Numerical Methods of Higher-Order Accuracy for Diffusion-Convection Equations

H. S. PRICE
JUNIOR MEMBER AIME
J. C. CAVENDISH
R. S. VARGA

-F RESEARCH & DEVELOPMENT CO.
ITTSBURGH, PA.
CASE WESTERN RESERVE U.
CLEVELAND, OHIO

ABSTRACT

A numerical formulation of high order accuracy, based on variational methods, is proposed for the solution of multidimensional diffusion-convection-type equations. Accurate solutions are obtained without the difficulties that standard finite difference approximations present. In addition, tests show that accurate solutions of a one-dimensional problem can be obtained in the neighborhood of a sharp front without the need for a large number of calculations for the entire region of interest.

Results using these variational methods are compared with several standard finite difference approximations and with a technique based on the method of characteristics. The variational methods are shown to yield higher accuracies in less computer time.

Finally, it is indicated how one can use these attractive features of the variational methods for solving miscible displacement problems in two dimensions.

INTRODUCTION

The problem of finding suitable numerical approximations for equations describing the transport of heat or mass by diffusion and convection simultaneously has been of interest for some time. Equations of this type, which will be called diffusion-convection equations, arise in describing many diverse physical processes. Of particular interest here is the equation describing the process by which one miscible liquid displaces another liquid in a one-dimensional porous medium. The behavior of such a system is described by the following parabolic partial differential equation:

$$\frac{\partial c(x,t)}{\partial t} = \frac{\partial^2 c(x,t)}{\partial x^2} - \lambda \frac{\partial c(x,t)}{\partial x}$$

$$\lambda > 0,\ 0 < x < \ell,\ t > 0 \quad . \ . \ . \ . \quad (1)$$

where the diffusivity is taken to be unity and $c(x,t)$ represents a normalized concentration, i.e., $c(x,t)$ satisfied $0 \leq c(x,t) \leq 1$. Typical boundary conditions are given by

$$\begin{cases} c(x,0) = 0;\ 0 < x \leq \ell, \\ \\ c(0,t) = 1;\ t > 0,\ \dfrac{\partial c(\ell,t)}{dx} = 0;\ t > 0. \end{cases}$$

$$. \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \quad (2)$$

Our interest in this apparently simple problem arises because accurate numerical approximations to this equation with the boundary conditions of Eq. 2 are as theoretically difficult to obtain as are accurate solutions for the general equations describing the behavior of two-dimensional miscible displacement. This is because the numerical solution for this simplified problem exhibits the two most important numerical difficulties associated with the more general problem: oscillations and undue numerical dispersion. Therefore, any solution technique that successfully solves Eq. 1, with boundary conditions of Eq. 2, would be excellent for calculating two-dimensional miscible displacement.

Many authors have presented numerical methods for solving the simple diffusion-convection problem described by Eqs. 1 and 2. Peaceman and Rachford[7] applied standard finite difference methods developed for transient heat flow problems. They observed approximate concentrations that oscillated about unity and attempted to eliminate these oscillations by "transfer of overshoot". It was later shown by Price, Varga and Warren[9] that these

Original manuscript received in Society of Petroleum Engineers office Sept. 10, 1967. Revised manuscript received June 24, 1968. Paper (SPE 1877) was presented at SPE 42nd Annual Fall Meeting held in Houston, Tex., Oct. 1-4, 1967.

This paper will be printed in *Transactions* volume 243, which will cover 1968.

[7]References given at end of paper.

oscillations were a direct result of spatial discretization and could be eliminated by using a fine spatial mesh. These latter authors also presented a second order correct spatial discretization that was non-oscillatory for any mesh spacing. However, numerical dispersion was still a problem. Garder, Peaceman and Pozzi[5] recognized that problems described by Eq. 1 for large values of λ behaved more like hyperbolic than parabolic differential equations. This led them to propose a method based on characteristics. Although non-rigorous, this approach using characteristics looked promising for equations of this type. Other authors—Stone and Brian[12] and Perrine and Gay[8]—have proposed other solution techniques for diffusion-convection-type equations; however, each of these has a serious drawback. A satisfactory extension to problems with dimension higher than one has not yet been found for the scheme of Stone and Brian, and Perrine and Gay state that their technique requires a large amount of computer time and storage.

This paper presents some numerical methods of high order accuracy for solving the diffusion-convection equation. These numerical approximations are based on variational methods, and have rigorously been shown by Ciarlet, Schultz and Varga[2] to be of high order accuracy for a wide class of nonlinear elliptic boundary value problems. These methods are still being tested. However, some preliminary results will be presented here. The approximate solution for the problem described by Eqs. 1 and 2 for various values of λ is calculated by these variational methods, some standard finite difference techniques, and by the Garder, Peaceman and Pozzi[5] method based on characteristics. These approximate solutions are compared with a known analytical solution and the variational methods are verified to be considerably more accurate and to require a great deal less computer time than any of the others considered. Finally, it is shown how these variational methods can be used in the calculation of two-dimensional miscible displacement.

ONE-DIMENSIONAL EXPERIMENTATION

The problem chosen here as the basis of comparison of the various methods is one that many authors have used for testing new numerical techniques.[5,7,9,11,12] This problem is to solve numerically by each of the methods listed below the equation describing the process by which one miscible liquid displaces another liquid in a constant cross-section homogeneous porous medium where both fluids are assumed incompressible. This system is described by the following equation.

$$\phi \frac{\partial \tilde{c}(x,t)}{\partial t} = D \frac{\partial^2 \tilde{c}(x,t)}{\partial x^2} - u \frac{\partial \tilde{c}(x,t)}{\partial x}, \quad 0 < x < \ell, \ t > 0, \quad \ldots\ldots\ldots (3)$$

with boundary conditions given by

$$\begin{cases} \tilde{c}(x,0) = 0; \ 0 < x \leq \ell \\ \tilde{c}(0,t) = 1; \ t > 0, \ \dfrac{\partial \tilde{c}}{\partial x}(\ell,t) = 0; \ t > 0, \\ \qquad \ldots\ldots\ldots\ldots\ldots\ldots (4) \end{cases}$$

With the following change of variables

$$\xi = x/\ell, \ \tau = Dt/\phi\ell^2, \ \tilde{c}(x,t) \equiv c(\xi,\tau),$$

Eqs. 3 and 4 become, respectively,

$$\frac{\partial c(\xi,\tau)}{\partial \tau} = \frac{\partial^2 c(\xi,\tau)}{\partial \xi^2} - \lambda \frac{\partial c(\xi,\tau)}{\partial \xi}, \quad 0 < \xi < 1 \ldots\ldots\ldots (5)$$

$$\begin{cases} c(\xi,0) = 0; \ 0 < \xi \leq 1 \\ c(0,\tau) = 1; \ \tau > 0, \ \dfrac{\partial c(1,\tau)}{\partial \xi} = 0, \ \tau > 0, \\ \qquad \ldots\ldots\ldots\ldots\ldots\ldots (6) \end{cases}$$

where

$$\lambda \equiv \frac{u\ell}{D} \quad \ldots\ldots\ldots\ldots\ldots (7)$$

Eqs. 5 and 6 were solved numerically on the IBM 360/75 by each of the methods described below for various values of the parameter λ. Each method was accurately timed and the accuracy of the solutions was obtained at various pore volumes injected. Numerical solutions are shown at three different times in Figs. 4 through 8, when the number of pore volumes injected are equal to 0.3, 0.5 and 0.8.

To characterize the accuracy of these methods, we need a very accurate approximation to Eqs. 5 and 6 and a measure of the accuracy. The following approximate solution to Eq. 5 for a semi-infinite strip was used:

$$C(\xi,\tau) = \frac{1}{2}\operatorname{erfc}\left(\frac{\xi-\lambda\tau}{2\sqrt{\tau}}\right) + \frac{1}{2}\exp(\lambda\xi)\operatorname{erfc}\left(\frac{\xi+\lambda\tau}{2\sqrt{\tau}}\right) \quad \ldots (8)$$

It should be noted that $C(\xi,\tau)$ satisfies Eq. 5 and all the boundary conditions of Eq. 6 *except* the right-hand end condition. However, for up to 2/3 of a pore volume, Eq. 8 is essentially an exact solution for Eqs. 5 and 6, and was used as a basis of comparison. The measure of accuracy E is given by the following:

$$E(\tau) \equiv \max_{0 \leq \xi \leq 1} |C(\xi,\tau) - c(\xi,\tau)| \quad . (9)$$

where $C(\xi,\tau)$ is the solution given by Eq. 8 and $c(\xi,\tau)$ is the numerical approximation. Since the methods described by A, B and C below are discrete approximations, Eq. 9 was modified to:

$$E(\tau) \equiv \max_{1 \le i \le n} |c_i(\tau) - c_i(\tau)| \quad . \quad (10)$$

Notice that the variational methods (see Appendix) give continuous solutions and, in these cases, Eq. 9 was used explicitly. We should remark that the measure of accuracy of Eq. 9 or Eq. 10 is not the only one that could be used. A *least-squares* measure of accuracy is more natural in terms of the mathematical error bounds that can be derived, but the measure of accuracy of Eq. 9 or Eq. 10 is, from a computing standpoint, the easiest to determine.

The following methods were considered in the comparison.*

1. Central Difference Approximations (CDA)

Using the shorthand notation, the CDA method is just the standard, second order correct, central finite difference approximations of Eqs. 5 and 6 for a uniform mesh of length h given by

$$\frac{dc_i(\tau)}{d\tau} = \frac{1}{h^2}\left[(1 - \frac{\lambda h}{2})c_{i+1}(\tau) - 2c_i(\tau) + (1 + \frac{\lambda h}{2})c_{i-1}(\tau)\right], \quad 1 < i \le n-1$$

$$\frac{dc_n(\tau)}{d\tau} = \frac{-2c_n(\tau) + 2c_{n-1}(\tau)}{h^2}$$

where $h = \frac{1}{n}$, $c_i(\tau) \equiv c(ih,\tau)$,

and $c_0(\tau) \equiv 1$.

2. Non-Central Finite Difference Approximations (NCDA)

These difference approximations, called NCDA, were shown by Price, Varga and Warren[9] to be second order correct and non-oscillatory and are given by

$$\frac{dc_1(\tau)}{d\tau} = \frac{1}{h^2}\left[-(2+\lambda h)c_1(\tau) + c_2(\tau) + (1+\lambda h)c_0(\tau)\right]$$

$$\frac{dc_i(\tau)}{d\tau} = \frac{1}{h^2}\left[-\frac{\lambda h}{2}c_{i-2}(\tau) + (1+2\lambda h)c_{i-1}(\tau) - (2 + \frac{3\lambda h}{2})c_i(\tau) + c_{i+1}(\tau)\right], \quad 2 \le i \le n-1$$

$$\frac{dc_n(\tau)}{d\tau} = \frac{2c_{n-1}(\tau) - 2c_n(\tau)}{h^2}$$

3. Method of Characteristics (MOC)

These approximations based on the Method of Characteristics are described in detail by Garder, Peaceman and Pozzi[5] and will not be repeated here. For this method, the fixed grid spacing was of length $h = 1/n$, and four moving points per grid interval were used.

4. Variational Methods (see Appendix for derivation)

Since these difference approximations are, in general, too long to describe, only the approximations using chapeau base functions will be discussed here. They are given by

$$\frac{2}{3}\frac{dc_1(\tau)}{d\tau} + \frac{1}{6}\frac{dc_2(\tau)}{d\tau} = -\frac{2}{h^2}c_1(\tau) + \frac{(1 - \frac{\lambda h}{2})}{h^2}c_2(\tau) + \frac{(1 + \frac{\lambda h}{2})}{h^2}c_0(\tau);$$

$$\frac{2}{3}\frac{dc_i(\tau)}{d\tau} + \frac{1}{6}\left(\frac{dc_{i+1}(\tau)}{d\tau} + \frac{dc_{i-1}(\tau)}{d\tau}\right) = -\frac{2}{h^2}c_i(\tau) + \frac{(1 - \frac{\lambda h}{2})}{h^2}c_{i+1}(\tau) + \frac{(1 + \frac{\lambda h}{2})}{h^2}c_{i-1}(\tau); \quad 2 \le i \le n-2$$

$$\frac{8}{15}\frac{dc_{n-1}(\tau)}{d\tau} + \frac{1}{6}\frac{dc_{n-2}(\tau)}{d\tau} + \frac{2}{15}\frac{dc_n(\tau)}{d\tau} = -\frac{7}{3h^2}c_{n-1}(\tau) + \frac{(\frac{4}{3} - \frac{\lambda h}{2})}{h^2}c_n(\tau) + \frac{(1 + \frac{\lambda h}{2})}{h^2}c_{n-2}(\tau);$$

$$\frac{8}{15}\frac{dc_n(\tau)}{d\tau} + \frac{2}{15}\frac{dc_{n-1}(\tau)}{d\tau} = -\frac{(\frac{4}{3} + \frac{\lambda h}{2})}{h^2}c_n(\tau) + \frac{(\frac{4}{3} + \frac{\lambda h}{2})}{h^2}c_{n-1}(\tau);$$

where $c_0(\tau) = 1$.

The associated equations for general subspaces, including piecewise-polynomial function subspaces, are derived in the Appendix. In the sequel, when referring to each of these approximations, they will be called by the name of the base functions used (i.e., chapeau, [illegible] and nonsmooth quintics).

* The variable τ in approximation to the time derivative was used instead of t in all cases, so only the semi-discretization is discussed here.

5. Variable Interpolation

For this case, two types of base functions were used. In the neighborhood of the front (i.e., two mesh lines on either side of the 50 percent concentration point) nonsmooth quintics were used, and for the rest of the region, chapeau base functions were used. The quintics moved in time with the front. The 50 percent concentration at any time τ was located by adding $u\Delta\tau$ to its position at the preceding time step.

This approach, using an accurate approximation in the neighborhood of steep gradients, is not new; see for example Warren, *et al.*,[14] McFarlane, Mueller and Miller[6] or Price and Donohue;[10] however, these procedures are difficult to implement in two dimensions. This is because if a front is not oriented along one of the coordinate axes, a fine mesh would be needed almost everywhere. For the variational methods, however, this idea can be extended to two spatial dimensions without difficulty because a high-order piecewise-polynomial surface may be used in any mesh block without affecting the neighboring blocks. Therefore, a fixed coarse mesh can be chosen and, when the front is located, only the mesh blocks containing it need have high-order piecewise-polynomial surfaces. The remaining blocks are unaffected. This allows us to work with problems of few unknowns and still obtain an accurate approximation.

DISCUSSION OF RESULTS

In this section, we discuss the numerical results obtained for the methods described in the previous section. The order of accuracy of each method is established and the various methods are compared both on the accuracy of the approximation and the amount of computer time necessary to obtain a given accuracy. These results are obtained for a wide range of the parameter λ since, in any realistic two-dimensional problem, λ is not a constant and may vary widely from one point in the region to another. The definition of λ in Eq. 7 shows that changes in λ reflect changes in either u, ℓ or D.

ORDER OF THE APPROXIMATION

It can be shown for each of the methods of approximation described above, except the method of characteristics, that the error $E(\tau)$ defined by Eq. 9 or Eq. 10 is of the form:

$$E(\tau) = K_1(\tau)h^{\alpha} + K_2(\tau)(\Delta\tau)^2 + \epsilon(\tau, h, \Delta\tau) \quad \ldots\ldots\ldots (11)$$

where $\epsilon(\tau, h, \Delta\tau)$ represents terms of higher order. Therefore, if $\Delta\tau << h$ and both h and $\Delta\tau$ are chosen small enough so that ϵ may be ignored, we have

$$E(\tau) \simeq K_1(\tau)h^{\alpha} = K_1(\tau)\left(\frac{1}{n}\right)^{\alpha}.$$

Taking logs of both sides, we have for any fixed value of τ

$$\log E(\tau) \simeq \log K_1(\tau) - \alpha \log n.$$

Thus, a plot of log E vs log n should result in a straight line with slope $-\alpha$. This is illustrated by Figs. 1 through 3. The departure from the straight line for small values of n is because the term $\epsilon(\tau, h, \Delta\tau)$ in Eq. 11 is not yet small enough to be ignored. The departure for large values of n is because $\Delta\tau$ is not small compared with $1/n$. Note that all the variational methods approach the same constant values as $n \to \infty$, which is simply the time truncation error.

Before discussing the experimentally determined slopes in relation to the values theoretically determined for elliptic problems, we should explain the strange behavior of the MOC curve in Fig. 1. As mentioned above, no theoretical determination of the error has been made for approximations of this type. Moreover, Fig. 1 illustrates that the error for this approximation does not get small as the mesh spacing decreases. This causes some concern about using approximations of this type, since it should be possible to reduce the error of these approximations some way, but we have been unable to do so for the method of characteristics. Because of this, for all time comparisons using MOC in the sequel, we have chosen 50 fixed grid points and four moving points for each mesh block.[5] The error of the MOC approximation was found to be somewhat dependent on $\Delta\tau$. That is, as $\Delta\tau$ was decreased from some large value, the error decreased until it reached about 0.02. Further decreasing $\Delta\tau$, however, had no effect. The choice of $\Delta\tau$ for the time comparisons will be discussed later.

From the other curves in Figs. 1 through 3, we have tabulated the slopes in Table 1. Note that in all cases the experimentally determined slopes are at least as large as those predicted by the theory for elliptic problems. The experimentally determined value for the smooth cubics is $\alpha = 4$; i.e., smooth cubics correspond to a high-order method. Although this is the result one would hope to be able to prove theoretically, this has not been accomplished to date. Another result shown in Table 1 is that the CDA, the NCDA and the chapeau approximations

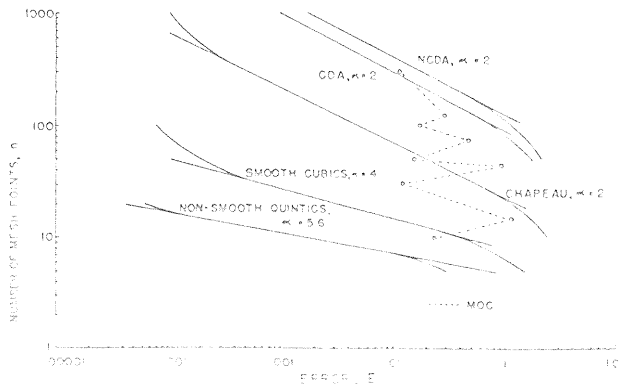


FIG. 1 — RESULTS OF MESH SPACING ON ERROR FOR A = 877.0.

all have slope $\alpha = 2$; however, Figs. 1 through 3 show the chapeau approximations to be far superior. This indicates that $K_1(\tau)$ in Eq. 11 is much smaller for the chapeau approximation than for the others. The fact that this approximation takes no more computing time than the other two for the same h and $\Delta\tau$ is, in itself, an excellent improvement (see Fig. 4).

One final point is that as λ increases, the slopes of these curves do not change even though the error does increase. This is as expected since the constant $K_1(\tau)$ involves high space derivatives of the concentration and these would increase as the front becomes sharper. The increase of $K_1(\tau)$ with λ is not a serious drawback because we were able to obtain excellent results for values of λ up to 100,000 (see Fig. 5).

COMPARISON OF THE VARIOUS TECHNIQUES

Tables 2 through 4 give a comparison of the computer time necessary to attain a given accuracy. These results were obtained at 1/2 PV injected; however, the comparison does not change much with time. All runs were made with the same time steps, chosen small enough so that the run times were accurate and the time truncation error was negligible.

Table 2 shows for $\lambda = 877.9$ that if an error of 0.02 is acceptable, chapeau base functions are the

TABLE 1 — COMPARISON OF COMPUTED AND THEORETICAL ORDERS OF ACCURACY

	Computed Slope $\lambda = 877.9$	Computed Slope $\lambda = 4389.5$	Computed Slope $\lambda = 8779$	Theoretical Slope
NCDA	2.0	2.2	—	2.0
CDA	2.0	2.0	2.1	2.0
Chapeau	2.0	2.2	2.3	2.0
Smooth Cubics	4.0	3.9	4.1	3.0
Nonsmooth Quintics	5.6	5.9	5.4	5.0

TABLE 2 — COMPARISON OF RUN TIMES, $\lambda = 877.9$ (TIME IN MINUTES $\times 10^{-2}$)

Accuracy	.02	.01	.005	.001	.0005	.0001
NCDA	350	512	765	1625	2600	5000
CDA	203	290	410	980	1400	3280
Chapeau	50	70	98	215	300	630
Smooth Cubics	51.5	64.5	77.4	116	138	212
Nonsmooth Quintics	75	87.5	100	137	150	200
Variable Interpretation	52	53	54	57	58	62
MOC	410	2400	—	—	—	—

TABLE 3 — COMPARISON OF RUN TIMES, $\lambda = 4389.5$ (TIME IN MINUTES $\times 10^{-2}$)

Accuracy	.02	.01	.005	.001	.0005	.0001
NCDA	1320	1795	2538	5480	7190	—
CDA	700	1000	1480	3420	5000	—
Chapeau	121	167	226	460	615	1280
Smooth Cubics	120	146	172	258	305	456
Nonsmooth Quintics	181	212	238	312	362	475
Variable Interpretation	61	63	65	71	75	84
MOC	410	—	—	—	—	—

TABLE 4 — COMPARISON OF RUN TIMES, $\lambda = 8779$ (TIME IN MINUTES $\times 10^{-2}$)

Accuracy	.02	.01	.005	.001	.0005	.0001
CDA	1350	1920	2720	6150	—	—
Chapeau	199	255	340	660	895	1720
Smooth Cubics	172	206	241	348	413	612
Nonsmooth Quintics	262	288	325	437	500	650
Variable Interpretation	67	69	72	81	86	98
MOC	410	—	—	—	—	—

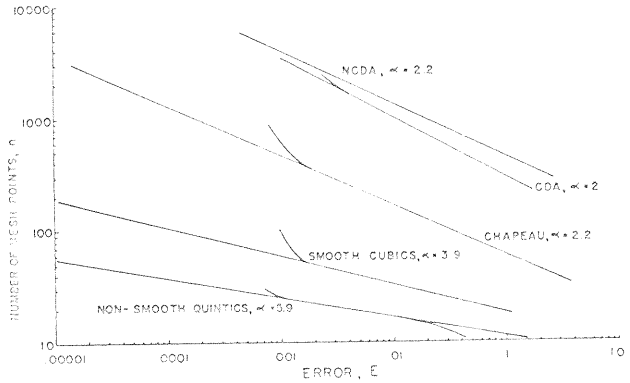


FIG. 2 — RESULTS OF MESH SPACING ON ERROR FOR $\lambda = 4389.5$.

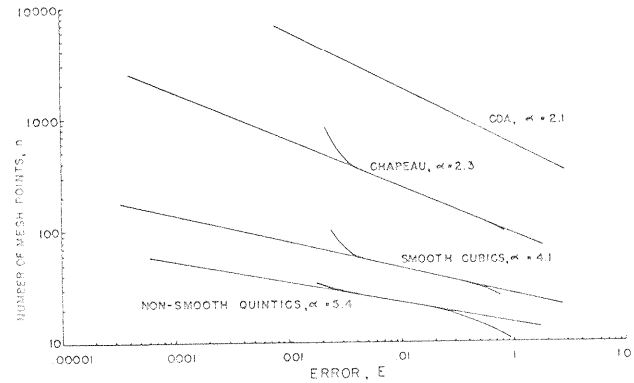


FIG. 3 — RESULTS OF MESH SPACING ON ERROR FOR $\lambda = 8779$.

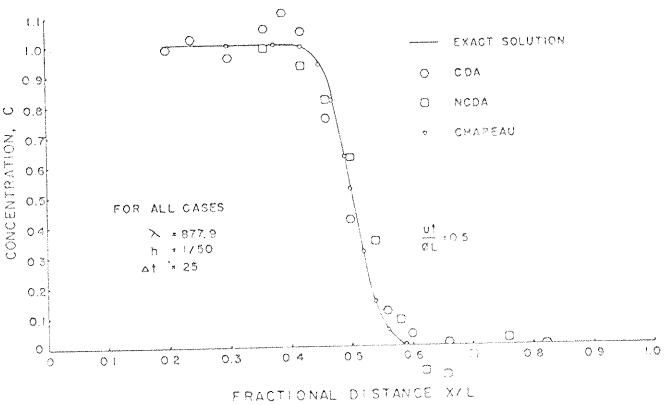


FIG. 4 — COMPARISON OF SOME SECOND ORDER NUMERICAL SOLUTIONS.

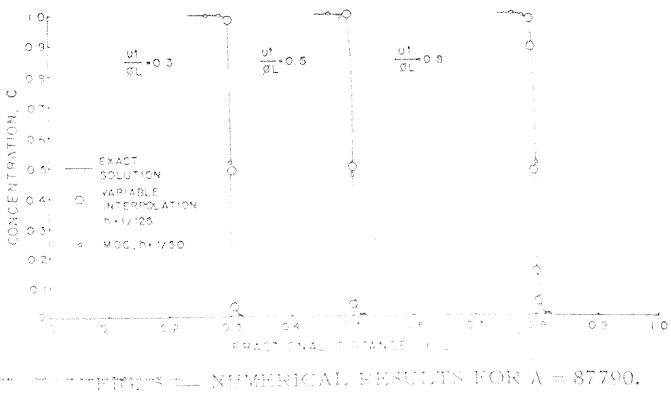


FIG. 5 — NUMERICAL RESULTS FOR $\lambda = 87790$.

best choice since this run takes only 0.5 minutes. However, if an accuracy greater than 0.005 is needed, variable interpolation should be used. As λ increases, variable interpolation is better for higher and higher accuracies (see Tables 3 and 4). The Method of Characteristics has only one entry in these tables because accuracies greater than 0.02 were unattainable, although, as λ increases, the accuracy of MOC becomes more and more competitive. However, even for $\lambda = 87{,}790$, variable interpolation is still slightly faster. Fig. 5 displays the results for these two cases. Here, $\Delta\tau$ was chosen as large as possible for each case to give acceptable accuracy. As can be seen from the curves, both are within engineering accuracy but variable interpolation took seven-eighths as long to run as MOC.

For smaller values of λ, MOC is less and less competitive. Fig. 6 displays the results for $\lambda = 877.9$ (the Peaceman and Rachford data: $u = 0.0048$, $D = 0.001$, $\ell = 182$ cm). Both methods gave acceptable accuracy; however, the results obtained by variable interpolation required only 6 grid points and took only one-tenth as much computer time as MOC. Again, $\Delta\tau$ was chosen as large as possible for each case to obtain the time comparisons.

Fig. 7 shows the results of reducing λ still further. Using variable interpolation required *only three grid points* as compared with 30 for the chapeau and 50 for MOC. These mesh spacings were found by experimentation to be close to the largest that would give acceptable accuracy. The same caution was used in selecting $\Delta\tau$.

For this case, the chapeau base functions were slightly faster than variable interpolation; however, both were 15 times faster than MOC. Another result of these curves relates to work in two dimensions. Using smooth bi-quintic Hermite polynomials with three grid points in each direction, only 169 unknowns need to be computed as compared with 900 for the chapeau (30×30) and at least 2,500 (50×50) for MOC. Solving this small system of linear equations should give a large savings in computer time if iterative methods are used.

Note (Fig. 8) the inaccuracy of the MOC method at 25 grid points for $\lambda = 175$, illustrating that for small values of λ, MOC is even less desirable than the CDA method. All the variational methods are superior to this.

Before concluding this numerical comparison, some mention must be made of the error as a function of time. It can be seen from Figs. 2 through 4 of Garder, Peaceman and Pozzi's paper[5] and from the comments of these authors, that the error of MOC fluctuates between some upper and lower limit as a function of time. We have observed similar fluctuations for this error; however, these fluctuations can be traced to the position of the front relative to the fixed mesh. When the 50 percent concentration point falls on a grid line, the error is greater than when it falls between grid lines (see Fig. 9). This is because the use of a

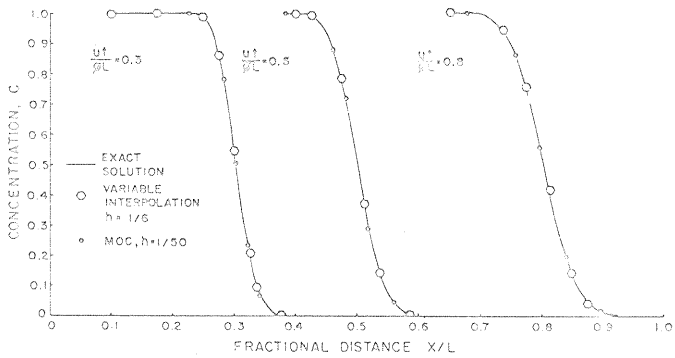


FIG. 6 — NUMERICAL RESULTS FOR $\lambda = 877.9$.

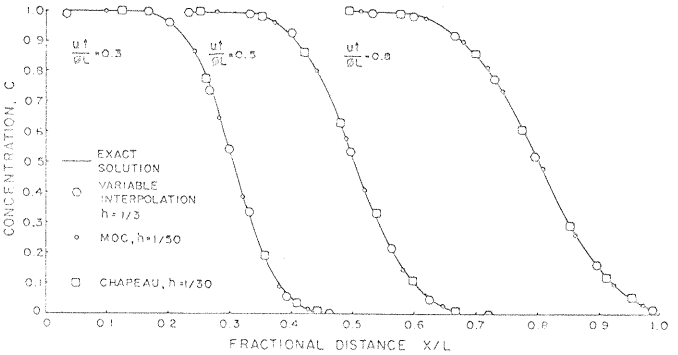


FIG. 7 — NUMERICAL RESULTS FOR $\lambda = 175$.

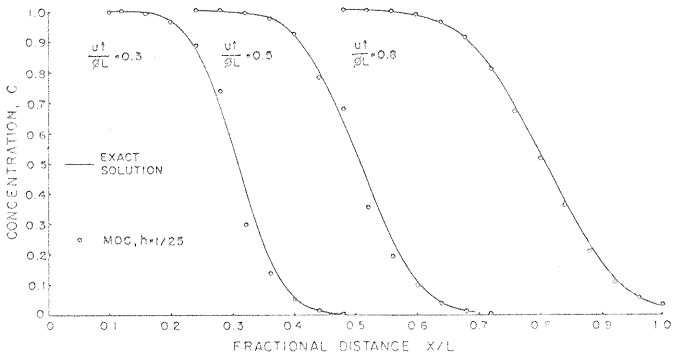


FIG. 8 — NUMERICAL RESULTS FOR $\lambda = 175$.

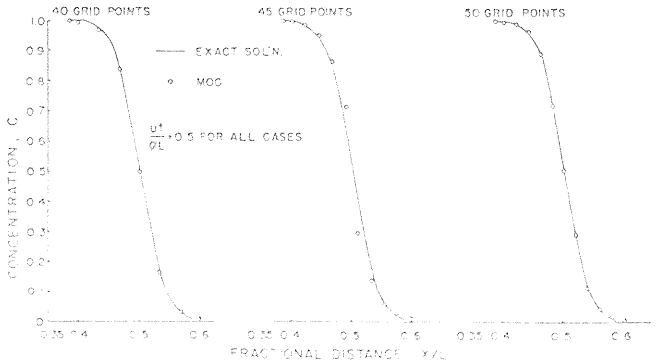


FIG. 9 — EFFECT OF MESH SPACING ON THE METHOD OF CHARACTERISTICS FOR $\lambda = 877.9$.

dual mesh system in MOC results in large, nonphysical fluctuations of the fluid quantities described by the moving points around the 50 percent concentration point. These fluctuations tend to even out when they are averaged at each grid point. If some consistent interpolation scheme were used for the entire region, and the error defined by Eq. 9 were used, these fluctuations would be damped out and the error would never get much better than the upper bound (about 0.02).

The error for the variational methods behaves quite differently with time (see Fig. 10). Notice that it starts at some large value, declines to a minimum and then increases again. The increase above the minimum for large times is caused by time truncation and round-off errors that accumulate as well as by the fact that the solution of Eq. 8 does not satisfy the right-hand boundary condition and therefore is not too reliable for large times. Of more concern are the large errors for small times. These occur because the error defined by Eq. 9 finds the maximum error over the entire region and not just at the mesh points. Therefore, to obtain initial conditions for the variational methods one must fit the initial data of the problem (see Appendix). The initial data for the problem of Eqs. 3 and 4 produce a square wave that is difficult to fit accurately. Therefore, the error at early times reflects this poor fit to the initial data. However, this is of no great concern because the initial conditions of the problem only approximate a smoother physical situation and these errors damp out quickly with time. In fact, if the error were examined only at the grid points for the variational methods as was done for the approximations, this large initial error would disappear and the curve of Fig. 10 would flatten out with a maximum around 0.003.

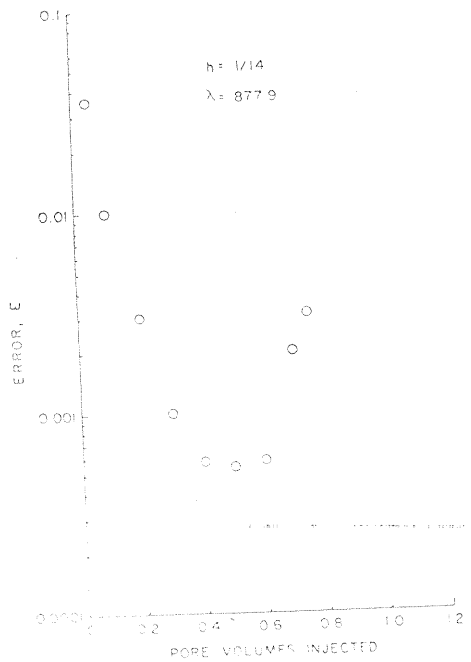


FIG. 10 — RESULTS OF ERROR VS TIME FOR VARIABLE INTERPOLATION.

EXTENSION TO TWO DIMENSIONS

A general development of the equations describing two-dimensional miscible displacement is given by Peaceman and Rachford[7] and Warren, Price, Skiba and Varga.[14] The equations presented in the latter paper are:

$$\frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z} = 0, \quad \ldots\ldots\ldots\ldots (12)$$

$$\frac{\partial}{\partial x}\left[D_x \frac{\partial c}{\partial x}\right] - u_x \frac{\partial c}{\partial x} + \frac{\partial}{\partial z}\left[D_z \frac{\partial c}{\partial z}\right] - u_z \frac{\partial c}{\partial z} = \phi \frac{\partial c}{\partial t}, \quad \ldots\ldots\ldots (13)$$

where

$$\begin{cases} u_x = -\dfrac{k_x}{\mu}\dfrac{\partial p}{\partial x}, \\ u_z = -\dfrac{k_z}{\mu}\left(\dfrac{\partial p}{\partial z} + \rho g\right), \\ D_x = D' + \dfrac{\alpha_x}{\phi}|u_x|, \\ D_z = D' + \dfrac{\alpha_z}{\phi}|u_z|. \end{cases} \quad \ldots (14)$$

Once suitable boundary conditions are specified, the following technique, commonly employed for nonlinear coupled systems of equations,[4,14] may be employed. For example, once the concentration $c(x,z,t)$ is known, Eq. 12 becomes a simple elliptic equation that can easily be solved for the pressure distribution using high-order variational techniques.[7] Once the pressure $p(x,z,t)$ is obtained from Eq. 12, u and D may be evaluated by Eq. 14 and then Eq. 13 may readily be solved by the methods presented in this paper. Some predictor-corrector scheme, or an iteration between the two equations, may be used to improve the time truncation error, but this should not affect the spatial approximations in any way.

Notice that because of the flexibility of these variational methods, the approximate pressure distribution obtained from Eq. 12 can be chosen to be at least a differentiable function. Therefore, the functions obtained from Eq. 14 will be at least continuous. That is, *no approximation to the derivatives occurring in Eq. 14 need be made.* Since some smoothness condition on the coefficients is necessary to obtain the high-order accuracy, some care must be used in selecting base functions for each equation; however, this presents no insurmountable difficulty.

CONCLUSIONS AND RECOMMENDATIONS

Some numerical methods of high-order accuracy, based on a variational approach for elliptic differential equations, are proposed for the solution of parabolic partial differential equations arising from problems describing fluid flow in porous media. Extensive testing on the equation describing the process by which one miscible liquid displaces another liquid in a one-dimensional porous medium shows:

1. The approximations are accurate over a wide range of values of the parameter λ, i.e., a wide range of values of either the dispersion coefficient, the fluid velocity, or the length of the system. The approximations are accurate even for values of $\lambda = 100{,}000$, which is a realistic number for field-size problems.

2. A variable interpolation method, which extends easily to two dimensions, is proposed. This approach allows high-order accuracy in the neighborhood of fronts and low-order accuracy in regions where the solution is smoother, keeping the over-all dimensionality of the system small.

3. Acceptable accuracy can be obtained for reasonable values of λ with these methods using from one-tenth to one-third the computing time of the other standard approaches tested. This savings in computer time should be essentially squared when going to two dimensions, resulting in savings from one to two orders of magnitude.

Following are general observations concerning these methods and recommendations for future work.

1. These methods should be applicable to a wide variety of *nonlinear* equations describing fluid flow in porous media. They should be especially useful in examining flow around a wellbore where quantitative effects are usually masked by truncation errors.

2. Convergence of these methods has been proved for a general class of nonlinear elliptic problems as well as for some simple parabolic equations. The convergence of these methods for a general class of nonlinear parabolic equations should provide a fruitful research project for the future.

3. Even larger savings in computer time than those estimated here should result from using the idea of variable interpolation for nonlinear problems. This is because reducing the total number of discrete variables in the variational formulation reduces total computational effort, in general.

NOMENCLATURE

c = volumetric concentration
D = dispersivity
D' = diffusion coefficient
E = maximum error in numerical solution, Eqs. 9 and 10
erfc = complimentary error function,

$$\text{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy$$

g = gravitational constant
h = mesh spacing
k = permeability
$K(\tau)$ = some function of τ independent of h and $\Delta\tau$
ℓ = length of system
n = number of grid intervals
p = pressure
t = time
u = flux
x = horizontal coordinate
z = vertical coordinate
α = mixing coefficient
ϕ = porosity
μ = viscosity
ρ = density
ξ = nondimensional variable $(= x/\ell)$
τ = nondimensional variable $(= Dt/\phi\ell^2)$
λ = parameter reflecting magnitude of frontal gradient in solution to Eqs. 3 and 4 $(= u\ell/D)$
$\epsilon(\tau, h, \Delta\tau)$ = terms of high order of Eq. 11

ACKNOWLEDGMENT

The authors wish to express their appreciation to the management of Gulf Research & Development Co. for permission to publish the results of this study.

REFERENCES

1. Birkhoff, G., Schultz, M. H. and Varga, R. S.: "Piecewise Hermite Interpolation in One and Two Variables With Applications to Partial Differential Equations", to appear in *Numerische Mathematik*.
2. Ciarlet, P. G., Schultz, M. H. and Varga, R. S.: "Numerical Methods of High-Order Accuracy for Nonlinear Boundary Value Problems", *Numerische Mathematik* (1967) Vol. 9, 394-430.
3. Collatz, L.: *The Numerical Treatment of Differential Equations*, 3rd Ed., Springer-Verlag, Berlin (1960) 598 pp.
4. Douglas, J., Peaceman, D. W. and Rachford, H. H.: "A Method for Calculating Multi-Dimensional Immiscible Displacement", *Trans.*, AIME (1959) Vol. 216, 297-308.
5. Garder, A. O., Jr., Peaceman, D. W. and Pozzi, A. L., Jr.: "Numerical Calculation of Multidimensional Miscible Displacement by the Method of Characteristics", *Soc. Pet. Eng. J.* (March, 1964) 26-36.
6. McFarlane, R. C., Mueller, T. D. and Miller, F. G.: "Unsteady-State Distributions of Fluid Compositions in Two-Phase Oil Reservoirs Undergoing Gas Injection", *Soc. Pet. Eng. J.* (March, 1967) 61-74.
7. Peaceman, D. W. and Rachford, H. H.: "Numerical Calculation of Multi-Dimensional Miscible Displacement", *Soc. Pet. Eng. J.* (1962) 327-338.
8. Perrine, R. L. and Gay, G. M.: "Unstable Miscible Flow in Heterogeneous Systems", *Soc. Pet. Eng. J.* (Sept., 1963) 228-238.
9. Price, H. S., Varga, R. S. and Warren, J. E.: "Application of Oscillation Matrices to Diffusion-Convection Equations", *J. of Math. and Phys.* (Sept., 1966) Vol. 45, 301-311.

10. Price, H. S. and Donohue, D. A. T.: "Isothermal Displacement Processes with Interphase Mass Transfer", *Soc. Pet. Eng. J.* (June, 1967) 205-220.

11. Shamir, U. and Harleman, D. R. F.: "Numerical Solutions for Dispersion in Porous Mediums", *Water Res. Research* (1967) Vol. 3, No. 2, 557-581.

12. Stone, H. L. and Brian, P. L. T.: "Numerical Solution of Convective Transport Problems", *AIChE Jour.* (1963) Vol. 9, No. 5, 681-688.

13. Varga, R. S.: "Hermite Interpolation-Type Ritz Methods for Two-Point Boundary Value Problems", *Numerical Solution of Partial Differential Equations* (J. H. Bramble, Ed.) Academic Press, New York (1966) 365-373.

14. Warren, J. E., Price, H. S., Skiba, F. F. and Varga, R. S.: "Miscible Displacement: The Liquid-Liquid Case", paper SPE 118, presented at SPE Annual Fall Meeting, Dallas, Tex., Oct. 8-11, 1961.

APPENDIX

In the region $R = \{(x,t)/0 < x < 1,\ 0 < t < \infty\}$, we seek a solution of the equation

$$L[u] \equiv \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} + \lambda \frac{\partial u}{\partial x} = 0, \quad (x,t) \in R, \qquad \text{. . . . . (A-1)}$$

subject to the initial-boundary conditions

$$u(x,0) = 0, \quad 0 < x \leq 1 \qquad \text{. . . . . (A-2)}$$

and

$$u(0,t) = 1.0, \quad \frac{\partial u}{\partial x}(1,t) = 0, \quad t > 0.$$

In particular, we discuss a numerical technique for approximating such a solution. (See Varga[13] and Ciarlet, Schultz and Varga[2] for a more general derivation for nonlinear elliptic equations.)

Our approach is to "discretize" first only the space variables, leaving the time variable continuous, by means of a Galerkin-type process (see, for example, Page 31, Ref. 3). The resulting system of ordinary differential equations is then discretized in the time variable to obtain a discrete approximation to Eq. A-1, Eq. A-2 which may be solved on a digital computer. In particular, we discuss here the use of high-order polynomials and/or piecewise-polynomials in the Galerkin process for obtaining high-order semidiscretizations.

Let S denote the class of all real-valued, piecewise continuously differentiable functions, $w(x)$, on $[0,1]$ such that $w(0) = 0$, $\left(\frac{dw}{dx}\right)_{x=1} = 0$. Let S_m be an m-dimensional subspace of S spanned by the m basis functions $\{w_j(x)\}_{j=1}^{m}$. We seek an approximation to the solution of Eq. A-1, Eq. A-2 in the form

$$u_m(x,t) = \sum_{k=1}^{m} C_{m,k}(t) w_k(x) + w_0(x)$$

where the coefficients $C_{m,k}(t)$ are determined by the condition that $L[u_m]$ be orthogonal (in L_2) to S_m for all $t > 0$, and that u_m satisfy the boundary conditions Eq. A-2; i.e.,

$$\int_0^1 \left\{ L[u_m] \right\} w_k dx = 0, k=1,2,\ldots,m, \quad t > 0 \qquad \text{. . . . . (A-4)}$$

and

$$\begin{cases} w_0(0) = 1.0, \ \dfrac{dw_0(1)}{dx} = 0, \\ \displaystyle\sum_{k=1}^{m} C_{m,k}(t) \left(\frac{dw_k(x)}{dx} \right)_{x=1} = 0, \end{cases} \qquad \text{. . . (A-5)}$$

and the coefficients $C_{m,k}(0)$ are uniquely determined such that

$$\left|\left| w_0(x) + \sum_{k=1}^{m} C_{m,k}(0) w_k(x) \right|\right|_{L^2} = \min_{c_i} \left|\left| w_0(x) + \sum_{k=1}^{m} c_k w_k(x) \right|\right|_{L^2}, \qquad \text{. . (A-6)}$$

Eq. A-4 takes the form

$$\int_0^1 \sum_{j=1}^{m} \left\{ C'_{m,j}(t) w_j(x) - C_{m,j}(t) w''_j(x) + \lambda C_{m,j}(t) w'_j(x) \right\} w_k(x) dx = \int_0^1 w''_0(x) w_k(x) dx - \lambda \int_0^1 w'_0(x) w_k(x) dx$$

$k = 1, 2, \ldots m$, or after integrating by parts and noting that

$$w_k(0) = \frac{dw_k(1)}{dx} = 0, \quad k=1,2,\ldots,m,$$

we have

$$\sum_{j=1}^{m} C'_{m,j}(t) \int_0^1 w_j(x) w_k(x)\,dx$$

$$+ \sum_{j=1}^{m} C_{m,j}(t) \int_0^1 w'_j(x) w'_k(x)\,dx$$

$$+ \lambda \sum_{j=1}^{m} C_{m,j}(t) \int_0^1 w'_j(x) w_k(x)\,dx =$$

$$- \int_0^1 w'_0(x)\,(\lambda w_k(x) + w'_k(x))\,dx, \quad k=1,2,\ldots m$$

. . . . . (A-7)

or

$$\begin{bmatrix} \int_0^1 w_1 w_1\,dx & \cdots & \int_0^1 w_m w_1\,dx \\ \vdots & & \vdots \\ \int_0^1 w_1 w_m\,dx & \cdots & \int_0^1 w_m w_m\,dx \end{bmatrix} \cdot \begin{bmatrix} C'_{m,1}(t) \\ \vdots \\ C'_{m,m}(t) \end{bmatrix}$$

$$+ \begin{bmatrix} \int_0^1 (w'_1 w'_1 + \lambda w'_1 w_1)\,dx & \cdots & \int_0^1 (w'_m w'_1 + \lambda w'_m w_1)\,dx \\ \vdots & & \vdots \\ \int_0^1 (w'_1 w'_m + \lambda w'_1 w_m)\,dx & \cdots & \int_0^1 (w'_m w'_m + \lambda w'_m w_m)\,dx \end{bmatrix} \cdot \begin{bmatrix} C_{m,1}(t) \\ \vdots \\ C_{m,m}(t) \end{bmatrix} =$$

$$\begin{bmatrix} -\int_0^1 w'_0(x)\,(\lambda w_1(x) + w'_1(x))\,dx \\ \vdots \\ -\int_0^1 w'_0(x)\,(\lambda w_m(x) + w'_m(x))\,dx \end{bmatrix} \quad . . . \text{(A-8)}$$

The following example using chapeau base functions should illustrate the use of Eq. A-8 to derive a semidiscretization of Eq. A-1, Eq. A-2.

The unit interval is divided into m mesh blocks of length h, and the chapeau basis functions are defined to be

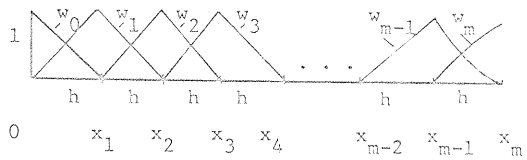
$$w_0(x) \begin{cases} = -(x-h)/h & , \ 0<x<h \\ = 0 & , \ h \le x \le 1 \end{cases}$$

$$w_i(x) \begin{cases} = (x-(i-1)h)/h & , \ (i-1)h \le x \le ih \\ = ((i+1)h-x)/h & , \ ih \le x \le (i+1)h \\ & \quad i = 1,2,\ldots, m-2 \\ = 0 & \quad \text{elsewhere} \end{cases}$$

$$w_{m-1}(x) \begin{cases} = (x-(m-2)h)/h & , \ (m-2)h \le x \le (m-1)h \\ = (mh-x)^2/h^2 & , \ (m-1)h \le x \le mh=1 \\ = 0 & \quad 0 \le x \le (m-2)h \end{cases}$$

$$w_m(x) \begin{cases} = -((x-mh)^2-h^2)/h^2, & (m-1)h \le x \le mh=1, \\ = 0 & \quad 0 \le x \le (m-1)h, \end{cases}$$

which is illustrated below,



We see, by performing the simple integrations in Eq. A-8 that this leads to the following matrix differential equation,

$$B \frac{d\underline{C}_m}{dt} = -A\,\underline{C}_m + \underline{S},$$

where

$$B = h \begin{bmatrix} \frac{2}{3} & \frac{1}{6} & & & & & \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & & & & \\ & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & & & \\ & & \frac{1}{6} & \frac{8}{15} & \frac{2}{15} & & \\ & & & \frac{2}{15} & \frac{8}{15} & & \end{bmatrix};$$

$$A=\frac{1}{h}\begin{bmatrix} 2 & -(1-\alpha) & & & & \\ -(1+\alpha) & 2 & -(1-\alpha) & & & \\ & -(1+\alpha) & 2 & -(1-\alpha) & & \\ & & -(1+\alpha) & \frac{7}{3} & -(\frac{4}{3}-\alpha) & \\ & & & -(\frac{4}{3}+\alpha) & (\frac{4}{3}+\alpha) & \end{bmatrix}$$

$$S=\frac{1}{h}\begin{bmatrix} (1+\alpha) \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \text{ and } \alpha = \frac{\lambda h}{2}.$$

Following are definitions of the other base functions used in the text.

SMOOTH CUBICS

If the unit interval is divided into $m/2$ mesh blocks of length h, the smooth cubic base functions are given by

$$w_0(x)\begin{cases} = (2x+h)(x-h)^2/h^3, & 0 \le x \le h \\ = 0, & h \le x \le 1 \end{cases}$$

$$w_1(x)\begin{cases} = x(x-h)^2/h^2, & 0 \le x \le h \\ = 0, & h \le x \le 1 \end{cases}$$

$$w_{2i}(x)\begin{cases} = (-2x+(1+2i)h)(x-(i-1)h)^2/h^3, \\ (i-1)h \le x \le ih \\ = (2x+(1-2i)h)(x-(i+1)h)^2/h^3, \\ ih \le x \le (i+1)h \\ i=1,2,\ldots,\frac{m}{2}-1 \\ = 0 \quad \text{elsewhere} \end{cases}$$

$$w_{2i+1}(x)\begin{cases} = (x-ih)(x-(i-1)h)^2/h^2, \\ (i-1)h \le x \le ih \\ = (x-ih)(x-(i+1)h)^2/h^2, \\ ih \le x \le (i+1)h \\ i=1,2,\ldots,\frac{m}{2}-1 \\ = 0 \quad \text{elsewhere} \end{cases}$$

$$w_m(x)\begin{cases} = (-2x+(1+m)h)(x-(\frac{m}{2}-1)h)^2/h^3, \\ (\frac{m}{2}-1)h \le x \le 1 \\ = 0 \quad \text{elsewhere} \end{cases}$$

NONSMOOTH QUINTICS

If the unit interval is divided into m equally spaced blocks of length h, the $5m+1$ basis functions will be the following:

$$w_0(x)\begin{cases} = (h-x)/h, & 0 \le x \le h \\ = 0, & h \le x \le 1 \end{cases}$$

$$\left.\begin{aligned} w_{5(i-1)+1}(x)&\begin{cases} = (x-(i-1)h)(x-ih), \\ (i-1)h \le x \le ih \\ = 0 \quad \text{elsewhere} \end{cases} \\ w_{5(i-1)+2}(x)&\begin{cases} = (x-(i-1)h)(x-ih)x, \\ (i-1)h \le x \le ih \\ = 0 \quad \text{elsewhere} \end{cases} \\ w_{5(i-1)+3}(x)&\begin{cases} = (x-(i-1)h)(x-ih)x^2, \\ (i-1)h \le x \le ih \\ = 0 \quad \text{elsewhere} \end{cases} \\ w_{5(i-1)+4}(x)&\begin{cases} = (x-(i-1)h)(x-ih)x^3, \\ (i-1)h \le x \le ih \\ = 0 \quad \text{elsewhere} \end{cases} \end{aligned}\right\} 1 \le i \le m$$

$$\left.\begin{aligned} w_{5i}(x)&\begin{cases} = (x-(i-1)h)/h, \\ (i-1)h \le x \le ih \\ = ((i+1)h-x)/h, \\ ih \le x \le (i+1)h \\ = 0 \quad \text{elsewhere} \end{cases} \\ w_{5m}(x)&\begin{cases} = (x-(m-1)h)/h, \\ (m-1)h \le x \le 1 \\ = 0 \quad \text{elsewhere} \end{cases} \end{aligned}\right\} 1 \le i \le m-1$$

* * *

Stresses Around Wellbores in Nonlinear Rock

M. A. MAHTAB
R. E. GOODMAN

U. OF CALIFORNIA
BERKELEY, CALIF.

ABSTRACT

The state of stress around a vertical wellbore in rock following nonlinear stress-strain laws is examined by means of finite element analysis. The wellbore is considered an axisymmetric body with axisymmetric loading. The initial vertical and horizontal stresses are "locked" in the rock elements around the wellbore and a new state of stress is generated by the displacements which occur around the borehole. A point-wise variation of the elastic moduli is made on the basis of the new stress state and the triaxial data. The initial stresses are now reintroduced along with the changed moduli and original boundary constraints. This procedure is repeated until convergent stresses are reached.

The effect of nonlinearity on stresses is examined for a 6,000-ft wellbore in a schistose gneiss and Berea sandstone using results of laboratory triaxial compression tests. The results show that the effect is restricted to one well radius from the bottom periphery of the hole. Beyond a distance of one-quarter radius, the effect of nonlinearity on stresses is almost always less than 5 percent for the cases considered. The consideration of a static pressure inside the well does not magnify the effect of nonlinearity on borehole stresses.

INTRODUCTION

The terms "wellbore" and "borehole" here designate cylindrical openings in the ground with vertical axis and a circular cross-section. A knowledge of the stress redistribution that occurs on excavating a wellbore is important in understanding (1) the behavior of the lined or unlined hole, (2) hydraulic fracture response, and (3) the effect of stress redistribution on drillability; also it is important in predicting initial stresses in the virgin ground, and in analyzing the response of measuring instruments placed in the borehole.

Our knowledge of the state of stress around a wellbore has been restricted to homogeneous, isotropic, elastic material and derives chiefly from the analysis by Miles and Topping[1] and the photoelastic work of Galle and Wilhoit[2] and Word and Wilhoit.[3] In this investigation the state of stress is examined for a nonlinear elastic material by means of finite element analysis.

Many rocks possess stress-strain curves that depart notably from straight lines in their initial or final portions. While the literature contains abundant stress-strain data from triaxial tests (axisymmetric loading) on cylindrical rock specimens, there is little information on rock deformability under nonaxisymmetric loading conditions such as occur at each point around the bottom of a wellbore. Although there is some knowledge of the effect of intermediate principal stress on rock strength, there is virtually nothing known about its effect on rock deformability; therefore, we have assumed here that the effect of intermediate principal stress can be ignored.

A schistose gneiss[4] and Berea sandstone[5] were selected as representative rocks for this analysis. The traditional graphs of deviator stress ($\sigma_1 - \sigma_3$) vs axial strain were reworked to give the tangent modulus as a function of the deviator stress for varying values of the minor principal stress. The result is a nesting family of skewed, bell-shaped curves for the gneiss (Fig. 1A) and the sandstone (Fig. 2A). A similar replotting of the lateral strain data defines the variation of Poisson's ratio (ν) with the deviator stress and confining pressure. These curves, shown in Fig. 1B for the gneiss and in Fig. 2B for the sandstone, are not so well ordered as the tangent modulus curves. However, all of these display an increase of ν with deviator stress application, but the rate of increase diminishes with confinement. The E_T and ν curves for the two rock types are tabulated in Tables 1A and 1B for use in a digital computer so that material properties corresponding to a given state of stress can be assigned by interpolation.

STATEMENT OF THE PROBLEM

Given a state of stress in the ground prior to drilling a hole (initial stresses), we assume that

Original manuscript of SPE 2005 received in Society of Petroleum Engineers office Sept. 24, 1967. Revised manuscript received June 25, 1968.

This paper will be printed in *Transactions* volume 243, which will cover 1968.

[1] References given at end of paper.