R. P. Harms, and G. T. Lewis. Edited by B. H. McCormick and B. J. Nordmann. Report No. 433, Department of Computer Science, University of Illinois, Urbana, Ill.
6. Rosenbloom, P. C. 1942. Post algebras I. Postulates and general theory. *Amer. J. Math.* 64:167-88.
7. Minsky, N. 1972. Rotating storage devices as partially associative memories. Technical Report, Department of Computer Information and Control Sciences, University of Minnesota, Mpls., Minnesota.

8. Minsky, N. 1972. Rotating storage devices as partially associative memories. Proc. AFIPS 1972 Fall Joint Computer Conf., Vol. 41, Part I:587-95. *Computing Reviews*, 1973, 14:25,560.

A Note on f-Ideals in Rings of Continuous Functions

W. Thurmon Whitley
Department of Mathematics
Marshall University, Huntington, West Virginia 25701

Abstract

In this paper, we study the relationships between semiprime ideals and f-ideals in the ring of continuous realvalued functions on a topological space. In this regard, we are able to obtain a converse to a theorem of J. G. Horne, and we are able to characterize semiprime ideals in terms of f-ideals.

Introduction

Let X be a topological space, and let C(X) denote the ring of continuous, real-valued functions on X. An ideal I in C(X) is an *f-ideal* if for each f,g in I, there is c in I such that f = cx and g = cy have solutions in C(X) [1]. We say that I is *semiprime* if it is an intersection of prime ideals.

Unless otherwise noted, all other terms will be as given in [2].

The Results

In 1958, J. G. Horne [3] proved that every semiprime ideal in C(X) is an f-ideal. In this section, we give an independent proof of Horne's result, and a converse to Horne's result.

The following lemmas will be useful.

Lemma 1. If f,g in C(X), with $|f| \leqslant |g|^r$ for some real $r > 1$, then f is a multiple of g. [2;21].

Lemma 2. An ideal I in a commutative ring is semiprime if and only if a^2 in I implies a is in I. [2;31].

Lemma 3. Every prime ideal in C(X) is absolutely convex. [2;69].

Theorem 1 (Horne's Theorem). If I is a semiprime ideal in C(X), then I is an f-ideal.

Proof

In view of lemma 3, it is clear that I is absolutely convex. Let f,g be in I. Then |f|, |g| in I, so $|f| + |g| = [(|f| + |g|)^{1/2}]^2$ is in I. Hence, by lemma 2, $(|f| + |g|)^{1/2}$ is in I. Clearly, $|f| \leqslant [(|f| + |g|)^{1/2}]^2$, so by lemma 1, f is a multiple of $(|f| + |g|)^{1/2}$. Similarly, g is a multiple of $(|f| + |g|)^{1/2}$. Therefore, I is an f-ideal.

We first note that the direct converse of Theorem 1 is not true. If (i) is the

principal ideal generated by the identity function i on the real line, we easily see that (i) is an f-ideal which is not semiprime.

Lemma 4. If I is an f-ideal such that $I^n = I$ for some positive integer n, then for each f in I, there exists c in I such that $f = c^n x$ has a solution in C(X).

The proof of lemma 4 is straightforward, and hence omitted. We might remark that lemma 4 remains true in general commutative rings with unity.

Lemma 5. An idempotent f-ideal I in C(X) is absolutely convex.

Proof

We first show that I is convex. It is enough to show that g in I and $0 \leqslant f \leqslant g$ imply that f is in I. Since $I^2 = I$, we have $I^4 = I$, so by lemma 4, there exist c in I and k in C(X) such that $g = c^4 k$. Then $0 \leqslant f \leqslant c^4 k$ implies $0 \leqslant f^{1/2} \leqslant c^2 k^{1/2} = (ck^{1/4})^2$. Then by lemma 1, $f^{1/2}$ is a multiple of $ck^{1/4}$, and hence a multiple of c. Thus, $f^{1/2}$ is in I, so f is in I.

Now to show that I is absolutely convex, it is enough to show that f in I implies |f| in I, since I is convex. Again using lemma 4, if f is in I, there is c in I and k in C(X) with $f = c^2 k$. Thus, $|f| = c^2 |k|$. Hence, |f| is in I.

Theorem 2. If I is an idempotent f-ideal in C(X), then I is semiprime.

Proof

We use lemma 2. Suppose f^2 is in I. By lemma 4, there exist c in I and k in C(X) such that $f^2 = c^2 k$. Then $|f| = |c|\,|k|^{1/2}$. Since I is absolutely convex by lemma 5, |c| is in I since c is in I. Hence, |f| is in I. Therefore, f is in I since I is convex. This proves the theorem.

Corollary 1. An ideal in C(X) is semiprime if and only if it is an idempotent f-ideal.

Proof

This is immediate from Theorems 1 and 2, and from the easily seen fact that semiprime ideals are idempotent.

The author is indebted to the referee for useful suggestions about this paper.

Literature Cited

1. Aull, C. E. 1967. Ideals and filters. *Composito Math.* 18:79-86.
2. Gillman, L., and M. Jerison. 1960. Rings of continuous functions. Van Nostrand. Princeton, New Jersey, 300p.
3. Horne, J. G., Jr. 1958. On 0_ω ideals in C(X). *Proc. Amer. Math. Soc.* 9:511-18.

Torsion Submodule and Injective Hull

David Cusick
Department of Mathematics
Marshall University, Huntington, West Virginia 25701

Abstract

The torsion submodule and the torsion ideal are defined. We then prove a theorem relating these objects to the injective hull. This yields information about the injective hull of an arbitrary ring. Finally, we prove a theorem concerning the Jacobson radical of the ring of endomorphisms of an injective module.

The study of rings of quotients began with the familiar theorem which states that a commutative integral domain can be embedded in a field. The proof is just as familiar; one makes "fractions" using the elements in the integral domain. In more general rings this process encounters difficulties. To overcome these problems, several generalizations of quotient structure have been introduced and studied. Some of these generalizations have, themselves, been generalized to the study of modules. Important examples of this are the concepts of essential extension and injective hull, which we shall consider here. Our results arose from this "quotient ring" context.

Preliminaries

Throughout this discussion R will be an associative ring which is not assumed to have a multiplicative identity. A prefix of R before any term will denote that the term applies when the underlying ring is R. The symbol 0 will be used for zero elements and for zero ideals and submodules; the context will prevent ambiguity. As usual, ker(f) will denote the kernel of the homomorphism f, and $\mathrm{Hom}_R(A,B)$ will denote the set of all R-homomorphisms from A into B. $\mathrm{End}_R(A) = \mathrm{Hom}_R(A,A)$. It is quite well-known that every R-module M can be embedded in an injective R-module E, such that the image of M is large in E. This module, E, is unique up to isomorphism. It is called the *injective hull* of M; and we denote it by $E_R(M)$, omitting the subscript R if the ring is clear from context.

Let M be an R-module, let x be an element of M, and let N be an R-submodule of M. We denote by (N:x) the set of elements r in R such that xr is an element of N. It is well-known that (N:x) is a right ideal of R. Let N be an R-submodule of M. We define the *closure* of N in M to be the set of elements x of M such that (N:x) is large in R. We will write cl (N) if the M is obvious. A submodule N of M is called *closed* in M if $\mathrm{cl}_M(N) = N$. It is clear that $N \subseteq \mathrm{cl}_M(N)$. In 1964, A. W. Goldie [1] proved that cl(cl(N)) is closed. Define $Z_R(M) = \mathrm{cl}_M(\{0\})$; and define $Z_R^*(M) = \mathrm{cl}_M(Z_R(M))$, called the *singular* and *torsion* submodules, respectively. It follows from the above Goldie's proposition that $M/Z_R^*(M)$ has a zero singular submodule (and hence, a zero torsion submodule). In the literature, $Z_R^*(M)$ is usually denoted $Z_2(M)$. As before, we will drop the subscript R if there will be no ambiguity.

These are actually R-submodules of M as their names imply. If M = R, then

these objects are called the singular and torsion *ideals* since they will be two-sided ideals of R. In fact, if I is a two-sided ideal of R, then so is $cl_R(I)$. So in particular, R/Z(R) and R/Z*(R) are both rings. Easy calculations also verify that $M \cdot Z(R) \subseteq Z(M)$ and $M \cdot Z^*(R) \subseteq Z^*(M)$. Since we will be interested in modules over R/Z*(R), we note that if I is a two-sided ideal of R, then an R-module M is also a natural R/I-module if $M \cdot I = 0$. Hence, M/Z*(M) is naturally an R/Z*(R)-module. It can also be shown that the R/Z*(R)-torsion submodule of M/Z*(M) is zero. In 1959, G. Azumaya [3] proved that if I is an ideal of R, and if M is an R/I-module, then $E_{R/I}(M) = \{x \in E_R(M) \mid xI = 0\}$. Thus, $E_{R/I}(M)$ is the largest R/I-module contained in $E_R(M)$.

Results

We begin with the following result.

Theorem: For any R-module M, the following isomorphisms hold: $E_R(M/Z^*(M)) \cong E_R(M)/E_R(Z_R^*(M)) \cong E_{R/Z^*(R)}(M/Z^*(M))$.

Proof

Let $M^* = M/Z_R^*(M)$, let $R^* = R/Z_R^*(R)$, and let $Z^* = Z^*(R)$. Since $M^* \cdot Z^*(R) \subseteq Z^*(M^*) = 0$, we know that M* is naturally an R*-module. Then by Azumaya's result, we see that $E_{R^*}(M^*) \subseteq E_R(M^*)$. To show the opposite containment, it is sufficient to show that Z* annihilates $E^* = E_R(M^*)$. Suppose not. Then $0 \neq E^* \cdot Z^* \subseteq Z_R^*(E^*)$. Therefore, $0 \neq M^* \cap Z_R^*(E^*)$ since M* is R-large in E*. But on the other hand, $0 = Z_R^*(M^*)$ by Goldie's theorem, and $Z^*(M^*) = M^* \cap Z^*(E^*)$ since, in general, $X \subseteq Y$ implies $Z^*(X) = X \cap Z^*(Y)$. This contradiction betrays our supposition; consequently, Z* annihilates E*, as required. This proves that $E^* = E_{R^*}(M^*)$.

To show that the second module is isomorphic to the first, we should remark that $E(Z^*(M)) = Z^*(E(M))$. To see this, we note that Z*(E(M)) is closed in E(M), and therefore, Z*(E(M)) is a direct summand of E(M). (See Harada [4], prop. 1.5). Thus, Z*(E(M)) is injective. Since $Z^*(M) = M \cap Z^*(E(M))$, we see that Z*(M) is R-large in Z*(E(M)). Consequently, Z* (E(M)) is the injective hull of Z*(M), or in symbols, $Z^*(E(M)) = E(Z^*(M))$. It is clear that the other direct summand is isomorphic to $E(M)/E(Z^*(M))$, or equivalently, isomorphic to $E(M)/Z^*(E(M))$. Let $E = E_R(M)$, and consider figure 1 a, where p is the natural epimorphism and i is the inclusion map. Then since E(M*) is injective, we can complete the diagram as in figure 1 b so that the completed diagram commutes. Let K be the kernel of f. Then $f(E) \cong E/K$ and $M^* \subseteq f(E) \subseteq E(M^*)$; thus, by identifying, we assume that $M^* \subseteq E/K \subseteq E(M^*)$. But then these inclusions are both R-large since M* is R-large in E(M*). Therefore, $Z(E/K) = 0$ due to the fact that $0 = Z(M^*) = M^* \cap Z(E/K)$. Consequently, K is closed in E since for a general homomorphism φ, $\varphi^{-1}(Z(A/B)) = cl_A(B)$. The commuting diagram in figure 1 b implies that $Z^*(M) = K \cap M$. Then Z*(M) is large in K since M is large in E. Thus, we are able to conclude that K is contained in E(Z*(M)) since $E(Z^*(M)) = Z^*(E(M))$ is the unique injective hull of Z*(M) which is contained in E(M) and that the containment is large. Hence, $K = E(Z^*(M))$ since a module which is closed in E contains all of its essential extensions which lie in E. Thus, $E/K = E/E(Z^*(M))$, which is injective. But since E/K was demonstrated to be large in E(M*), we see that $E/K = E(M^*)$. Thus, $E/E(Z^*(M)) = E(M/Z^*)$, completing the proof.

In the following easy corollaries, we appeal to a statement of a theorem of Johnson and Wong [6]. The statement can be found also in Faith [2, p. 69].

Corollary. Let R be any ring. Then we have the following R-isomorphisms: $E_R(R)/E_R(Z^*(R)) \cong E_R(R/Z_R^*(R)) \cong E_{R/Z^*(R)}(R/Z^*(R))$. In addition, these modules are also self-injective regular rings which are ring isomorphic to their rings of endomorphisms. Moreover, they are Jacobson semi-simple, i.e., the Jacobson radicals of these rings are all zero, and the ring multiplications extend the R-module multiplications.

Proof

The R-isomorphisms follow from applying the theorem when M is set equal to R. The remainder of the corollary follows from the theorem of Johnson and Wong.

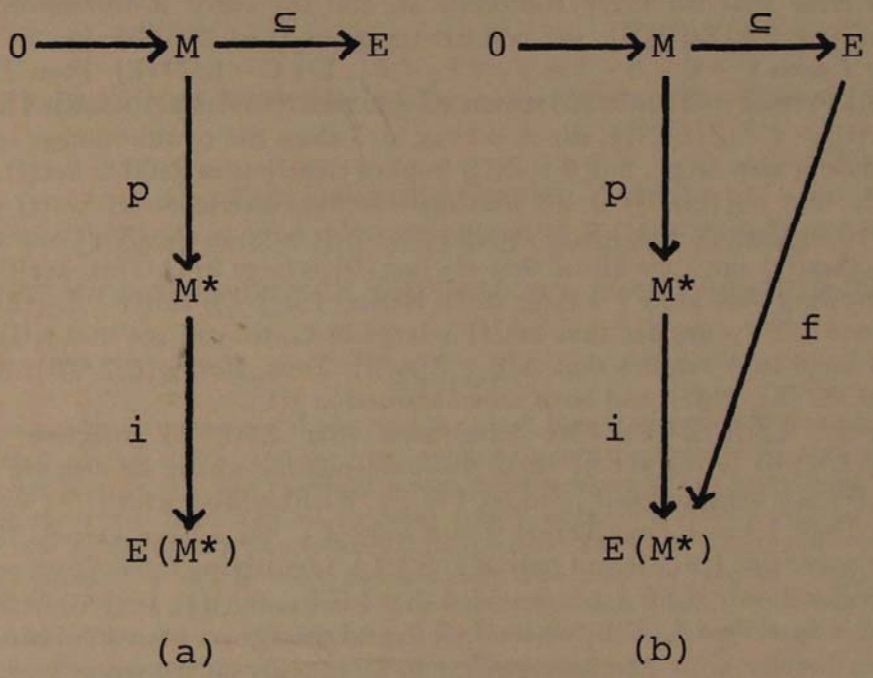


FIGURE 1. An application of injectivity.

Corollary. Let R be any ring. Then the R-injective hull of R contains—as an R-direct summand—a self-injective regular ring which is isomorphic to its ring of endomorphisms and which is Jacobson semi-simple.

Proof

We know that $E(R) \cong E(Z^*(R)) \oplus E(R)/E(Z^*(R))$. The second summand was shown in the last corollary to have the necessary properties.

We turn our attention now to an application of these results. In the proof of the next theorem we repeatedly make use of the well-known result: if $N_i \subseteq M_i$ for $i = 1,2$, then $N_1 \oplus N_2$ is large in $M_1 \oplus M_2$ if and only if N_i is large in M_i for $i = 1,2$.

Theorem. Let E be an injective R-module, and let $L = \mathrm{End}_R(E)$. Then $J(L) = J(\mathrm{End}_R[Z_R^*(E)]) \oplus \mathrm{Hom}_R(E/Z_R^*(E), Z_R^*(E))$, where J denotes the Jacobson radical. Moreover, $\mathrm{Hom}_R(E/Z_R^*(E), Z_R^*(E)) = \mathrm{Hom}_R(E/Z_R^*(E), Z_R(E))$.

Proof

It is well-known that $\{f \in L \mid \ker(f)$ is R-large in $E\} = J(L)$, the Jacobson radical of L. Since $E = Z^*(E) \oplus E/Z^*(E)$, it is clear that $L \cong \mathrm{End}_R(Z^*(E)) \oplus \mathrm{End}_R(E/Z^*(E)) \oplus \mathrm{Hom}_R(E/Z^*(E), Z^*(E)) \oplus \mathrm{Hom}_R(Z^*(E), E/Z^*(E))$; say $L \cong F_1 \oplus F_2 \oplus F_3 \oplus F_4$ for short.

But if $f \in F_4$, then $f: Z^*(E) \to E/Z^*(E)$. Clearly, homomorphisms take Z^* into Z^*; so $f(Z^*(E)) \subseteq Z^*(E/Z^*(E)) = 0$, implying $f = 0$. Thus, $F_4 = 0$; that is, $\mathrm{Hom}_R(Z^*(E), E/Z^*(E)) = 0$. Hence, $L \cong F_1 \oplus F_2 \oplus F_3$.

If $f \in F_2$, then we can identify f with $f' = 0 + f + 0$ in the $F_1 \oplus F_2 \oplus F_3$ representation of L. Then $f'(Z^*(E)) = 0$. So $\ker(f')$ is large in E if and only if $\ker(f)$ is large in $E/Z^*(E)$. But if $\ker(f)$ is large in $E/Z^*(E)$, then $f(E/Z^*(E)) = Z(f[E/Z^*(E)]) \subseteq Z(E/Z^*(E)) = 0$. By identifying f and f', we see that $0 \neq f \in F_2$ implies that f is not in J(L)—that is, $F_2 \cap J(L) = 0$.

It is clear that for every R-module M, and for every R-homomorphism f, $\mathrm{cl}_M(\ker(f)) = f^{-1}[Z(f(M))]$, the counter-image in M of $Z(f(M))$. Let $f \in F_3$ and identify f with $f' = 0 + 0 + f$ in $F_1 \oplus F_2 \oplus F_3$. Let $G = E/Z^*(E)$. Then $Z(f(G)) = Z(Z^*(E)) \cap f(G) = Z(E) \cap f(G)$ which is large in $Z^*(E) \cap f(G) = f(G)$. Therefore, $\mathrm{cl}_G(\ker(f)) = f^{-1}(Z[f(G)])$, which is large in G since the counter-image of a large submodule is also large. But $0 = Z(G)$ implies clearly that $Z(G) \subseteq \ker(f)$. We can conclude that $\mathrm{cl}_G(\ker(f))$ is the maximal essential extension of $\ker(f)$ which is contained in G since $Z(M) \subseteq N$ implies that N is large in $\mathrm{cl}_M(N)$. Consequently, $G = \mathrm{cl}_G(\ker(f))$ since we know that $\mathrm{cl}_G(\ker(f))$ is large in G. This, $\ker(f)$ is large in G, showing that $\ker(f')$ is large in E. Hence, $f \in J(L)$, and $F_3 \subseteq J(L)$. Moreover, since $f \in F_3$ implies that $\ker(f)$ is large in G, we can see that $f(G) \subseteq Z(E)$ since B large in A implies that $A/B = Z(A/B)$. Thus, $\mathrm{Hom}_R(E/Z^*(E), Z^*(E)) = \mathrm{Hom}_R(E/Z^*(E), Z(E))$, and both are contained in J(L).

Consider $\mathrm{End}_R(Z^*(E))$. We have seen that $Z^*(E)$ is injective; so that $J(\mathrm{End}_R(Z^*(E)))$ is the set of those endomorphisms whose kernels are large in $Z^*(E)$. We are considering F_1; so let $f \in F_1$. We identify f with $f' = f + 0 + 0$, as before. Then f has a large kernel if and only if f' has a large kernel. To put it another way, $f \in J(F_1)$ if and only if $f' \in J(L)$. Identifying f and f', we see that $f \in J(F_1)$ if and only if $f \in J(L)$, provided that $f \in F_1$. So $J(F_1) = J(L) \cap F_1$.

Let $f = f_1 + f_2 + f_3 \in L$, where $f_i \in F_i$ and the f_i's are identified canonically and notationally with homomorphisms in L. If $f(x) = 0$ for some x in E, then $f_2(x) = -f_1(x) - f_3(x)$. Since the left side of the previous equation is in $E/Z^*(E)$ and the right side is in $Z^*(E)$, we know that both sides must be zero because $E/Z^*(E)$ and $Z^*(E)$ have zero intersection. This shows that $\ker(f) \subseteq \ker(f_2)$. But if $f_2 \neq 0$, then we know that $\ker(f_2)$ is not large in E since $Z(f_2(E)) = 0$. So $f \in J(L)$ implies that $f = f_1 + f_3$. But since we have shown that $F_3 \subseteq J(L)$, we can see that $f_1 = f - f_3$ is in J(L)—which tells us that $f_1 \in J(L) \cap F_1 = J(F_1)$. Putting all of this together, $J(L) \subseteq J(F_1) \oplus F_3$. The reverse containment is clear since we have shown that both the direct summands on the right are contained in J(L).

Acknowledgements

A portion of this paper represents part of the author's doctoral dissertation, which was written under the direction of Professor Goro Azumaya. The remainder was written with the support of a National Science Foundation grant, which was awarded through the Research Board of Marshall University.

Literature Cited

1. Azumaya, G. 1959. A Duality Theory for Injective Modules. *American Journal of Mathematics* 81:249-78.
2. Faith, C. 1967. Lectures on Injective Modules and Quotient Rings. Springer-Verlag. New York, 140p.
3. Goldie, A. W. 1964. Torsion-Free Modules and Rings. *Journal of Algebra* 1:268-87.
4. Harada, M. 1965. Note on Quasi-Injective Modules. *Osaka Journal of Mathematics* 2:351-56.
5. Johnson, R. E. 1951. Extended Centralizer of a Ring Over a Module. *Proceedings of the American Mathematical Society* 2:891-95.
6. Johnson, R. E., and E. T. Wong. 1959. Self-Injective Rings. *Canadian Mathematical Bulletin* 2:167-173.

An Evaluation of Normalizing Transformations for Skewed Data

Cynthia Deitz and Stanley Wearden

Department of Statistics and Computer Science

West Virginia University, Morgantown, West Virginia 26506

Abstract

A comparison of the empirical power of the t-test, the t-test applied to data transformed by expected normal scores, and the t-test applied to data transformed by random normal deviates has been made. The empirical power curves created in this study show a lack of power of the t-test when applied to small sample sizes from skewed populations. This indicates the necessity of a transformation of the data. The expected normal scores transformation provides a locally more powerful test than does the random normal deviates transformation. Since the difference in power curves is very slight, the simplicity of the random normal deviates transformation justifies its use. Various sample sizes, mean separations, and populations are examined to illustrate these results.

Situations are encountered in data analysis in which parametric tests cannot be appropriately performed because the assumptions required for these tests are not met. This paper is concerned with one such assumption—that the data must be normally distributed. This may not always occur; for example, if the parent population were IQ scores, these scores might follow a skewed distribution if mentally retarded or gifted people appeared in the population.

Another example would be a parent population which would follow an exponential distribution. If parametric tests are performed on data from a nonsymmetric, or skewed, distribution, a significant loss of power of the test occurs. *Power* is the probability of rejecting the null hypothesis when it is actually false.

One solution of the difficulty is to transform the data so that they meet the normality criterion. To determine empirically which degrees of skewness necessitate a transformation, the parametric t-test was performed on sets of raw data

and those results compared to those obtained from the t-test applied to the transformed data. We also examined the effect sample size has on power when there is skewness.

The transformations studied were the expected normal scores (C_1), and ranked random normal deviates (RND). Expected normal scores entail ranking the observations and, to the i^{th} smallest observation, assigning the i^{th} expected normal score, which is obtained from special tables [Fisher and Yates, 4]. These expected values are the means of the random normal deviates which would fall in a certain interval depending on rank of the observation and sample size. A chief disadvantage in using these scores is that to obtain them, tables have to be consulted. Furthermore, the distribution from which they are chosen is not continuous, and this violates another assumption of the parametric tests which could be performed on these scores.

Bell and Doksum [1] suggested that ranked random normal deviates be used instead of expected normal scores. A sample of random normal deviates is generated and ranked, and the observation with the smallest rank receives the smallest random deviate. These deviates are much easier to obtain, and they constitute a continuous distribution. There is the disadvantage that the test statistic applied depends on the random deviates selected, which inserts another factor of randomization or variation in the sample. The amount of variation that several different sets of deviates could introduce and the influence of this possible variation has not been examined empirically. However, a theoretical consideration of this problem may be found in Jogdeo [6].

To investigate and examine the two transformations, many populations were generated with different degrees of skewness. From these populations two equal size samples were drawn. This was done for various sample sizes, and the t-test, the t-test on the substituted expected normal scores, and the t-test on the substituted random normal deviates were performed. The power of each was examined, and comparisons were made to determine how skewness and sample size affect power.

For this study sample sizes of 10, 15, 20, and 25 were selected and population distributions chosen were chi-square distributions of 2, 4, 6, 8, 10, and 12 degrees of freedom. A parent population of 10,000 chi-square variates was generated for each distribution. From this parent population were randomly selected two samples (which will be referred to as X and Y). In order to determine the approximate power of the statistical tests it is necessary that a known difference exist between X and Y. Therefore, constants were generated so that $Pr(X > Y) =$ .5, .4, .3, .2, and .1. Constants were generated in this fashion to conform with a study done by Frush [5]. The method of generating these constants was developed by Wang [8]. The appropriate constant was then added to each element of the Y sample creating a separation between X and Y means, and then the three statistical tests were performed. The null hypothesis was that the X and Y populations were identical, and the alternative hypothesis was that they were identical except for a shift in location. For each test was counted the number of times the null hypothesis was rejected.

The procedure described above was followed 5000 times for each constant within each sample size for each distribution. The number of times the null hypothesis was rejected appears in Table 1. The resulting power curves, the constants generated, and the skewness associated with each population can be found in Deitz [3]. In discussing the results and conclusions C_1' will refer to the

Table 1.

		$Pr(X > Y) = .5 = \delta_1$				$Pr(X > Y) = .4 = \delta_2$			
		Sample Size				Sample Size			
		10	15	20	25	10	15	20	25
		No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.
Dist.	Test								
χ^2_2	T	245	214	273	252	427	501	603	634
	C_1'	257	245	280	265	636	858	1105	1286
	RND'	255	244	276	255	621	807	1115	1280
χ^2_4	T	221	232	234	230	457	636	753	1024
	C_1'	238	260	245	252	570	809	1026	1246
	RND'	236	259	258	252	536	784	1005	1234
χ^2_6	T	211	234	252	240	505	680	808	1010
	C_1'	237	247	257	255	584	804	978	1221
	RND'	237	252	257	255	568	757	958	1196
χ^2_8	T	243	237	265	283	543	723	925	1077
	C_1'	253	241	261	293	597	831	1063	1207
	RND'	255	236	269	285	557	812	1043	1176
χ^2_{10}	T	236	252	267	265	592	741	889	1106
	C_1'	249	256	282	270	631	787	975	1223
	RND'	247	251	282	270	604	781	956	1194
χ^2_{12}	T	221	269	243	270	541	751	898	1119
	C_1'	231	265	262	266	576	805	957	1189
	RND'	242	253	258	275	548	799	947	1190

Table 1.–*Continued*

Dist.	Test	Pr(X > Y) = .3 = δ_3 Sample Size				Pr(X > Y) = .2 = δ_4 Sample Size			
		No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.	No. Times H_0 Rej.
χ^2_2	T	1196	1574	1971	2260	2808	3540	4029	4368
	C'_1	1770	2518	3124	3653	3383	4255	4668	4866
	RND'	1730	2475	3071	3590	3294	4185	4604	4839
χ^2_4	T	1343	2004	2484	2900	3026	3892	4296	4693
	C'_1	1620	2510	3143	3694	3338	4262	4705	4905
	RND'	1583	2465	3080	3632	3235	4197	4659	4896
χ^2_6	T	1563	2136	2688	3151	3262	4034	4552	4798
	C'_1	1741	2458	3162	3635	3460	4295	4739	4888
	RND'	1681	2308	3089	3581	3392	4247	4703	4900
χ^2_8	T	1572	2265	2788	3245	3366	4177	4602	4787
	C'_1	1721	2521	3150	3597	3480	4329	4740	4877
	RND'	1636	2455	3049	3562	3349	4237	4710	4856
χ^2_{10}	T	1724	2220	2816	3440	3450	4207	4642	4860
	C'_1	1790	2409	3054	3705	3543	4350	4736	4918
	RND'	1733	2363	3033	3685	3429	4278	4698	4903
χ^2_{12}	T	1638	2391	2899	3407	3434	4263	4701	4868
	C'_1	1730	2512	3107	3659	3521	4359	4754	4922
	RND'	1670	2451	3059	3621	3398	4276	4734	4915

Table 1.—*Continued*

$Pr(X > Y) = .1 = \delta_5$

Dist.	Test	Sample Size 10 No. Times H_0 Rej.	15 No. Times H_0 Rej.	20 No. Times H_0 Rej.	25 No. Times H_0 Rej.
χ^2_2	T	4551	4871	4950	4985
	C'_1	4639	4952	4996	5000
	RND'	4576	4925	4991	4998
χ^2_4	T	4622	4917	4983	4995
	C'_1	4667	4961	4995	5000
	RND'	4608	4945	4991	5000
χ^2_6	T	4730	4949	4992	4999
	C'_1	4725	4958	4998	5000
	RND'	4657	4949	4995	4999
χ^2_8	T	4765	4957	4992	4994
	C'	4773	4973	4999	4999
	RND'	4728	4964	4998	4999
χ^2_{10}	T	4789	4974	4999	5000
	C'_1	4783	4986	4998	5000
	RND'	4729	4975	4997	5000
χ^2_{12}	T	4816	4969	4996	5000
	C'_1	4803	4981	4998	5000
	RND'	4749	4966	4998	5000

t-test performed on the expected normal scores and RND′ will refer to the t-test performed on the random normal deviates.

To determine if one test was significantly more powerful than another the sign test was performed on the powers of the three tests in two ways. Tests were compared two at a time by counting the number of times the power of one exceeded the power of the other. This was done by holding sample size constant and comparing the four mean separations, as well as by holding the mean separations constant and comparing the four sample sizes. In each case, one test would be significantly more powerful than the other only if the power of the first exceeded that of the second in all four comparisons, or in terms of power curves, if one curve was entirely above the other. Graphs of the power curves and tables of the sign test may be found in Deitz [3].

The most extreme situation examined was the $X^2_{(2)}$ which has a degree of skewness of 1.9996. As was expected because of the skewness of this distribution, not only were C_1' and RND′ significantly more powerful than t but these power differences were largest for this parent population. Since this population is farthest from the normal of those studied, its results indicate the relative power of the three tests in the most extreme empirical situation. These results support the previous assertions that the power of parametric tests such as t are low when the symmetry aspect of normal assumptions is violated. It also shows the need for "normalizing" transformations since their power was shown to be greater than that of t. For this degree of skewness, there is another indication of greater power of the two tests performed on the transformed data in that their curves plateau (or approach 1) sooner and then level off.

Examination of the distributions which were the last skewed $X^2_{(10),(12)}$, demonstrated that at best the t-test was as good as the C_1' and RND′ which supports the contention that the asymptotic relative efficiency of C_1' with respect to t is greater than or equal to 1 [Chernoff and Savage, 2].

The other distributions studied ($X^2_{(4),(6),(8)}$) were not as extremely skewed as $X^2_{(2)}$, and did not approach normality as closely as $X^2_{(10),(12)}$. Therefore, our main interest lies in the results obtained from these distributions.

The largest and smallest mean separation ($Pr(X > Y) = .1$ and .4) exhibit no real difference in power among the three statistical tests. This could be attributed to the fact that (1) for small mean separations, it is difficult to distinguish between two samples; and (2) large mean separations are easier to detect irrespective of assumptions This means there is little opportunity to affect low or high power situations.

The other mean separations studied ($Pr(X > Y) = .3$ and .2) are important for two reasons: (1) they exhibit no extreme conditions and (2) relatively little research has been undertaken with regard to these conditions. It is hardest to detect a power difference among the three tests for the smallest sample size (n = 10), but this could be attributed to the fact that there are not enough data available. As sample size increases within these two probability conditions, the difference in power of the t-test and that performed on the transformed data is greatest. Even though the sign test indicates C_1' is significantly more powerful than RND′, this difference is so small (and continues to decrease as sample size increases) that it is considered, for all practical purposes, unimportant. When sample size reached 25, all three tests demonstrated large powers and small power differences. This was expected for large sample size increases the power of most tests.

It can be concluded that C_1' and RND′ are significantly more powerful than t

for skewed distributions, which supports the assertion that transformation is desirable for data from a skewed distribution. These results would tend to indicate that data with a degree of skewness ≥ 1.0 should be transformed by expected normal scores or random normal deviates. Even though C_1' is significantly more powerful than RND′ in several cases, the difference in power is so small that the two tests may be considered equal. Thus, the more convenient RND′ is recommended. The RND′ test is much easier to perform since tables do not have to be used to transform the data, and normal deviates can be generated by computers with minimum effort.

Certain questions were raised by the methods employed in this study. One such question is that a t-test was performed on the transformed data instead of the normal scores test as defined by Terry [7], but results were discussed as if the conventional C_1-test had been performed. This can be justified by the following reasoning: The t-test applied to data which have been transformed by expected normal scores is a strictly increasing function of the normal scores C_1-test. This implies that if the null hypothesis is rejected by performing a t-test, it would also have been rejected if the C_1-test had been performed.

Another important question to consider in evaluating the results of this study is how close the empirical power of the tests is to the true power. The probability of approximating the true power of the statistical test within a given bound and for a given number of times the test was performed can be determined in the following manner. For a given test procedure assign a value of 1 if the null hypothesis was accepted, thus producing a Bernouli variable. Therefore, how close X/N (N = number of times the test was performed) is to the true probability, within some interval, is given by

$$\Pr\left(-\epsilon \leq \frac{X - Np}{\sqrt{Np(1-p)}} \leq \epsilon\right) \sim N(0,1)$$

$$\leq \Pr\left(-b \leq \frac{X}{N} - p \leq b\right) \quad \text{where} \quad b = \frac{\epsilon}{2\sqrt{N}}.$$

Solving for ϵ gives $\epsilon = 2b\sqrt{N}$. To evaluate the above interval and thereby obtain the desired probability, N and b must be specified. For our study N was 5000. A b of 0.013 yields a probability of 0.9426. That is, $\Pr(-0.013 \leq \frac{X}{N} - p \leq 0.013) \geq .9426$ or the probability of the power obtained from empirical results is within ± 0.013 of the true power of the test. The cases of interest exhibited power differences greater than 0.013.

Finally, we mention the work of Jogdeo [6], who has shown that the power of the Bell and Doksum Randomized Rank Scores procedures remains bounded away from unity for the entire parameter space of the alternative hypothesis. That is, the effect of the superimposed noise created by the additional sample which was employed in the test creates undesirable properties of the test. Our simulation has shown this concern to be of no practical consequence, for the empirical power of the RND′ test is very close to 1 in several cases and is equal to 1 in three cases.

Literature Cited

1. Bell, C. B. and K. A. Doksum. 1964. Some New Distribution Free Statistics. *Annals of Mathematical Statistics* 36:203-14.

2. Chernoff, Herman and I. Richard Savage. 1958. Asymptotic Normality and Efficiency of Certain Non-Parametric Test Statistics. *Annals of Mathematical Statistics* 29:972-94.
3. Deitz, Cynthia A. 1973. An Evaluation of the Bell-Doksum Procedure for Small Samples from a Skewed Distribution. Unpublished Master's Report, West Virginia University.
4. Fisher, Ronald A. and Frank Yates. 1963, 6th edition. Statistical Tables. Oliver and Boyd, London, 146 p.
5. Frush, William Harold. 1969. A Comparison of Power of Student's t-test, Mann-Whitney's U-test, and the Fisher-Yates test for Different Degrees of Skewness and Sample Size. Unpublished Master's Report, West Virginia University.
6. Jogdeo, Kumar. 1966. On Randomized Rank Score Procedures of Bell and Doksum. *Annals of Mathematical Statistics* 37:1697-1703.
7. Terry, Milton E. 1952. Some Rank Order Tests Which Are Most Powerful Against Specific Parametric Alternatives. *Annals of Mathematical Statistics* 23:346-66.
8. Wang, Ying Y. 1967. A Comparison of Several Variance Component Estimates. *Biometrika* 54:301-5.

Open Sets From Disjoint Closed Intervals

Michael Mays
John Schleusner
William Simons
Department of Mathematics
West Virginia University, Morgantown, West Virginia 26506

Abstract

In this paper we show by several different arguments that no open subset of the real line can be written as a union of disjoint closed subintervals unless uncountably many of the subintervals contain only one point.

Most texts concerned with general topology contain the theorem which states that each open subset of R, the real line, is the union of disjoint open intervals. Nevertheless, it seems difficult to find any reference for conditions under which an open subset of R can be the union of disjoint closed intervals. In this note we present such conditions.

If any open set $O \subset R$ is the union of disjoint closed intervals, then so is each component of O. Conversely, if any open interval is the union of disjoint closed intervals, then so is every open subset. Hence for our purposes, it suffices to consider the open interval (0,1).

Let $J_A = \{I_\alpha \mid I_\alpha = [a_\alpha, b_\alpha] \subset (0,1), \alpha \in A\}$ be a disjoint family of closed intervals. We will write $I_\alpha < I_\beta$ to mean I_α is to the left of I_β on R.

Theorem 1. If each $I_\alpha \in J_A$ *contains at least two points, then* $\bigcup_{\alpha \in A} I_\alpha \subsetneq (0,1)$.

Theorem 2. If $\bigcup_{\alpha \in A} I\alpha = (0,1)$, *then uncountably many of the* I_α *contain only one point.*

To prove the first theorem we assume $\bigcup_{\alpha \in A} I_\alpha = (0,1)$. Since (0,1) is open A must be infinite, and since each I_α contains a rational A must be countable. Thus we can write $(0,1) = \bigcup_{n=1}^{\infty} I_n$. We will generate a sequence from this family of intervals by first chosing any two, I_{n_1} and I_{n_2}, with $I_{n_1} < I_{n_2}$ and $a_{n_2} - b_{n_1} <$ ½. Having chosen $I_{n_1}, \ldots, I_{n_{2k}}$, we choose $I_{n_{2k+1}}$, and $I_{n_{2k+2}}$ so that $I_{n_{2k-1}} < I_{n_{2k+1}} < I_{n_{2k+2}} < I_{n_{2k}}$ and $a_{n_{2k}} - b_{n_{2k+1}} < 1/(2k+1)$, $a_{n_{2k+2}} \quad b_{n_{2k+1}} < 1/(2k+2)$. The sequences of endpoints, $\{b_{n_{2k-1}}\}_{k=1}^{\infty}$ and $\{a_{n_{2k}}\}_{k=1}^{\infty}$, are monotone increasing and decreasing, respectively, and both converge to the same point, which we denote by c. Since $c \in (0,1)$, there is some interval $I_N = [a_N, b_N] \in J_A$ containing c. However, if $a_N \leqslant c < b_N$ then $I_N \cap I_{n_{2k}} \neq \phi$ for k sufficiently large, and if $c = b_N$ then $I_N \cap I_{n_{2k-1}} \neq \phi$ for large k. Therefore the members of J_A are not disjoint, and hence the theorem is proved.

We now turn to the proof of Theorem 2. Let the index set $A = S \cup T$, where S is countable, and suppose $I_\sigma = [a_\sigma, a_\sigma]$ for $\sigma \in S$, and each I_τ, $\tau \in T$, has at least two points. If $\{a_\sigma | \sigma \in S\}$ is not dense in (0,1) then $(0,1) - \{a_\sigma | \sigma \in S\}$ contains an open interval which, by Theorem 1, cannot be covered by the I_τ, $\tau \in T$. However if $\{a_\sigma | \sigma \in S\}$ is dense in (0,1), then for each $\tau \in T$ there is some $\sigma \in S$ such that $a_\sigma \in I_\tau$. Hence $I_\tau \cap I_\sigma \neq \phi$ and the proof of Theorem 2 is complete.

Another interesting proof of Theorem I proceeds as follows: Again, assuming that each closed subinterval contains at least 2 points, suppose $(0,1) = \bigcup_{n=1}^{\infty} I_n$. Note that each x in (0,1) is contained in a unique $[a_x, b_x]$ and define the function f mapping (0,1) to (0,1) by $f(x) = a_x$. The f so defined is continuous because if $f((0,a_x)) \subseteq (0, a_x - \epsilon)$, then each point in $(0,a_x)$ is contained in a closed subinterval with left hand endpoint $\leqslant a_x - \epsilon$. But $(a_x - \epsilon, a_x)$ must contain one of the I_n and for x in this $I_n = (a_n, b_n)$, $f(x) = a_x > a_x - \epsilon$, which is a contradiction. Now all that needs to be observed is that (0,1) is connected, while f((0,1)) is countable and hence disconnected, which implies that f cannot be continuous. This contradiction establishes Theorem I.

On the Maximum Modulus for Meromorphic Univalent Functions

James Miller
Department of Mathematics
West Virginia University, Morgantown, West Virginia 26506

Abstract

We use a variational formula for a class of univalent meromorphic functions to determine the form of the extremal function that maximizes the modulus in that class.

Introduction

Let S(p) denote the class of all functions $f(z) = z + a_2 z^2 + \ldots$ that are analytic, except for a pole at $z = p$, and univalent in the unit disc E. Fenchel [1] obtained what Komatu [3] called a "distortion theorem", that is, the following inequalities:

$$(1) \qquad |f(z)| \geq \frac{p|z|}{p + (1+p^2)\,|z| + p|z|^2}, \quad |z| < 1$$

and, if $|a_2| > 2$,

$$(2) \qquad |f(z)| \leq \frac{|z|}{-1 + |a_2|\,|z| - |z|^2},$$

which holds for $\frac{1}{2}|a_2| - \sqrt{\left(\frac{|a_2|^2}{4} - 1\right)} < |z| < 1.$

The inequality (1) is attained by the function

$$f_0(z) = \frac{pz}{p - (p^2 + 1)\,z + pz^2}$$

However, neither the actual maximum nor the form of the extremal function, related to (2), are known.

In this paper we take a first step to complete this study by using a variational formula for the class S(p) to determine the form of the function in S(p) which maximizes the functional $F(f) = |f(z_0)|$. The variational formula was developed in [4].

Extremal function for $|f(z_0)| = \max$.

Let z_0 be fixed in $|z| < 1$ with $z_0 \neq 0, p$. We wish to find the function in S(p) which maximizes the functional $|f(z_0)|$. It is more convenient to consider the functional $\log |f(z_0)|$ instead of the functional $|f(z_0)|$. It is evident that a function in S(p) which is extremal for the latter is also extremal for the former and conversely. Let $f \in S(p)$ be the extremal function for $\log |f(z_0)| = \max$. Since $\log |f(z_0)|$ is linear in the small (see [4]) on S(p), we may apply Theorems 1 and 2 [4] to obtain

Theorem 1. Let $z_0 \in E$, $z_0 \neq 0, p$. If f is a solution to the extremal problem $|f(z_0)| = \max$. over the class S(p), then f satisfies the following differential equation

$$P(z) = \frac{-f(z_0)}{f(z_0) - f(z)} \left(\frac{z f'(z)}{f(z)} \right)^2 \tag{3}$$

where

$$P(z) = \frac{p}{z - p} + \frac{pz}{1 - pz} + \left(\frac{z_0}{z - z_0} - \frac{p}{z - p} \right) A + \left(\frac{\bar{z}_0 z}{1 - \bar{z}_0 z} - \frac{pz}{1 - pz} \right) \bar{A} \tag{4}$$

where $A = \dfrac{z_0 f'(z_0)}{f(z_0)}$. The domain D that is the image of E under the extremal function f is bounded by a finite number of analytic arcs and has no exterior points. Moreover, the function P(z) is non-negative on $|z| = 1$ and $P(z) = P(\frac{1}{z})$.

From (4) we see that P(z) can be written in the form

$$P(z) = \frac{pz^4 - (\text{lower order terms in } z)}{(z - z_0)(z - p)(1 - z_0)(1 - pz)} .$$

so that P(z) has 4 zeros in the z-plane. By (3) we see that P(z) has no zeros in E and therefore no zeros in $|z| > 1$. Since P(z) is non-negative on $|z| = 1$ and f maps $|z| = 1$ onto slits in the finite plane we conclude that P(z) has 2 double zeros $e^{i\alpha}$ and $e^{i\beta}$ on $|z| = 1$ and that f maps E onto a domain bounded by a single analytic arc with end points $f(e^{i\alpha})$ and $f(e^{i\beta})$. Hence we can write p(z) in the form

$$P(z) = \frac{p(z - e^{i\alpha})^2 (z - e^{i\beta})^2}{(z - z_0)(z - p)(1 - z_0 z)(1 - pz)} .$$

If we set $f(z) = w$ and $f(z_0) = w_0$ in (3), then we obtain

$$P(z) = \frac{-w_0}{w_0 - w} \left(\frac{z}{w} \frac{dw}{dz} \right)^2 \tag{5}$$

The left hand side of (5) in non-negative on $|z| = 1$ and the zeros and poles of $\frac{1}{z^2} P(z)$ are 0, p, z_0, $e^{i\alpha} e^{i\beta} \frac{1}{z_0}$ and $\frac{1}{p}$.

Let

$$v(z) = \int_p^z \sqrt{-P(z)} \, \frac{dz}{z} \tag{6}$$

and

$$v(w) = \int_\infty^w \left(\frac{w_0}{w_0 - w} \right)^{1/2} \frac{dw}{w} . \tag{7}$$

Then $v(z)$ and $v(w)$ are analytic in E and D respectively except at the zeros and poles of $\frac{1}{z^2} P(z)$. Integrating (7) gives us

$$v(w) = \log \frac{\sqrt{w_0 - w} - w_0}{\sqrt{w_0 - w} + w_0} ,$$

and therefore

$$w = \frac{-4 e^v w_0}{(e^v - 1)^2} = - w_0 \operatorname{csch}^2 \frac{v}{2} . \tag{8}$$

Using (6) we have w as a function of z for all $z \in E$ except for the branch points. By examining the mapping in (6) we show that the singularities of (8) may be removed.

We now state and prove the main result.

Theorem 2. Let $z_0 \in E$. Let f be the function in S(p) which maximizes the functional $|g(z_0)|$ for $g \in S(p)$. Then

$$f(z) = \frac{f_1(z) - f_1(0)}{f_1'(0)}$$

where

$$f_1(z) = -w_0 \operatorname{csch}^2 \frac{v}{2}$$

and v is given by (6) and (7).

Proof

We prove *Theorem 2* by examining the mapping properties of (6) and (7). Following the notation in [2] we see that $P(z)(\frac{dz}{z})^2$ is a positive quadratic differential on E. A trajectory of $P(z)(\frac{dz}{z})^2$ is a maximal regular curve on which $P(z)(\frac{dz}{z})^2 > 0$. An orthogonal trajectory is a maximal regular curve on which $P(z)(\frac{dz}{z})^2 < 0$.

The origin of the z-plane is a double pole of $\frac{1}{z^2}P(z)$. Thus from Theorem 3.4 [2] we conclude that in a neighborhood of $z = 0$ and for each ray emanating from $z = 0$, there is a trajectory of $P(z)(\frac{dz}{z})^2$ which has that ray as the limiting tangent. The structure is shown in Figure 1. The structure near $z = 0$ together

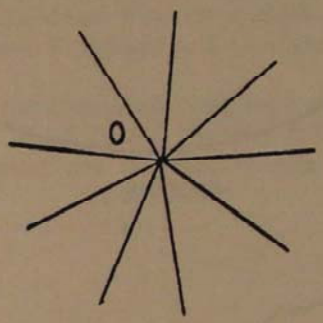


FIGURE 1.

with the Basic Structure Theorem [2] allows us to describe the structure of the trajectories that connect the zeros and poles of $\frac{1}{z^2}P(z)$ in $|z| \leqslant 1$ which are $e^{i\alpha}$, $e^{i\beta}$ and 0, z_0, p respectively. This structure is shown in Figure 2. The structure of the trajectories in the w-plane is shown in Figure 3. We see that the functions in (6) and (7) map E and D respectively onto two infinite strips as in Figure 4, where $e^{i\alpha}$ and $f(e^{i\alpha})$ each map onto α' and α'', and $e^{i\beta}$ and $f(e^{i\beta})$ each map onto β' and β'' as in Figure 4. Since the segment s_1 between α' and β' and the segment s_2 between α'' and β'' correspond to the trajectory connecting $f(e^{i\alpha})$ and $f(e^{i\beta})$, we see that the lengths of s_1 and s_2 are equal. In the w-plane we see that there is exactly one orthogonal trajectory with limiting point at $f(z_0)$. The other limiting end point is at infinity. Thus we can match the points α' and β' with α'' and β'' respectively. At the same time we see that p' and z_0' (image of p and z_0) differ by an imaginary number. Thus we can "glue" the two strips together as in Figure 5 where there is a cut along the horizontal line between α' and β'. Using the mapping in (6) we have $p' = 0$ and $z_0' = ic$ where $c > 0$. Therefore the mapping given by (8) will be univalent in E and map E onto the whole plane minus a slit if and only if $c = \pi$. Thus the function

$$f_1(z) = -w_0 \operatorname{csch}^2\{\frac{1}{2}\int_p^z (-P(z))^{1/2}\frac{dz}{z}\}$$

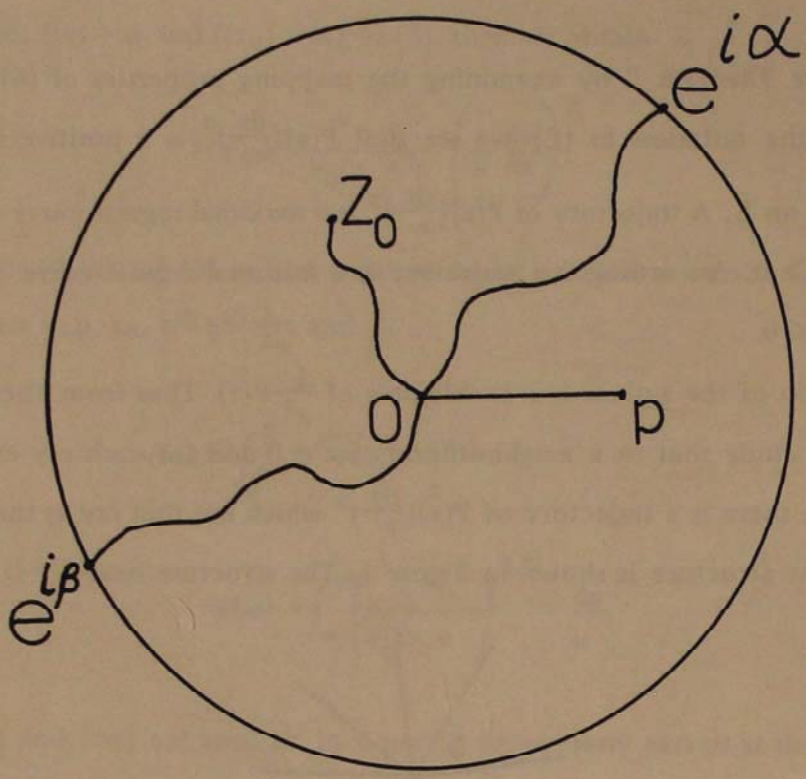


FIGURE 2.

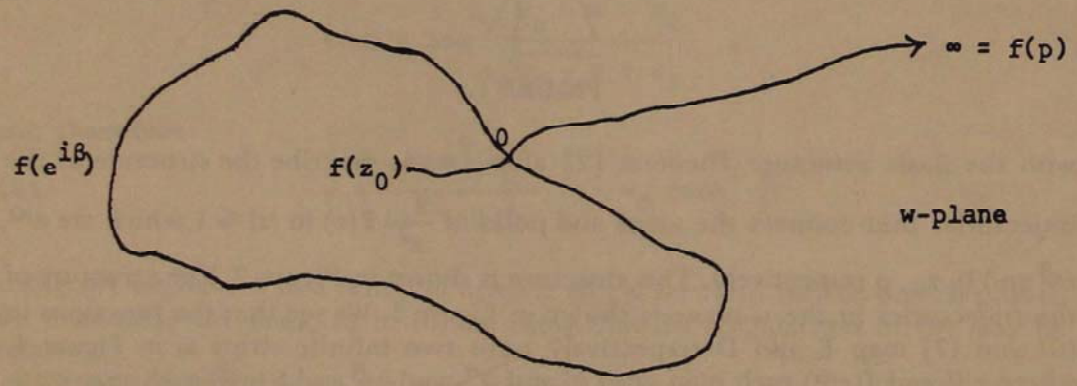


FIGURE 3.

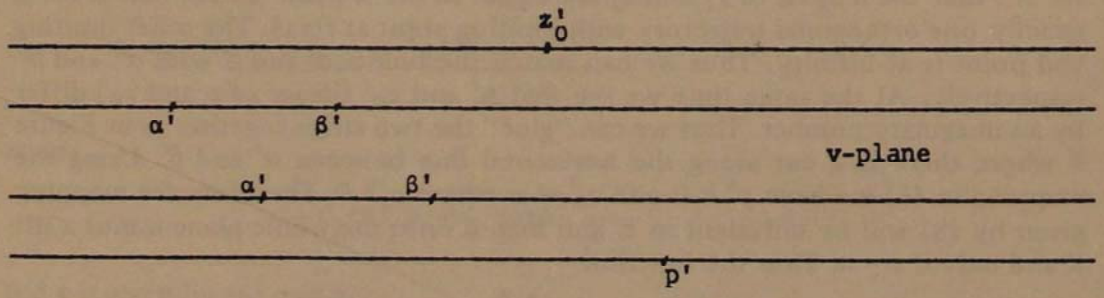


FIGURE 4.

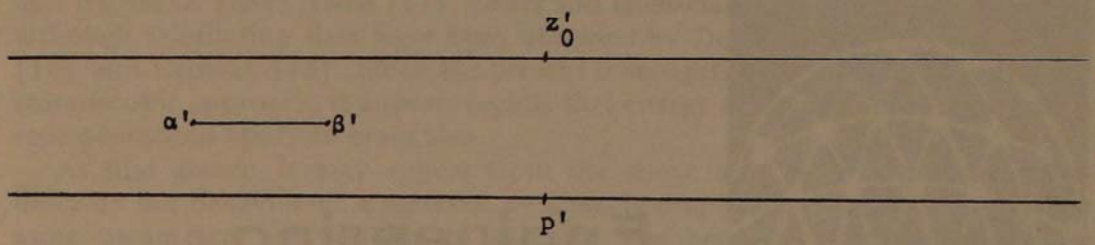


FIGURE 5.

will map the unit disk onto a domain bounded by a single slit of finite length. Therefore we have found the form of $f(z)$ such that f maximizes $|f(z_0)|$ in the class $S(p)$.

The author wishes to acknowledge the valuable suggestions made by the referee.

Literature Cited

1. Fenchel, W. 1931. Bemerkungen über die in Einheitskreis meromorphen schlichten Funktionen. Sitzungsber. *Preuss. Akad. Wiss. Phys.-Math. Kl. Heft* 22/23:431-36.
2. Jenkins, J. A. 1958. Univalent Functions and Conformal Mappings. Springer-Verlag. Berlin.
3. Komatu, Y. 1949. Note on the theory of conformal representation by meromorphic functions, I and II. *Proc. Imp. Acad. Tokyo.* 21:269-77; 278-84. Imprinted 1945.
4. Miller, J. 1971. Extremal functions for meromorphic univalent functions. *Journal d'Analyse Math.* 24:77-86.

Engineering

Constitutive Properties of Brain Tissue

M. R. Pamidi and S. H. Advani
Department of Mechanical Engineering and Mechanics
West Virginia University, Morgantown, West Virginia 26506

Abstract

Biological tissues, in general, exhibit non-linear viscoelastic behavior. This paper is concerned with the formulation of constitutive relations governing brain tissue. Essentially two methods for describing brain tissue properties are presented. First, the existing strain energy representations for finite deformations of elastic materials are modified to include time dependent material response. Second, the classical Boltzmann Integral form is refined by including a cubic strain functional in the kernel. In both cases, the material relaxation function is assumed to have the conventional exponential form. The mathematical model predictions are compared to experimentally determine *in vitro* properties of human brain tissue.

Introduction

Numerous studies have been devoted in recent times to the mechanical and acoustical characterization of the properties of biological tissues. Motivated by the fact that the response of materials can be deduced from a potential function (strain energy function), many studies have been directed at the derivation of constitutive relations from this quantity. Examples of these studies include research by Fung [7,8], Hildebrandt *et al* [14], Gottenberg *et al* [10], Smith [26], Patel and Vaishnav [19], Gou [11], Simon *et al* [25], Demiray [4], Haut and Little [13], Blatz *et al* [3] and Snyder [27]. A perusal of these works reveals, however, that most of the theories developed involve application of finite elasticity theory using strain invariants rather than a viscoelastic representation which should include the strains, their time history and possibly strain rates. In addition, even the strain energy forms that have been established have been tested for a specific loading pattern.

This work is based on the idea of the existence of a mechanical equation of state for a continuum relating deformation or its time rate, stress and time for isothermal processes. Such an idea for metals has been advanced by Vorotnikov

and Novinskii [28], Nadai [17], Zener and Holloman [30] and Holloman [15], although conflicting data have been obtained by Dorn *et al* [5], Johnson *et al* [16] and Orowan [18]. Since the present investigation is based on a continuum macroscopic approach, it appears logical that energy and polynomial constitutive representations apply to brain also.

At first glance, it may appear from the above arguments that attempts to describe the behavior of a material by a single simple equation of state with the same parameters are unlikely to succeed. The behavior of a material under different types of tests can be described by a generalized equation, but the parameters of this equation will have different numerical values for each individual type of test, reflecting the particular physical conditions prevailing during stress relaxation, creep, uniaxial strain, etc. However, in the sequel, a strain energy form is established which correlates uniaxial relaxation tests, constant strain-rate tests and dynamic torsion tests within reasonable limits.

Governing Equations

From the point of view of an engineering material, brain tissue is assumed to be isotropic, homogeneous and incompressible. Isotropy is based on microscopic studies [24] and homogeneity on the fact that the test specimens consist mostly of white or grey matters. Because of its high bulk modulus of 305,300 psi [29] and low shear modulus [24], incompressibility is justified. With these preliminaries, we can now formulate constitutive relations by two different methods.

(a) Uniaxial Tests

i) Strain Energy Approach: The conventional constitutive relation $\sigma ij = \partial W/\partial \in ij$ must be modified in view of the inclusion of strain rates in W. Consider, then, a continuum with kinetic energy density $W = W(q_K, \dot{q}_K)$ strain energy density $W = W(q_k \dot{q}_k)$ and external work $W_D = t_k q_k$, where q_k and $\dot{q}_k$ are generalized coordinates, t_K the generalized forces and ρ the continuum density. The Lagrangian is $L=T-(W-W_D)$ and Hamilton's principle reads

$$\int_{t_0}^{t_1} \delta L \, dV \, dt = \delta \int_{t_0}^{t_1} [T-(W-W_D)] \, dV \, dt = 0$$

Consequently, Euler-Lagrange equation yields

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_K} - \frac{\partial L}{\partial q_K} = 0$$

where $$L = \rho \dot{q}_K^2/2 - W(q_K, \dot{q}_K) + t_K q_K$$

Thus, in the absence of inertia forces, one finally obtains

$$t_K = \frac{\partial W}{\partial q_K} - \frac{d}{dt}\frac{\partial W}{\partial \dot{q}_K}$$

If the generalized coordinates are strains and their time rates, we have the 'engineering' stresses as

$$t_{ij} = \frac{\partial W}{\partial \epsilon_{ij}} - \frac{d}{dt}\frac{\partial W}{\partial \dot{\epsilon}_{ij}}$$

where $\in ij$ are the strains.
If the rate of energy, or power Q, is employed, we have

$$t_{ij} = \frac{\partial Q}{\partial \dot{\epsilon}_{ij}} - \frac{d}{dt}\frac{\partial Q}{\partial \ddot{\epsilon}_{ij}} \qquad (2\text{-}1)$$

Polynomial forms for the strain energy W as a function of the invariants I_i, i = 1, 2, 3, have been used in the representations by Mooney, Gou, Blatz *et al*, Demiray and Snyder. The invariants for an orthogonal deformation field are

$$\begin{aligned} I_1 &= \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \\ I_2 &= \lambda_1^2\lambda_2^2 + \lambda_2^2\lambda_3^2 + \lambda_3^2\lambda_1^2 \qquad (2\text{-}2) \\ I_3 &= \lambda_1^2\lambda_2^2\lambda_3^2 \end{aligned}$$

where λ_i 's are the principal stretches and $I_3 = 1$ for isochoric deformations. Because of its good agreement with certain isochoric deformations, we choose the Mooney form

$$W = C_1(I_1 - 3) + C_2(I_2 - 3)$$

and incorporate viscoelastic effects therein by defining the material parameters as

$$C_1 = \sum_{m=0}^{M} A_m \exp(-\lambda_m t) \; ; \quad C_2 = \sum_{n=0}^{N} B_n \exp(-\mu_n t)$$

where λ_m and μ_n are time constants. The power is

$$Q = \dot{W} = C_1\dot{I}_1 + C_2\dot{I}_2 - \sum_m \sum_n \left(A_m \lambda_m I_1 e^{-\lambda_m t} + B_n \mu_n I_2 e^{-\mu_n t}\right)$$

For the constant strain-rate tests of Estes [6], $\ell = \ell_o - vt$ or $t = (\lambda - 1)/\dot{\lambda}$, where ℓ_o and ℓ are the initial and instantaneous (at time t) lengths of the specimen, V the ram velocity and λ the stretch, ℓ/ℓ_o. The only non-zero 'true' stress is

$$\sigma_{11} = (\partial Q / \partial \dot{\epsilon}_{11})\,\lambda^2$$

which, after substitution and reduction, yields

$$\sigma_{11}/2A_o = (\lambda^2-1/\lambda)(1+B_o/A_o\lambda) + e^{\lambda_1(1-\lambda)/\dot\lambda}\left[(\lambda^2-1/\lambda)(A_1/A_o + B_1\exp((\lambda-1)(\lambda_1-\mu_1)/\dot\lambda)/A_o\lambda) + A_1\lambda_1(\lambda-1)\{2\dot\lambda(\lambda^2-1/\lambda) - \lambda_1(\lambda^3-3\lambda+2)\}/2A_o\dot\lambda^2 + B_1\exp((\lambda-1)(\lambda_1-\mu_1)/\dot\lambda)\mu_1\cdot(\lambda-1)\{2\dot\lambda(\lambda-1/\lambda^2) - \mu(2\lambda^2-3\lambda+1/\lambda)\}/2A_o\dot\lambda^2\right] \quad (2\text{-}3)$$

where it has been assumed for simplicity M = N = 1, $\lambda_o = \mu_o = 0$. The elastic solution [23] is obtained with $\lambda_1 = \mu_1 = 0$ and $A_1 = B_1 = 0$.

ii) Integral Approach: Following Ashton [2], Halpin [12], Schapery [20,21,22] and Fund [8], we modify the conventional Boltzmann integral form and propose the following representation:

$$\sigma_{ij}(t) = \int_0^t G(t-\tau)\,\frac{\partial f[\epsilon_{ij}(\tau)]}{\partial\tau}\,d\tau \quad (2\text{-}4)$$

and assume

$$G(t) = G(\infty) + \sum_{n=1}^{N} G_n \exp^{-\lambda_n t}$$
$$f[\epsilon_{ij}] = C_1\epsilon_{11} + C_2\epsilon_{11}^3 \quad (2\text{-}5)$$

It is noteworthy that in Eq. $(2.5)_2$ even powers in ϵ_{11} are not chosen. A quadratic form of the type $f(\epsilon) = C_1 + C_2\epsilon^2$ has been assumed by Haut and Little [13]. However it would appear to violate the requirement that the resulting strain energy be positive definite for all strains. After substitution, integration and simplication, Eq. (2-4) can be written as

$$\sigma_{11}(t)/2C_1\,G(\infty) = \gamma + \gamma^2/2 + 3C_2(4\gamma^3/3 + 2\gamma^4 + \gamma^5 + \gamma^6/6)/C_1 + \dot\lambda\sum_n G_n\Big[\gamma + (1-\delta)(1-e^{-\gamma/\delta}) + 3C_2\{4(\gamma^2-2\gamma\delta+2\delta^2(1-e^{-\gamma/\delta})) + 8(\gamma^3 - 3\gamma^2\delta + 6\delta^2\gamma - 6\delta^3(1-e^{-\gamma/\delta})) + 5(\gamma^4 - 4\delta\gamma^3 + 12\delta^2\gamma^2 - 24\delta^3\gamma + 24\delta^4(1-e^{-\gamma/\delta})) + \gamma^5 - 5\delta\gamma^4 + 20\delta^2\gamma^3 - 60\delta^3\gamma^2 + 120\delta^4\gamma - 120\delta^5(1-e^{-\gamma/\delta})\}/C_1\Big]\cdot/G(\infty)\lambda_n \quad (2\text{-}6)$$

where $\delta = \dot\lambda/\lambda_n$, $\gamma = \lambda - 1$.

In what follows, N = 1. Equations (2-3) and (2-6) will be used in the sequel to predict the behavior of brain tissue under a constant strain-rate loading.

(b) Torsion Tests. The complex shear modulus is $E(\omega) = E_1(\omega) + iE_2(\omega)$ $= i\omega \int_0^t E(t)e^{-i\omega t}dt$ where $E_1(\omega)$ and $E_2(\omega)$, the storage and loss moduli, are given by

$$E_1(\omega) = \sum_{n=0}^{N} \omega^2 E_n/(\omega^2 + \lambda_n^2); \quad E_2(\omega) = \sum_{n=0}^{N} \omega \lambda_n E_n/(\omega^2 + \lambda_n^2) \quad (2\text{-}7)$$

and $E(t) = E_o + \sum_{n=1}^{N} E_n e^{-\lambda_n t}$ is the shear relaxation modulus and equals $G(t)/3$ for incompressible materials. Eq. (2-7), with N = 1 and $\lambda_o = 0$, will be used to predict the response to dynamic torsion tests.

Experimental and Analytical Results

Uniaxial relaxation and constant strain-rate tests have been reported by Galford [9] and Estes [6], respectively. Figure 1 shows the relaxation curve along with the fit obtained using Eq. $(2\text{-}5)_1$ with N = 1. The final correlation is

$$\frac{G(t)}{G(\infty)} = 1 + 1.80\, e^{-0.1t} \quad (3.1)$$

Figure 2 shows true-stress true-strain curves for four different stretch rates, namely, -0.08, -0.83, -9.5 and -69.5 sec^{-1}. The first two are low enough to be assumed quasi-static and the correlation obtained using the predicted Boltzmann response, namely, Eq. (2-6) is

$$\frac{G(t)}{G(\infty)} = 1 + 1.14\, e^{-0.1t}$$

The corresponding correlation for the high strain-rates is

$$\frac{G(t)}{G(\infty)} = 1 + 6.12 e^{-5000t}$$

The modified Mooney response, Eq. (2-3), is also shown therein.

Dynamic shear tests on ½" diameter cylindrical specimens have been reported by Advani *et al* [1]. The curves for storage and loss moduli are shown in Figure 3 along with the response predicted by Eq. (2-7). The final correlation for is

$$\frac{G(t)}{G(\infty)} = 1 + 15 e^{-5000t}$$

Discussions and Conclusions

The strain energy approach has been shown to be a very useful method in predicting viscoelastic response. The strain energy form, established within a

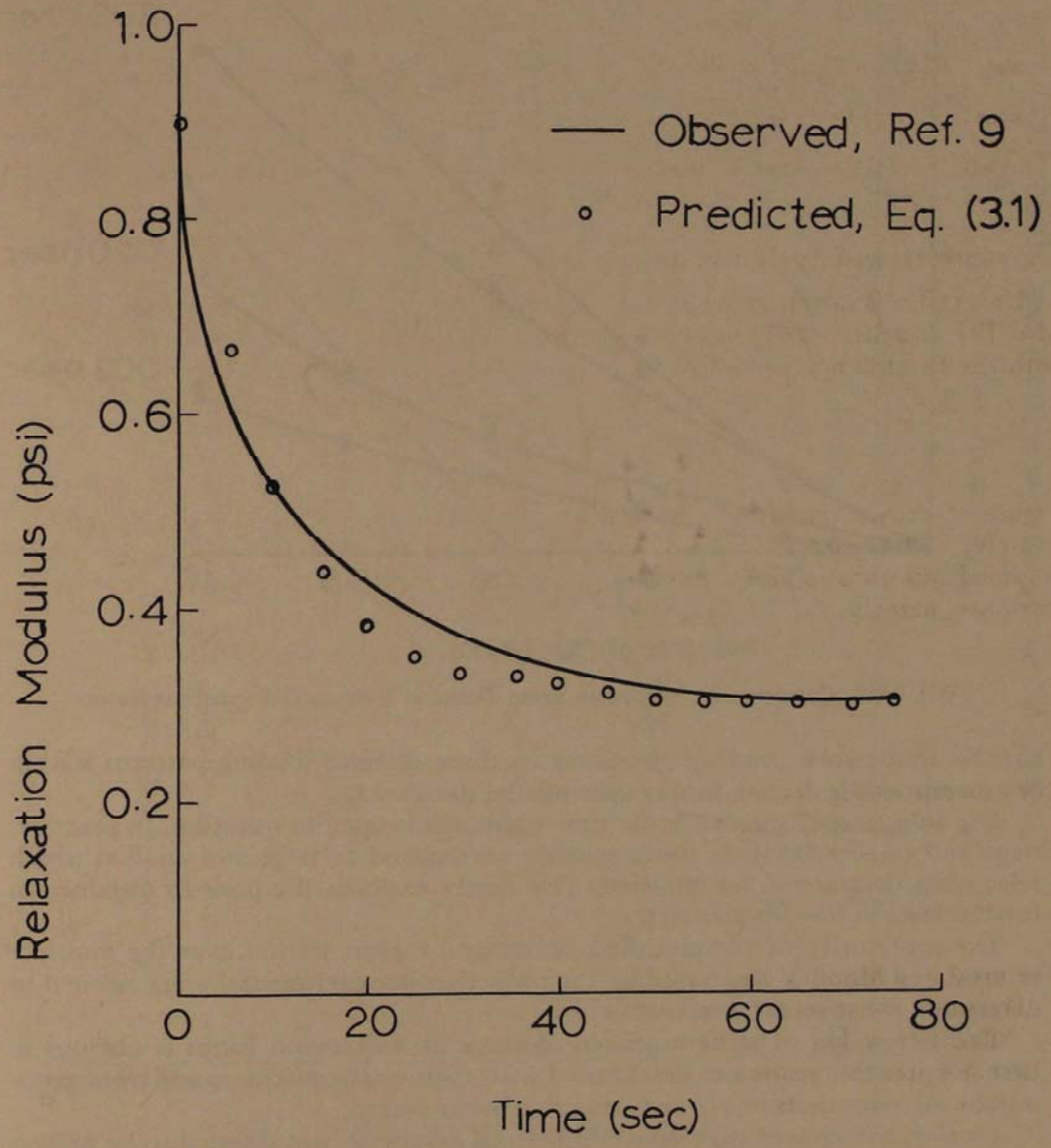


FIGURE 1. Relaxation Curves for Human Brain.

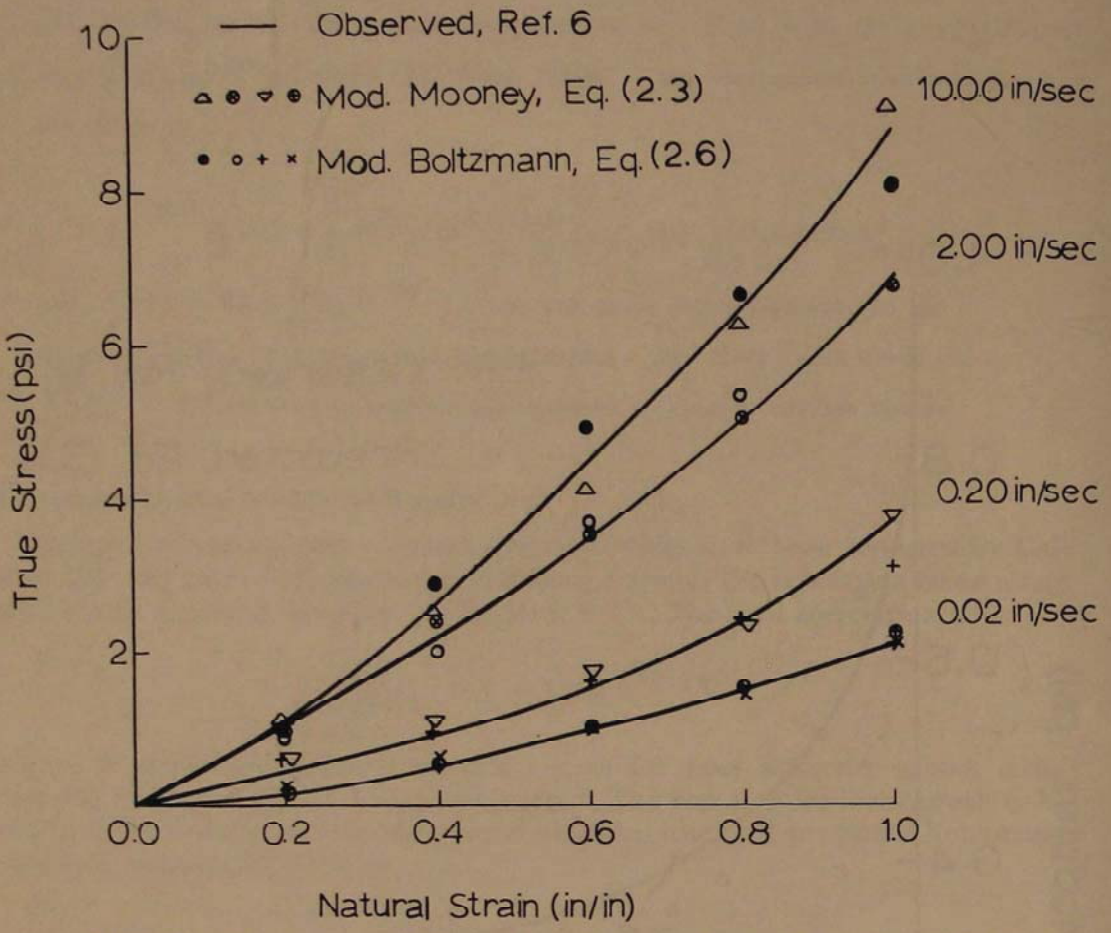


FIGURE 2. Compression of Human Brain Tissue at Various Deformation Rates.

narrow framework predicts responses in three distinct loading patterns within reasonable limits despite some experimental drawbacks.

The selection of values for the time constants is open to question. In practice, large and small relaxation times roughly correspond to large and small at which relaxation data are to be modeled. This partly explains the poor fit obtained in torsion tests at low frequencies.

The superiority of the modified Boltzmann representation over the multiple integral expansion is illustrated by the reduction in experimental work needed to determine the material constants.

The advantage of using modified Mooney or Boltzmann forms is obvious in that the present results can be checked with their elastic counterparts by neglecting the viscoelastic terms or reducing to a linear form.

Finally, the constitutive relations derived herein are not claimed to be universal and more experimental data are needed to support the present theory.

Acknowledgments

Part of the work presented herein was performed under the auspices of the National Institutes of Health Contract No. PH-43-67-1137. Computer time was provided by the West Virginia University Computer Center.

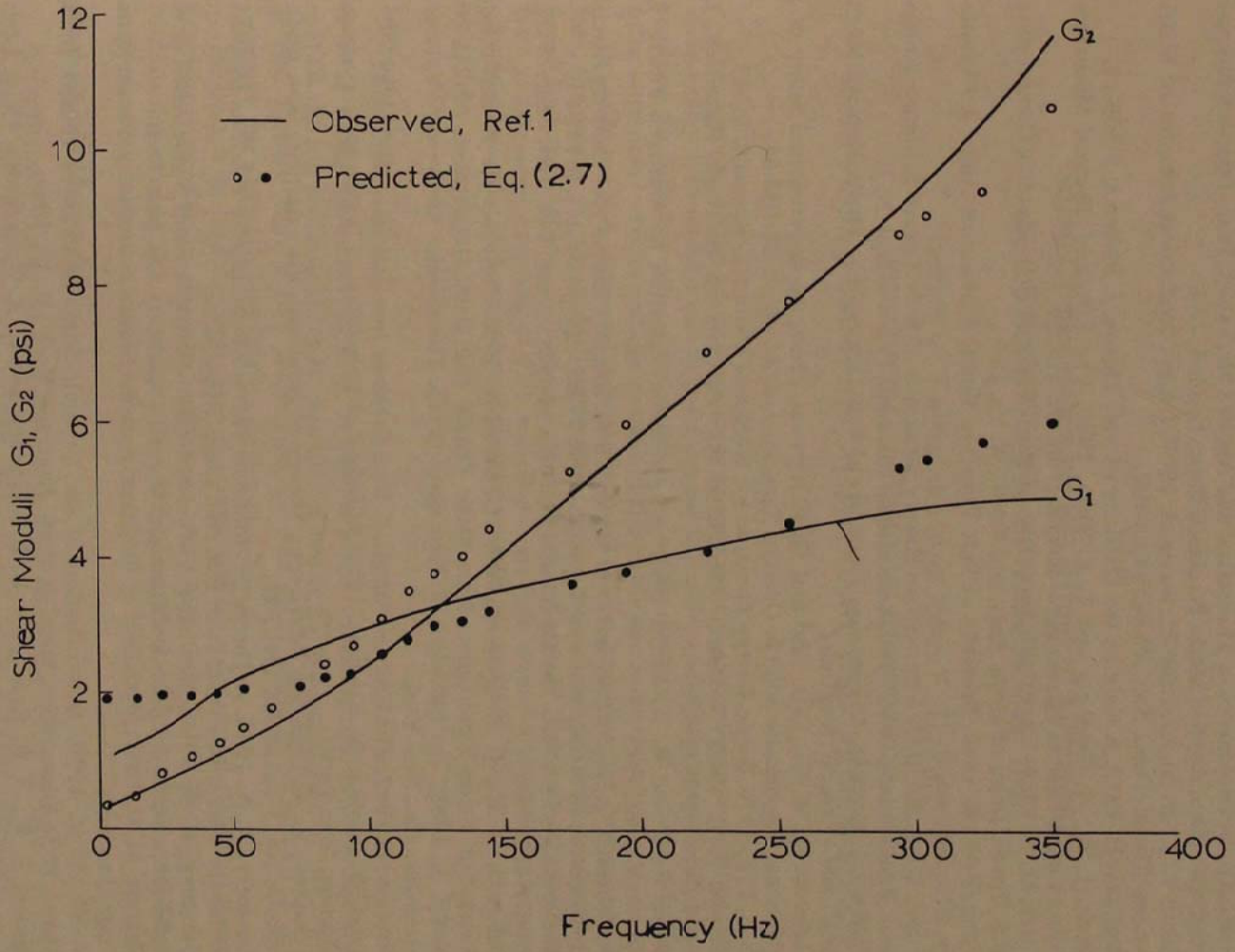


FIGURE 3. Dynamic Shear Moduli of Human Brain.

Literature Cited

1. Advani, S. H., R. P. Owings and L. Z. Shuck. 1972. Dynamic Response Evaluation of Translational and Rotational Head Injury Models. *Shock and Vibration Digest.* 4:3-16.
2. Ashton, J. E. 1970. An Elementary Boundary-Value Problem in Nonlinear Viscoelasticity. Developments in Theoretical and Applied Mechanics, Proc. Fourth Southeastern Conf. Theo. Applied Mech. Pergamon Press, London, 73-83.
3. Blatz, P. J., B. M. Chu and H. Wayland. 1967. On the Mechanical Behavior of Elastic Animal Tissue. *Trans. Soc. Rheol.* 13:83-102.
4. Demiray, H. 1972. A Note on the Elasticity of Soft Tissues. *J. Biomech.* 5:309-11.
5. Dorn, J. E., A. Goldberg and T. E. Tietz. 1949. The Effect of Thermal Mechanical History on the Strain Hardening of Metals. *J. Metals.* 1:325.
6. Estes, M. S. 1969. Viscoelastic Properties of Brain Tissue in Free-Standing Compression. M.S. Thesis, Dept. of Mech. Engr. and Mech., W. Va. Univ.
7. Fung, Y.–C. B. 1967. Elasticity of Soft Tissues in Simple Elongation. *Amer. J. Physiology.* 213:1532-44.
8. Fung, Y.–C. B. 1972. Stress-Strain-History Relations of Soft Tissues in Simple Elongation. Biomechanics: Its Foundations and Mechanisms, Prentice-Hall, N. J.
9. Galford, J. G. 1969. Viscoelastic Properties of Soft Tissues. M.S. Thesis, Dept. of Mech. Engr. and Mech., W. Va. Univ.
10. Gottenberg, W. G., J. O. Bird and G. L. Agrawal. 1969. An Experimental Study of a Nonlinear Viscoelastic Solid in Uniaxial Tension. *J. Appl. Mech.* 36:558-64.
11. Gou, P. F. 1970. Strain Energy Function for Biological Tissues, *J. Biomech.* 3:547.
12. Halpin, J. C. 1965. Nonlinear Rubberlike Viscoelasticity–A Molecular Approach. *J. Appl. Phys.* 36:2975-82.
13. Haut, R. C. and R. W. Little. 1972. A constitutive Equation for Collagen Fibers. *J. Biomech.* 5:423-30.
14. Hildebrandt, J., H. Fukaya and C. J. Martin. 1969. Simple Uniaxial and Uniform Biaxial Deformation of Nearly Isotropic Incompressible Tissues. *Biophysical J.* 9:781-91.
15. Holloman, J. H. 1946. The Mechanical Equation of State. Amer. Inst. Met. Engrs. Tech. Publ. No. 2034.
16. Johnson, A. E., J. Henderson and V. D. Matchur. 1959. Note on the Prediction of Relaxation Stress-Time Curves from Static Tensile Test Data. *Metallurgica.* 59:215-20.
17. Nadai, A. 1938. The Influence of Time upon Creep. Timoshenko Anniversary Volume, Macmillan Co., New York.
18. Orowan, E. 1947. The Creep of Metals. *J. West Scotland Iron and Steel Inst.* 54:45.
19. Patel, D. J. and R. N. Vaishnav. 1970. Rheology of Blood Vessels. ASME Winter Annual Meeting, New York.
20. Schapery, R. A. 1966. An Engineering Theory of Nonlinear Viscoelasticity with Applications. *Int. J. Solids Str.* 2:407-25.
21. Schapery, R. A. 1966. A Theory of Nonlinear Thermoviscoelasticity Based on Irreversible Thermodynamics. Proc. Fifth U.S. Nat. Cong. Appl. Mech.
22. Schapery, R. A. 1966. Nonlinear Viscoelastic Characterization and Stress Analysis of Solid Propellants. Solid Rocket Structural Integrity Abstracts, Cal. Inst. Tech.
23. Sharma, M. G. 1967. Mechanics of Polymers. Summer Short Course Notes, Pennsylvania State University.
24. Shuck, L. Z. 1970. Determination of the Dynamic Shear Modulus of Human Brain Tissue. Ph.D. Dissertation, Dept. of Mech. Engr. and Mech., W. Va. Univ.
25. Simon, B. R., A. S. Kobayashi, D. E. Strandness and C. A. Weiderhielm. 1970. Large Deformation Analysis of the Arterial Cross Section. ASME Winter Annual Meeting, New York.
26. Smith, T. L. 1962. Nonlinear Viscoelastic Response of Amorphous Elastomers to Constant Strain Rates. *Trans. Soc. Rheology.* 6:61-80.
27. Snyder, R. W. 1972. Large Deformation of Isotropic Biological Tissue. *J. Biomech.* 5:601-6.
28. Vorotnikov, G. S. and B. M. Rovinskii. 1966. Stress Relaxation, Creep and Uniaxial Strain: General and Special Features. *J. Appl. Mech. Tech. Phys.* (ZMTP). 7:13-17.

29. West Virginia University. 1971. Head Injury Model Construction Program, Final Report. Dept. of Mech. Engr. and Mech.
30. Zener, C. M. J. H. Holloman. 1946. Problems in Nonelastic Deformation. *J. Appl. Phys.* 17:69.

Some Observations on the Effect of Ultrasonics on the Solidification of Nodular Cast Iron

Harold V. Fairbanks and John F. Dakan
Department of Chemical Engineering
West Virginia University, Morgantown, West Virginia 26506

Abstract

The purpose of this investigation was to determine the influence the addition of ultrasonic energy into nodular cast iron during solidification would have upon the structure and resulting physical properties of the metal. A factorially designed set of experiments were run using ultrasonic intensity and the amount of alloy additions as the variables.

It was found that ultrasonic radiation retarded the decomposition of iron carbide. Therefore, the greater the ultrasonic intensity the harder the nodular cast iron produced. It was found that the addition of ultrasonic energy produced a small decrease in the nodular size. With an ultrasonic intensity of 120 watts for a one inch diameter cross-section there was a substantial increase in the number of graphite nodules per unit area formed.

Introduction

It is a well known fact that the application of ultrasonic radiation to molten metals during solidification is capable of reducing the grain size of the metal. It is also known that a fine dispersion of one solid phase in another can greatly improve the strength and quality of a metal.

Inasmuch as nodular cast iron is composed of a dispersed phase of spherical graphite particles in an iron matrix, it was conjectured that the application of ultrasonics during the formation of the spherical graphite particles could result in producing favorable mechanical properties.

Method and Materials

A factorially designed set of experiments were run using different ultrasonic intensities at a frequency of 20 kHz along with varying amounts of the magnesium-nickel alloy additive which aids in the production of the spherical nodular type of graphite in cast iron. The various levels of ultrasonic intensity and alloy additions used are given in Table 1.

Table I. Experimental Designs

Variables	*Levels*
Mg-Ni additions (#/ton)	17½, 20, 25
Ultrasonic intensity in watts	90, 120, 150

A vertical sand mold was used for casting a one inch in diameter test bar eight inches long. A steel rod was attached to the end of the ultrasonic horn and the free end of the steel rod was immersed approximately one-half inch beneath the top surface of the molten iron. The ultrasonics were applied during the solidification period of the iron. Figure 1 shows a schematic sketch of the apparatus used.

The alloy additions were made to the iron in the ladle just prior to pouring into the test mold. The analysis of the cast iron used is given in Table II.

Table II. Analysis of Cast Iron

Element	*Per Cent*
Carbon	3.71
Silicon	2.70
Manganese	0.36
Phosphorus	0.038
Sulfur	0.015

After solidification the bars were cross-sectioned and Brinell hardness tests made. A sample taken near the top of the test bar was polished for microscopic examination. The average size and distribution of the nodular graphite particles was determined.

Results and Discussion

Figure 2 shows both polished and etched samples taken from ultrasonically treated and untreated test bars. It can be seen that some decrease in nodular size was produced by the application of ultrasonic radiation during solidification of the cast iron. Also it can be seen in the ultrasonically treated and etched sample that some undecomposed iron carbide remained in the cast iron structure.

Figure 3 shows the effect of ultrasonic intensity upon the resulting hardness of the nodular cast iron. The greater the ultrasonic intensity, the harder the nodular cast iron produced. This result is directly related to the fact that ultrasonics retarded the decomposition of iron-carbide which is a very hard material.

Figure 4 shows the effect of ultrasonic intensity had on the size of the nodular graphite. There was a small decrease in graphite particle size due to ultrasonic radiation but not as great as had been conjectured.

Figure 5 gives the results of ultrasonic intensity upon the number of graphite particles per square millimeter. For the system used, it appears that 120 watts of ultrasonic power produced the maximum number of graphite particles. This may

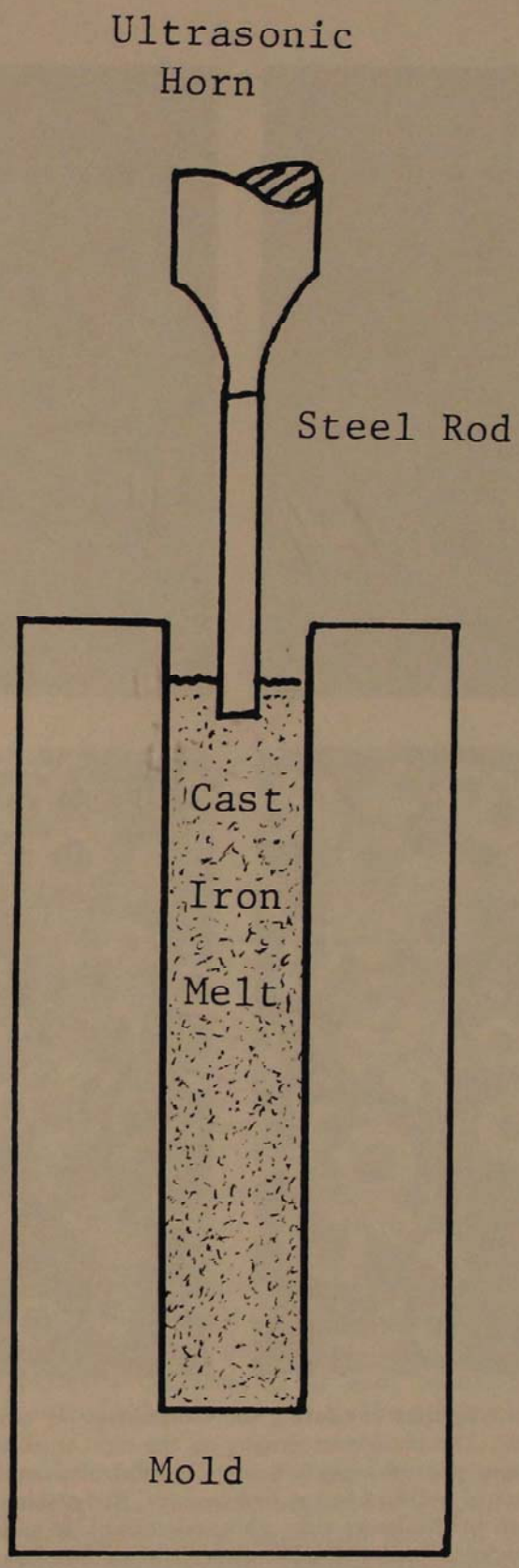


FIGURE 1. Schematic sketch of apparatus showing method of applying ultrasonic radiation during the solidification of nodular cast iron.

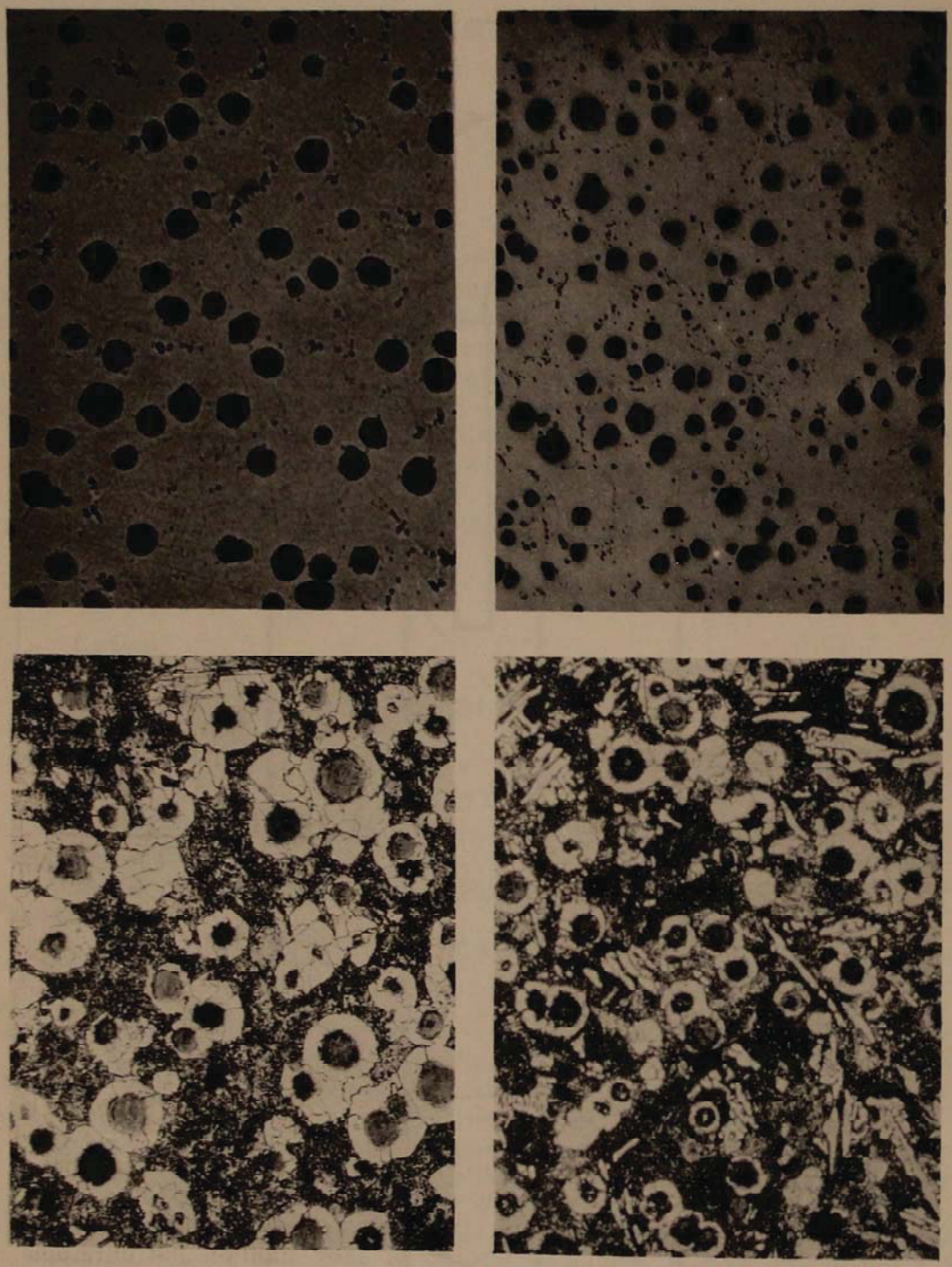


FIGURE 2. Photomicrographs comparing the morphology of ultrasonically treated metal with untreated metal. The photomicrographs on the right show the ultrasonically treated metal. The upper two photomicrographs are of polished samples, while the lower two photomicrographs are of polished and etched samples. The etchant used was Nital. The long white sections shown in the lower right photomicrograph is undecomposed iron carbide. The dark round particles is ferrite. Magnification of the photomicrographs is 100X.

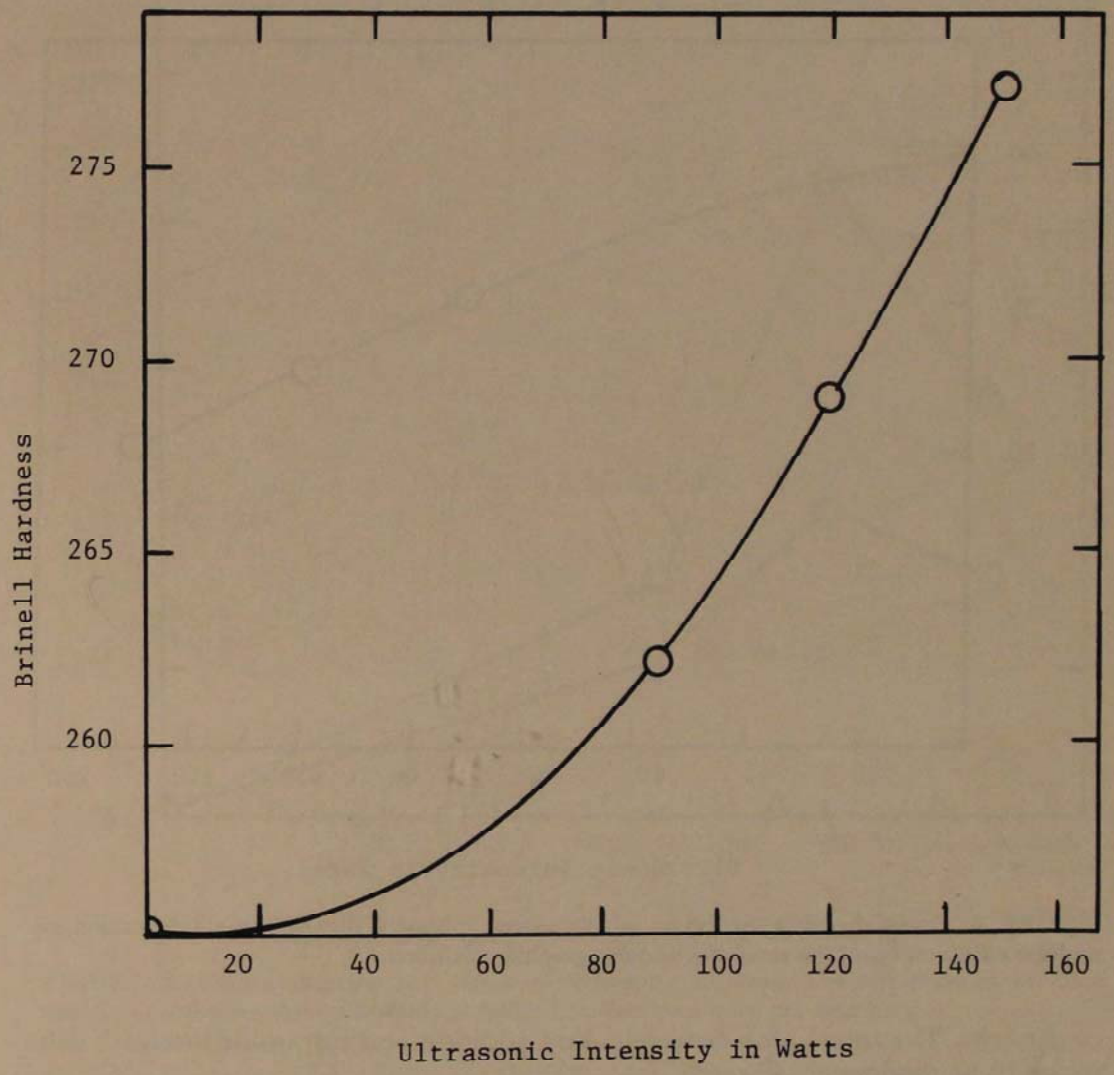


FIGURE 3. Curve showing the effect of ultrasonic intensity during the solidification of nodular cast iron upon the hardness of the resulting metal.

be due to the fact that more iron-carbide is retained at the higher ultrasonic intensities. Thus, the amount of carbon available in the form of graphite is diminished.

The calculated results of the experimental design for hardness showed that a maximum hardness of 500 Brinell should be obtained when using 190 watts of ultrasonic power and 30 pounds of alloy addition is made. A moderately positive interaction was found between the amount of alloy addition and ultrasonic intensity. The ultrasonic intensity was found to be the predominate variable for the set of conditions used.

The calculated results of the experimental design for the number of graphite particles per unit area showed that a maximum of 480 graphite nodules should be obtained per square millimeter when using 135 watts of ultrasonic power with an alloy addition of 16 pounds per ton. For this experimental design, it was found that the alloy addition was the predominate factor and had a negative

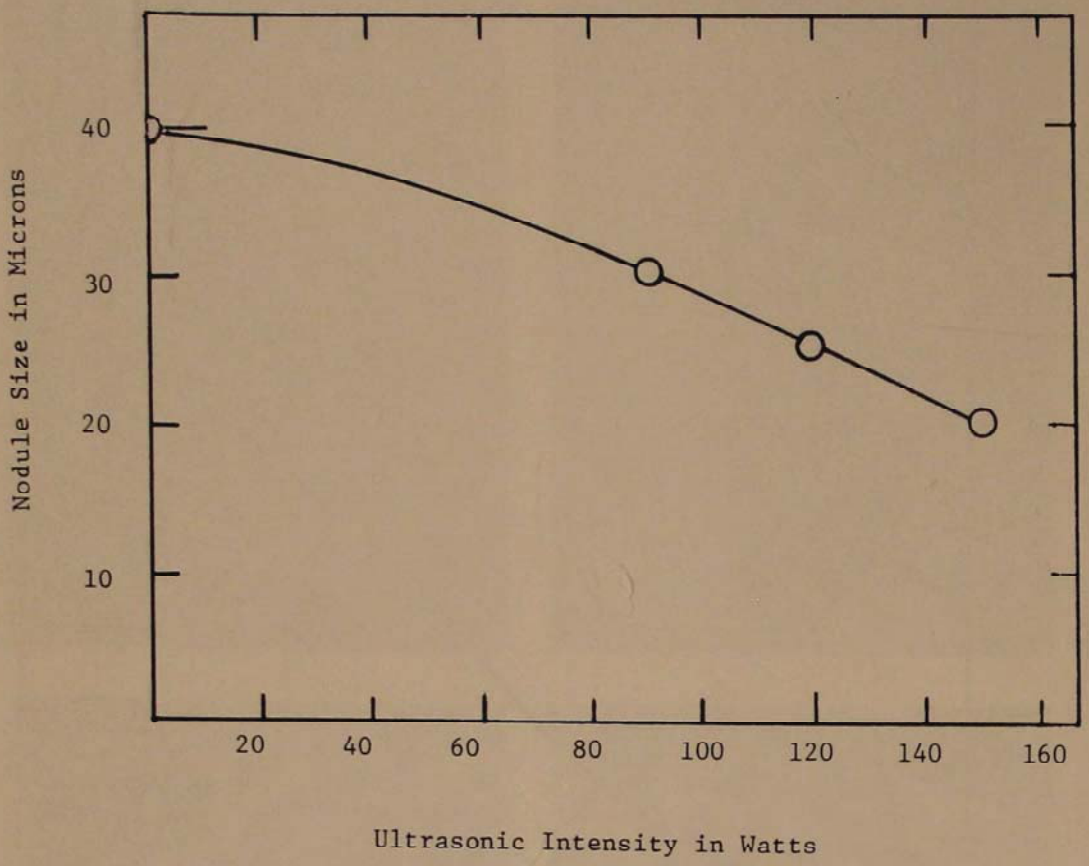


FIGURE 4. Curve showing the effect of ultrasonic intensity during the solidification of nodular cast iron upon the resulting nodular graphite diameter.

coefficient. The interaction between alloy addition and ultrasonic intensity was found to be moderately negative.

Conclusions

1. The application of ultrasonic energy during solidification of nodular cast iron retarded the decomposition of iron carbide, and thereby increased the hardness of the resulting material.
2. The application of ultrasonic energy during solidification of nodular cast iron resulted in a small but significant decrease in the size of the nodular graphite particles.
3. The application of ultrasonic energy during solidification of nodular cast iron resulted in an increase in the number of graphite nodules per unit area up to a given ultrasonic intensity. Ultrasonic intensity above this value produced more retained iron carbide and decreased the number of graphite nodules per unit area.
4. The addition of increased amounts of alloy along with ultrasonic radiation aided in the retention of iron-carbide and definitely decreased the number of graphite nodules produced per unit area.

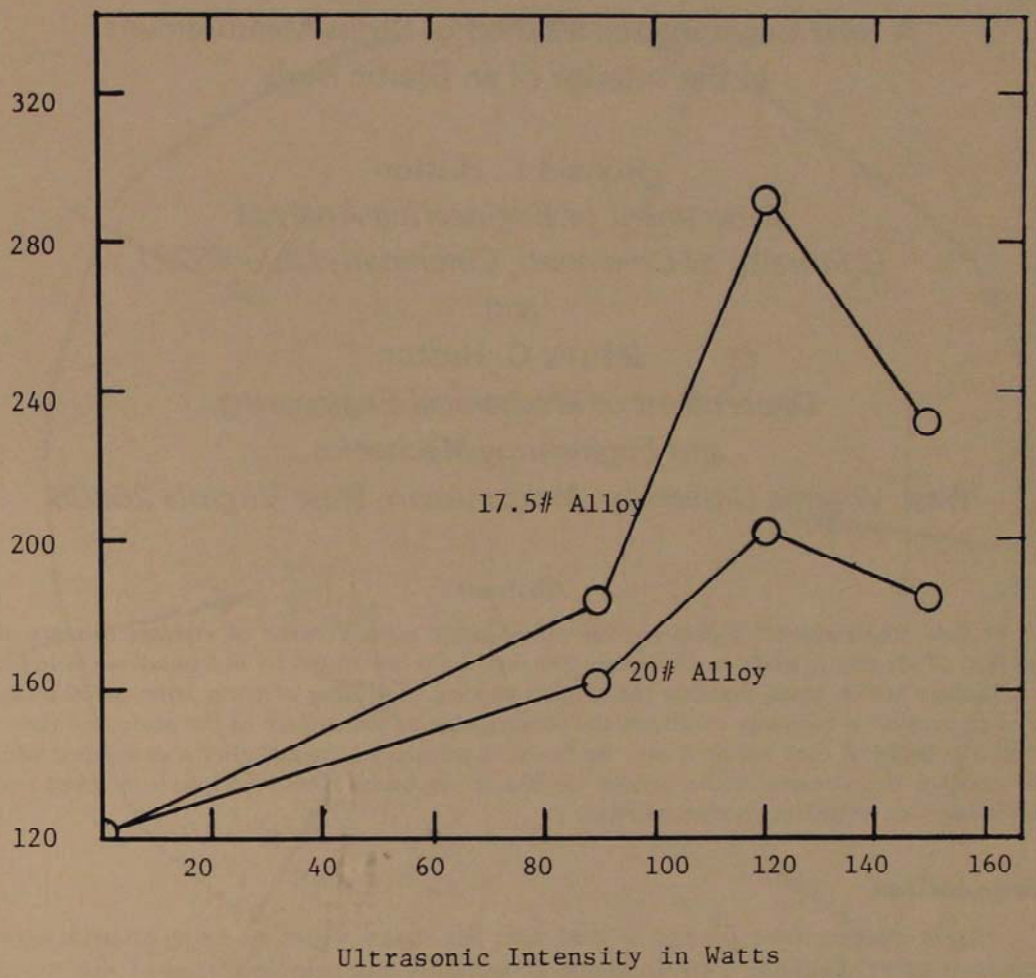


FIGURE 5. Curves showing the effect of ultrasonic intensity and the amount of alloy addition on the number of nodular graphite particles produced per unit area.

Acknowledgements

The authors wish to acknowledge the cooperation and use of facilities of the Kelly Foundry and Machine Company of Elkins, West Virginia. Also the authors wish to thank Branson Sonic Power Company of Danbury, Connecticut for supplying the instrumentation used in this investigation.

A New Experimental Method of Stress Measurement in the Interior of an Elastic Body

Ronald L. Huston
Department of Engineering Analysis
University of Cincinnati, Cincinnati, Ohio 45221
and
Jeffrey C. Huston
Department of Mechanical Engineering
and Engineering Mechanics
West Virginia University, Morgantown, West Virginia 26506

Abstract

A new experimental technique for strain-gauge measurement of stresses beneath the surface of an elastic body is discussed. The method is motivated by recent advances in laser technology which make possible the minute cutting or drilling of cores from metal bodies. The procedure is basically to attach the strain-gauge to the surface of the body and then to drill a cylindrical core about it into the body. A principle of stress relief is developed which determines the stresses at the various depths of the body. This principle is checked using techniques of reflective photoelasticity.

Introduction

Strain-gauges have played a vital role for many years in experimental stress measurement. Indeed, a strain-gauge is probably the principle tool of the typical experimentalist. However, strain-gauge techniques have the serious limitation of measuring surface strains only. That is, strain-gauge measurements generally do not provide information about the strain (or stresses) in the interior of the body—and many times these are the most important (i.e. maximum) strains (or stresses). It is the objective of this brief paper to introduce and discuss a new procedure which will extend the range of applicability of strain-gauge techniques so that interior strains may be measured.

This new technique is based upon the principle of residual stress relief through cylindrical core drilling. This core drilling is made possible by recent advances in laser technology which allows the drilling be done by minute local burning. The essence of the stress measurement principle is as follows: Consider an elastic body in a stress field as shown in Figure 1. The strain in the interior of the body beneath a typical point P on the surface is measured by drilling (coring) a cylinder about P into the body, as shown in Figure 2. This drilling operation relieves the stress in the cylindrical core. By strain-gauge measurement at P it is possible to predict the magnitude of the relieved stress along the axis of the cylinder. This, in effect, determines the interior stress distribution.

The following section of the paper, contains the theoretical basis for measuring this stress. This theory is then verified by experimental results discussed in the subsequent section. The final section of the paper contains a brief discussion of the scope and range of applicability of the method.

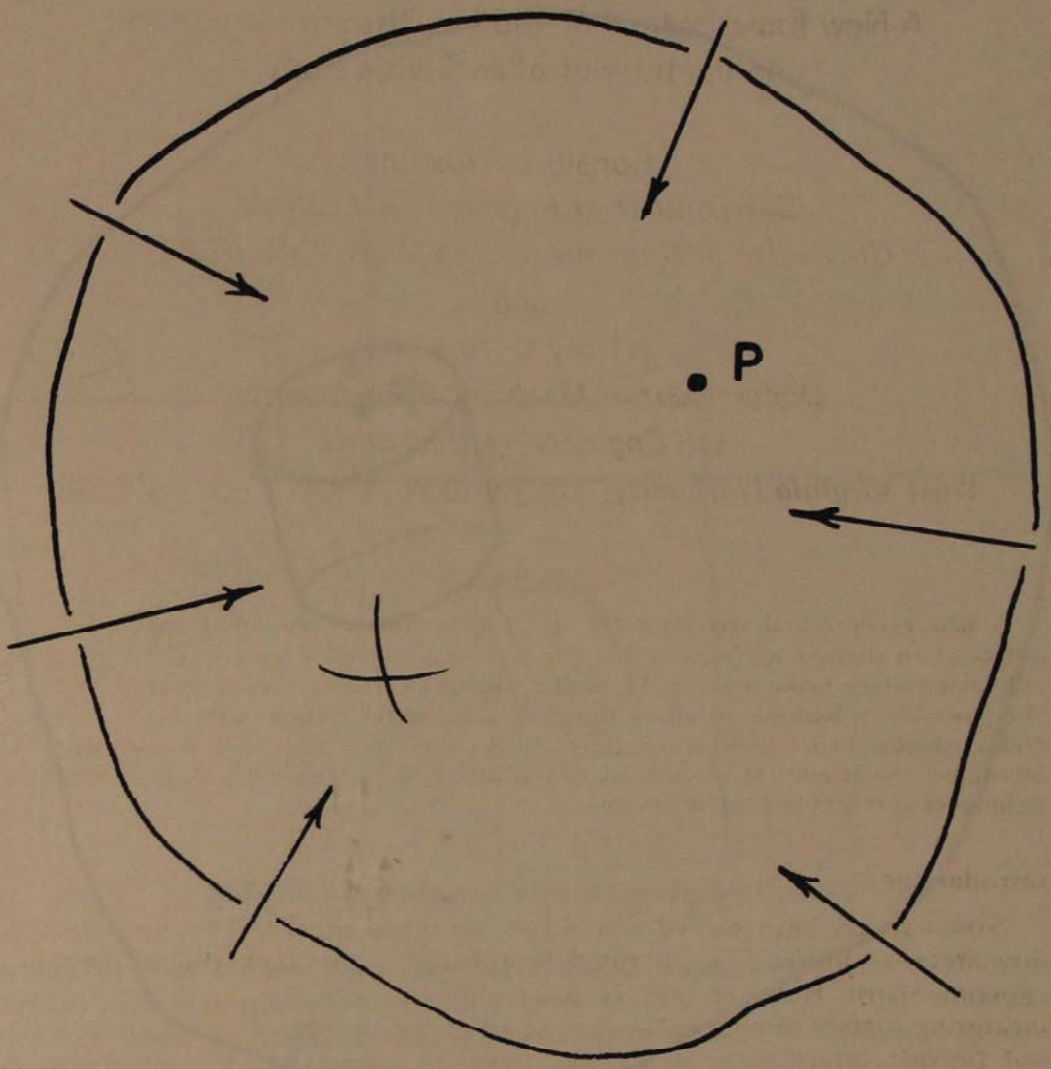


FIGURE 1. Elastic Body under Stress.

Theoretical Analysis

Consider an elastic body in a stress field as shown in Figure 1 and imagine the core-drilling of a cylinder into the body as shown in Figure 2. The basic theoretical problem is two-fold: First, the stress at the base of the cylinder must be analytically expressed in terms of the stress, and then the strain, on the top of the cylinder. Next, this stress at the cylinder base needs to be analytically related to the stress at the same point prior to the core drilling (stress relieving) operation.

The first problem may be expressed more succinctly by referring to the sketch of the core-drilled cylinder of Figure 3: That is, given the stress at P determine the stress at O. Similarly, the second problem becomes: given the stress at O, determine the stress at the same point just prior to the drilling of the cylindrical slot.

In the knowledge of the authors there does not currently exist exact, closed-form solutions to either of these problems. Indeed, to the authors' knowledge, there does not even exist analytical solutions to the simpler, two-dimensional

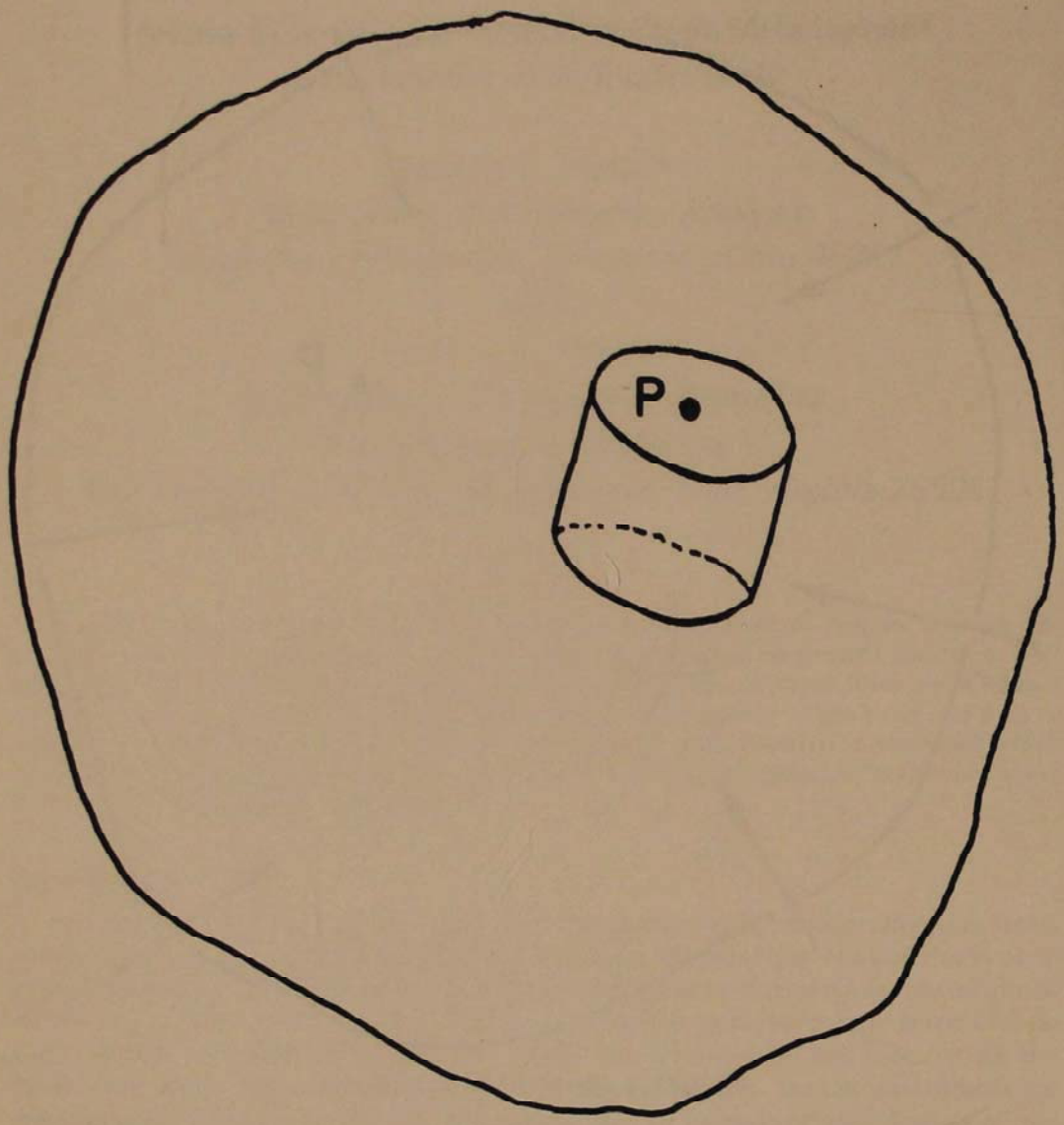


FIGURE 2. Core-Drilled Cylinder.

problems obtained by considering Figure 3 as a cross-section view of either a flat (plane stress) or infinitely thick (plane strain) body in the Z direction. Furthermore, no such solutions are presented here. However, much work has been done on similar, related problems, and by using the results of this work it is possible to extract quantitative as well as qualitative information about the solutions of the foregoing problems.

Specifically Klemm (3) indicates that the solution of the first problem is directly related to the "decay theories" emanating from St. Venant's Principle. The work of Knowles and Horgan (4) and of Timoshenko (5) substantiates this. Indeed the work of Knowles and Horgan (4) suggests that a close approximation to the solution of the first problem is:

$$\sigma_P = \sigma_O e^{-ky/R} \qquad (1)$$

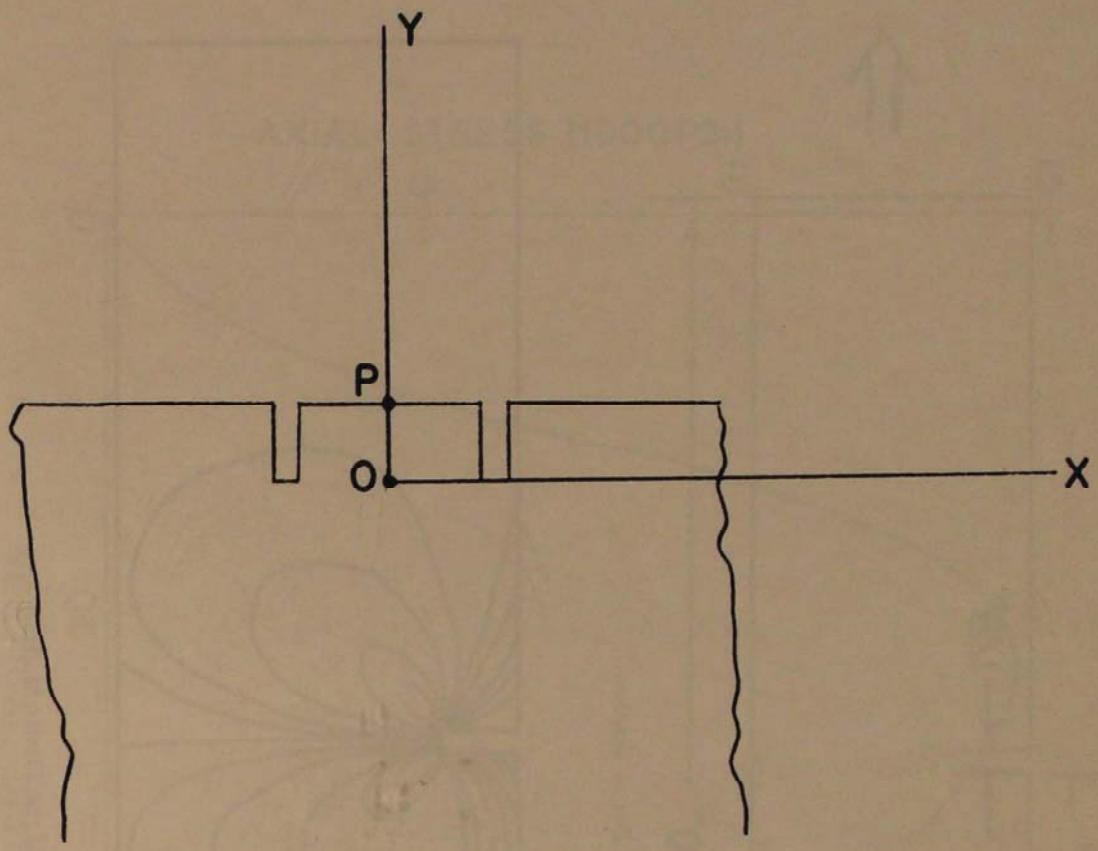


FIGURE 3. Cross-Section View of Core-Drilled Cylinder.

where σ_P and σ_O are the stresses at P and O respectively, R is the cylinder radius, and y is measured vertically from O (See Figure 3). The value of k depends upon the material. For a Poisson ratio of 0.25 to 0.30, k is approximately 1.4.

For the second problem, it is more difficult to use the results of other work to relate the stress at O to the stress at O prior to the core drilling. However, much work has been done on stress concentration factors due to notches and slots. This work—specifically, Green and Zerna (2), Faupel (1), and Yang (6)—suggest that the stress at O is slightly larger than the stress at O prior to the core drilling. That is, if this latter stress is called σ_O' then σ_O' and σ_O are related as:

$$\sigma_O' = K\sigma_O \tag{2}$$

where K is in the interval between 0.5 and 1.0.

To summarize these conclusions: The stress in the interior of an elastic body σ_O' is related to the measured stress on the surface σ_P by the expression:

$$\sigma_O' = K\sigma_P e^{ky/R} \tag{3}$$

where K, k, y, and R are as defined above. This expression states the basic theoretical principle of the proposed new experimental technique. Experimental verification of this result is presented in the next section.

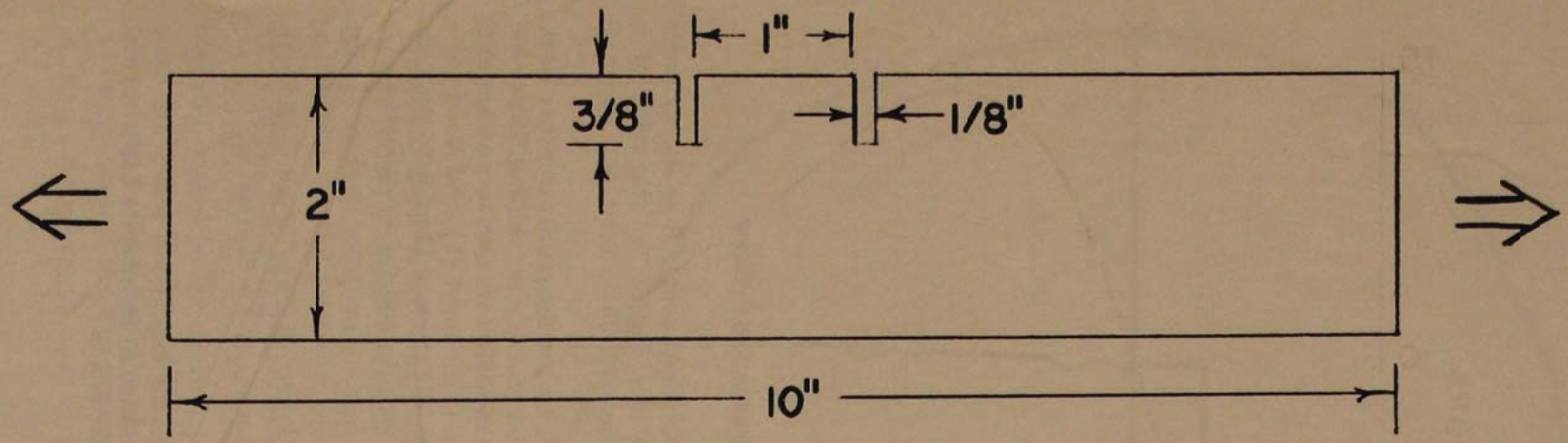


FIGURE 4. Plastic Bar in Tension.

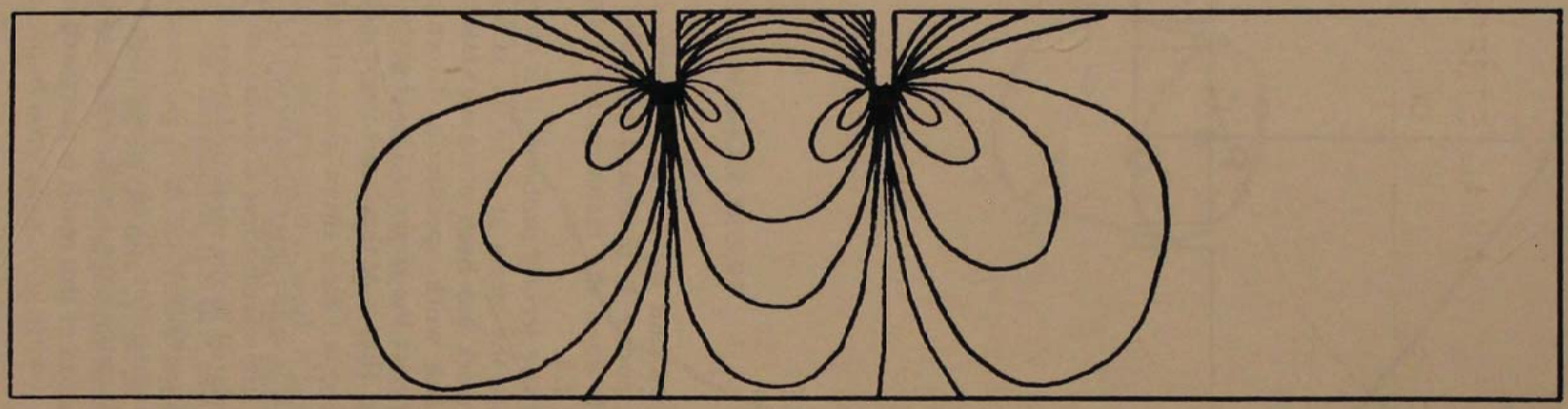


FIGURE 5. Isoclinics of the Plastic Model.

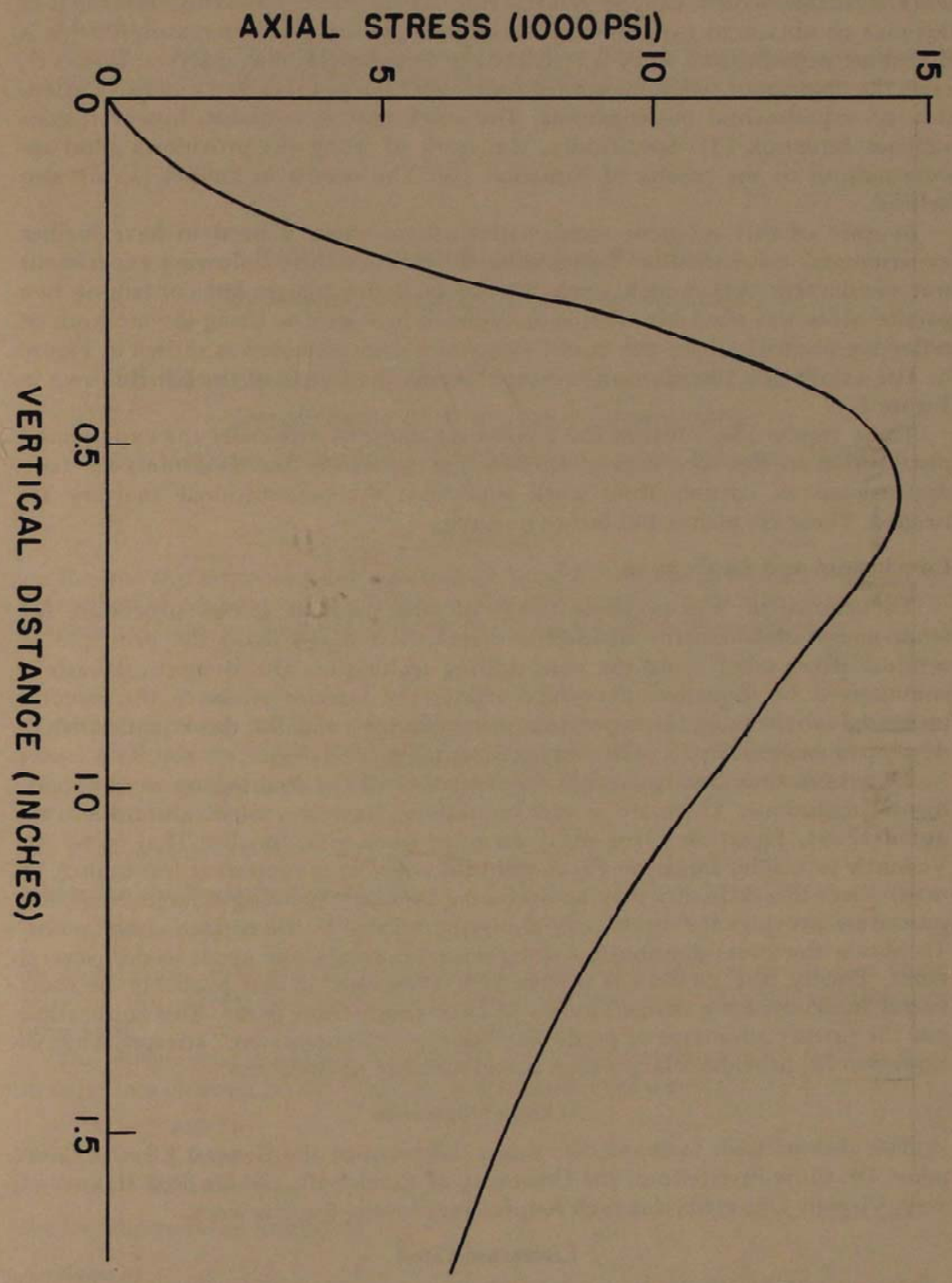


FIGURE 6. Axial Stress Distribution across the Center of the Bar.

Experimental Analysis

Just as with the theoretical problems above there is very little experimental work available which can be related directly to these problems. Hence, it is difficult to obtain an experimental check of Equation (3). One reason for this is that most experimental work is restricted to two-dimensional analyses. However, even the associated two-dimensional problems have not received extensive attention of experimental investigations. The work that is available, however, does validate Equation (3). Specifically, the work of Yang (6) provides a good approximation to the results of Equation (3). The results in Faupel (1) are also helpful.

In spite of this apparent confirmation, there exists a need to have further experimental confirmation. To partially fill this need the following experiment was conducted: A 1/8 inch thick 2 x 10 inch flat plastic bar containing two parallel slots was placed in tension as depicted in Figure 4. Using the methods of reflective photoelasticity the model's isoclinics were obtained as shown in Figure 5. The axial stress distribution measured across the center of the bar is shown in Figure 6.

These results also confirm the theoretical analysis—especially the exponential distribution in the core region. However, these results like the others are two-dimensional in nature. More work—especially three-dimensional analyses are needed. These are planned in future research.

Conclusions and Application

To summarize: the principle result of this study is a new procedure for strain-gauge measurement of interior stress. It is based upon the principle of residual stress relief using the core drilling technique. The theoretical basis is summarized by Equation (3) which relates the interior stress to the directly measured (strain-gauge) stress on top of the cylinder. Finally, this equation while developed analytically, is verified experimentally.

This result thus greatly extends the versatility of the strain-gauge as an experimental technique. There are several limitations, however, which should also be noted: First, Equation (3) is most accurate when y is "small." That is, by St. Venant's principle, Equation (3) is valid only if y/R is somewhat less than 2. In many cases this difficulty may be overcome by simply making R large. Next, the procedure provides the stress only along the normal to the surface at one point. To obtain the stress distribution along several normals, one needs to drill several cores. Finally, the method is destructive. Therefore, it will probably be most useful in determining design stresses in large production items. This application has the further advantage of predicting "safe" or "conservative" stresses. That is, Equation (3) provides a larger-than-actual measure of the stress.

Acknowledgments

The authors wish to thank Mr. Roger Adelman of the General Electric Company, Dr. Chris Passerello of the University of Cincinnati, and Dr. Rex Haynes of West Virginia University for their helpful suggestions for this work.

Literature Cited

1. Faupel, J. H. Engineering Design, John Wiley and Sons, Inc. New York Chapter 11.
2. Green, A. E. and W. Zerna. 1960. Theoretical Elasticity, Oxford, London, Chapter 8.
3. Klemm, J. L. and R. W. Little. 1971. "Saint Venant's Principle," Technical Report No. 10.1, Division of Engineering Research, Michigan State University.

4. Knowles, J. K. and C. O. Horgan. 1969. "On the exponential decay of stresses in circular elastic cylinders subject to self-equilibrated end loads," *Int. J. of Solids and Structures* 5:33-50.
5. Timoshenko, S. and J. N. Goodier. 1951. Theory of Elasticity, McGraw Hill. New York, Chapter 2.
6. Yang, C. 1963. "Stress Concentration Produced by Multiple Stress Raisers in Members under Uniaxial Tension," M.S. Thesis, West Virginia University.

Solution of an Eigenvalue Problem by Several Weighted Residual Methods

William Squire, E. E. Barth, H. F. Chou, and L. S. Fan
Department of Aerospace Engineering
West Virginia University, Morgantown, West Virginia 26506

Abstract

The first two eigenvalues of a second order linear differential equation are computed by the method of subdomains, collocation and Galerkin's method. Some tests of the effect of the choice of basis functions and nodes on the collocation results are presented.

Although there has been considerable recent interest (1, 2) in weighted residual methods for eigenvalue problems, it is difficult to find comparable results for the same problems by different methods. This paper presents calculations of the eigenvalues of

$$\ddot{y} + \lambda (1-x^2)\, y = 0 \qquad (1$$

subject to the boundary conditions

$$\dot{y}\,(0) = y\,(1) = 0 \qquad (2$$

by collocation, the method of subdomains, and Galerkin's method. Also some tests of the effect of changing the basis function and collocation points are presented.

The differential equation considered arises in connection with laminar heat transfer in a channel (3, 4). The first three eigenvalues are

$$\lambda_1 = 2.82778 \qquad (3a$$
$$\lambda_2 = 32.1478 \qquad (3b$$
$$\lambda_3 = 93.4664 \qquad (3c$$

the last figure being doubtful.

Analysis

The solution of a linear differential equation

$$(P + \lambda Q)\, y = 0 \qquad (4$$

subject to homogeneous boundary conditions is approximated by

$$y_N = c_j u_j \qquad (5$$

where the u_j's are a set of functions satisfying the boundary conditions. These are called the basis functions. A summation convention is used so the right side of eq. 5 is a sum of N terms. Because of the linearity, the residual is given by

$$r_N = c_j v_j \qquad (6$$

where

$$v_j = (P + \lambda Q)\, u_j. \qquad (7$$

Weighted residual methods obtain a set of N linear equations for the c_j's by introducing a set of N weight functions w_i and requiring

$$\int_0^1 w_i\, r_N\, dx = 0. \qquad (8$$

The condition for the existence that the determinant of the coefficients of the set of algebraic equations vanish, gives an equation for the eigenvalues.

Collocation corresponds to taking the w_i's as a set of Dirac delta functions so that the elements of the determinant are

$$d_{ij} = Pu_j + \lambda Qu_j \,|x = x_i. \qquad (9$$

The method of subdomains divides the range 0–1 into N strips and take w_i as 1 in the i^{th} strip, 0 elsewhere. Galerkin's method is based on

$$w_i = u_i. \qquad (10$$

In all three methods the elements of the determinant are linear in λ so the determinant corresponds to a polynomial of degree N in λ.

The method of least squares is based on minimizing

$$\int_0^1 r_N^2\, dx.$$

The equation obtained by setting the partial derivatives with respect to the c's equal to zero correspond to eq. 8 with

$$w_i = v_i. \qquad (11$$

In this case the elements of the determinant are quadratic in λ so the determinant corresponds to a polynomial of degree 2N in λ. It turns out that the roots are double roots. As this complicates the numerical solution, this method was not included in the calculations.

Numerical Results

A PL/1 computer program (available on request) was written which computes the elements of the determinant and then finds the zeros. The calculation of the determinant elements is easily modified to use any of the three methods or to handle other differential equations. Since multiplying out the determinant is laborious, the determinant is treated as a function of λ and the roots found by using an adaptive scanning procedure (5) to locate changes of sign in a specified range and then a combination of the secant rule and bisection is used to find the zero to a specified tolerance. This procedure does not require evaluation of derivatives which is rather difficult for determinants.

Table 1 compares the values of the first two eigenvalues obtained by the three methods using

$$u_j = 1 - x_j^2 \qquad (12$$

as the basis functions. The nodes of the collocation procedure were

$$x_i = (i - .5)/N. \tag{13}$$

Table 2 uses collocation with the above nodes and with node given by

$$x_i = (i - 1)/(N - 1) \tag{14a}$$

and

$$x_i = 0.5\,[1 + \cos(\frac{i-1}{N-1})\,\Pi]. \tag{14b}$$

The basis functions were

$$u_j = \cos(j - .5)\,\Pi x. \tag{15}$$

Table 1. First Two Eigenvalues Using $1-x_j^2$ as Basis Functions

	Collocation $x_i = (i - .5)/N$		*Subdomains*		*Galerkin*	
N/k	*1*	*2*	*1*	*2*	*1*	*2*
2	2.80117	50.9297	2.76846	67.4259	2.83286	45.9204
3	2.82915	30.2178	2.83202	28.5278	2.73050	33.4997
4	2.82768	32.7178	2.82739	34.3034	2.82776	32.2312
5	2.82776	32.0576	2.82778	31.6573	2.82776	32.1503
6	2.82776	32.1557	2.82776	32.2579	2.82776	32.1473

Table 2. Collocation Using cos (j-.5)Πx as Basis Functions

	$x_i = \frac{i-.5}{N}$		$x_i = \frac{i-1}{N-1}$		$x_i = 0.5\,[1+(\frac{i-1}{N-1})\Pi]$	
N/k	*1*	*2*	*1*	*2*	*1*	*2*
3	2.83025	32.4233	2.80914	33.1196	2.80914	33.1196
4	2.82854	32.2179	2.82474	31.7507	2.29082	30.7992
5	2.82808	32.1755	2.82686	32.0646	2.83465	32.4451

Discussion of Results

From Table 1 it appears that collocation, which is the simplest procedure, is certainly as good as the method of subdomains. It may be slightly inferior to Galerkin's method but it would seem well worth using a slightly larger N to save the trouble of evaluating the integrals involved in Galerkin's method.

Table 2 indicates that this choice of the basis function does not have a marked effect on the accuracy of the calculation but the choice of nodes does. The use of eq. 14b for even N results in a considerable decrease in the accuracy of the calculations. Apparently it is important to keep the residual small in the center of the range.

However, it would be dangerous to draw general conclusions from a single example. We plan to examine other cases to test the generality of these results.

Literature Cited

1. Becker, M. 1964. The Principles and Applications of Variational Methods, M.I.T. Press, Cambridge. 120 p. See Chapters 2 and 3.
2. Finlayson, B. A. 1972. The Method of Weighted Residuals and Variational Principles, With Application in Fluid Mechanics, Heat and Mass Transfer. Academic Press New York, 412 p.
3. Prins, J. A., J. Mulder, and I. Schenk. 1951. Heat Transfer to Laminary Flow Between Parallel Plates. *Appl. Sci. Res.* A2:431-38.
4. Sellars, J. R., M. Tribus, and J. S. Klein. 1956. Heat Transfer to Laminar Flow in a Round Tube or Flat Conduit—The Graetz Problem Extended. *Trans. A.S.M.E.* 78:441-48.
5. Shelud'ko, G. A. 1970. An Adaptive Method for Separating Real Roots of Algebraic and Transcendental Equations. *USSR Computational Mathematics and Mathematical Physics.* 10:265-73.

Cancer-Lymphocyte Interaction Analysis as a Problem in Differential Game Theory

Alvin M. Strauss

Department of Engineering Analysis

University of Cincinnati, Cincinnati, Ohio 45221

Abstract

Gurel in references [4] through [8] was the first to conclude, on the basis of results obtained by phase photomicrography of the motion of lymphocytes around and non-neoplastic and cancerous cells, that there exists an unstable biodynamic field in and around a cancerous cell. Particles that lie in the influence region of each individual cancer cell are affected by this field. Although attempts have been made to describe the cause of this field by Gurel the analysis concerns itself with the phenomenological theory of the cancer-lymphocyte interaction systems.

In this paper the interaction of cancer cells and lymphocytes is described within the framework of differential game theory. In this formulation we seek first to find the Nash equilibrium point where cancer cells and lymphocytes both try to optimize their payoffs irrespective of what the other is doing, but where neither will be able to unilaterally increase its payoff by changing its strategy.

Next, the transition is made to actual clinical problems by formulating the theory in terms of a multistage, discretetime, stochastic differential game model. The biodynamic field is assumed to be the mechanism whereby the cancer cell develops immunity to attack by lymphocytes. Thus, the cancer cell attempts to maximize its immunity and hence to maximize its biodynamic field. The lymphocyte, on the other hand, seeks to destroy the cancer cell and hence to minimize the effects of the field. The dynamics of this system is described by a set of linear difference equations. The effect of chemotherapy on this system is described where the chemotherapy is taken to be an independent noise sequence.

The Cancer Problem

Gurel [4] has presented evidence to indicate that there exists around cancer cells a biodynamic field. The basic dynamical behavior of a cancer cell is shown in Figures one through three. In Figure 1 a lymphocyte is seen moving towards a cancer cell in the intercellular fluid, reaching it, entering it and finally exiting

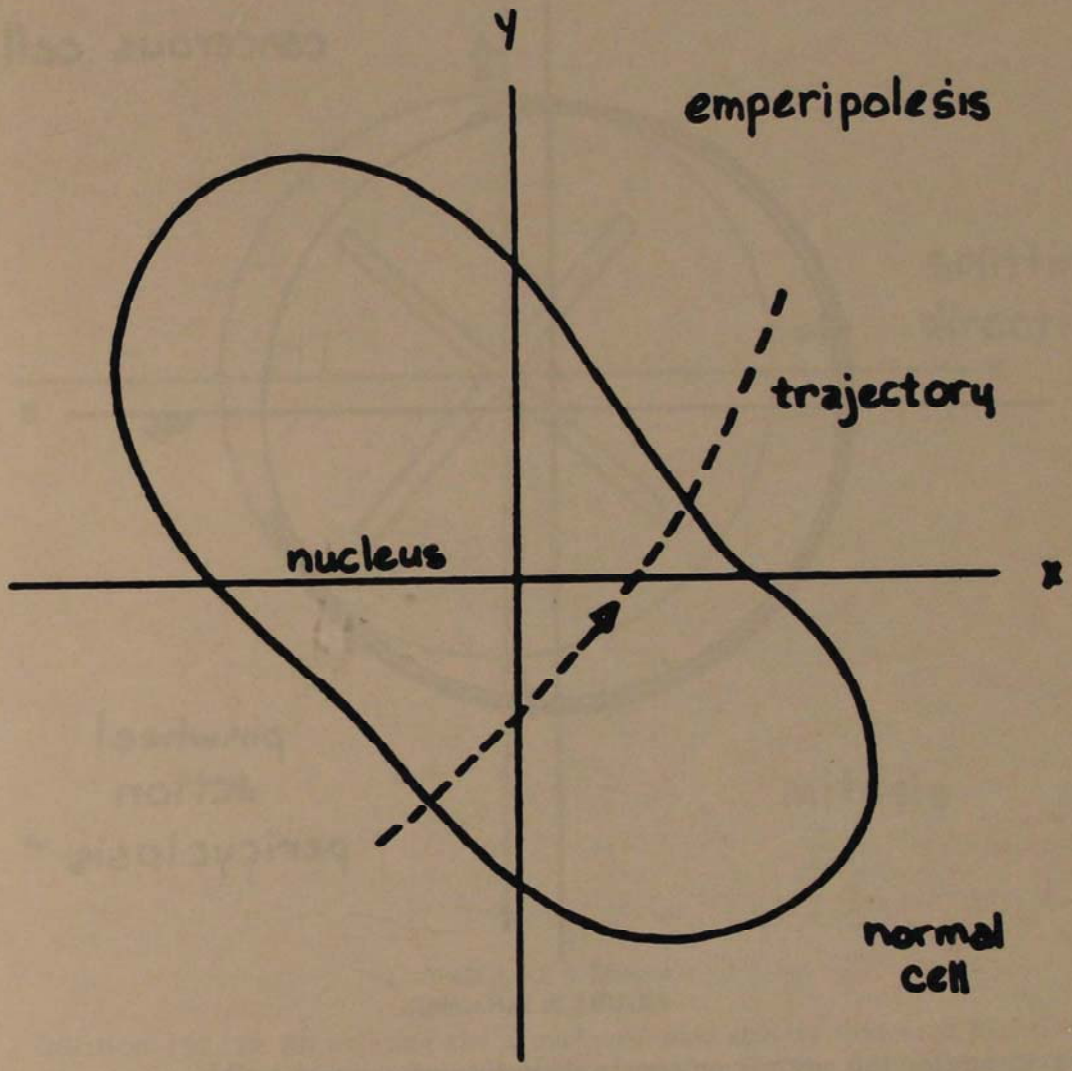


FIGURE 1. Emperipolesis.

from the cell. It has been observed that lymphocytes migrate towards certain cells, which are usually defective in some way, and destroy them. This migration is called clustering around the target cells. However, in the case of certain defective cells, namely cancer cells, lymphocytes migrate away from them. This is the strongest indication that some type of field effect exists around a cancer cell and that the lymphocyte can detect the fields of other defective cells. Further dynamic properties of a cancer cell that relate to the field concept are shown in Figures 2 and 3. In Figure 2 the rotational pinwheeling action of the peripheral portion of the cancer cell's cytoplasm is illustrated. This is called pericyclosis and has been related by Gurel to the limit cycle of a differential equation. Figure