A Study of a Numerical Solution to a Two-Dimensional Hydrodynamical Problem

By A. Blair, N. Metropolis, J. von Neumann,* A. H. Taub & M. Tsingou

1. Introduction. The purpose of this paper is to report the results obtained on Maniac I when that machine was used to solve numerically a set of difference equations approximating the equations of two-dimensional motion of an incompressible fluid in Eulerian coordinates. More precisely, the problem was concerned with the two-dimensional motion of two incompressible fluids subject only to gravitational and hydrodynamical forces which at time $t = 0$ were distributed as illustrated in Fig. 1.

This problem was discussed and formulated for machine computation by John von Neumann and others. His own original draft of a discussion of the differential and difference equations is given in Appendix I, and an iteration scheme for solving systems of linear equations is given in Appendix II. In the main body of this paper we shall outline the derivation of the equations employed by the computer and refer to these appendices for detailed discussions concerning them where necessary. Some of von Neumann's difference equations were modified in the course of the work. The reasons for these modifications and their nature will be enlarged upon in the course of the discussion.

2. The Equations of Motion and Boundary Conditions. We denote by x and y the Cartesian abscissa and ordinate of a point in a fixed coordinate system in a vertical plane oriented as in Fig. 1; that is, x and y are Eulerian coordinates. The velocity of the fluid at this point at time t will be said to have x and y components $u(x, y, t)$ and $v(x, y, t)$, respectively. The density of the fluid will be denoted by ρ, the pressure by p, and the acceleration of gravity by g.

The system of equations describing the motion of an incompressible fluid subject to the force of gravity in the vertical direction is then

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial x} \tag{2.1}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial y} + g \tag{2.2}$$

$$\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} = 0 \tag{2.3}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{2.4}$$

Equations (2.1) and (2.2) represent the conservation of momentum, equation (2.3) is the incompressibility condition, and equation (2.4) states the conservation of mass.

Received April 19, 1959. This work was done at Los Alamos Scientific Laboratory and the paper is a condensation of Report LA-2165 with the same title.

* Published posthumously.

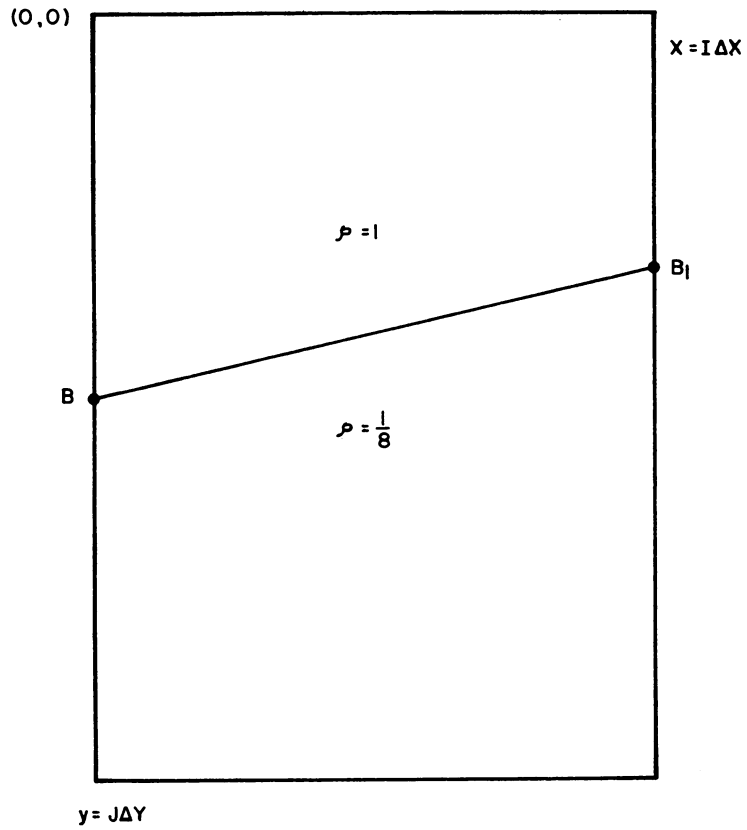

FIG. 1.—Initial density distribution. BB_1 denotes the boundary separating the fluid of density $\rho = 1$ from that of density $\rho = \frac{1}{8}$ at time $t = 0$.

At any exterior boundary of the fluid the component of velocity normal to the boundary vanishes. That is

$$u \cos(x, n) + v \cos(y, n) = 0 \tag{2.5}$$

where $\cos(x, n)$ and $\cos(y, n)$ are the cosines of the angles between the x axis and the normal to the boundary and the y axis and the normal to the boundary, respectively.

Along a curve across which there is a density discontinuity we must have the component of the velocity normal to the curve continuous. That is, we must have

$$[u] \cos(x, n) + [v] \cos(y, n) = 0 \tag{2.6}$$

where

$$[f] = f(x^+, y^+) - f(x^-, y^-)$$

and x^+, y^+ and x^-, y^- represent contiguous points on opposite sides of the curve of discontinuities.

3. Discontinuities. The only discontinuities present in incompressible fluid motion are density discontinuities and across these equation (2.6) must hold. In

the calculation to be described, such discontinuities are not explicitly taken into account. Indeed, it was one purpose of the calculation to see if this type of discontinuity could be followed in time from a plot of the density contours when no special provision was made to provide for the discontinuities.

Initially the density distribution assumed was that given in Fig. 1. Some provisions must be made in the difference equations to represent the spatial derivatives of the density on the curve BB_1 at time $t = 0$ (and at subsequent times on the curve into which BB_1 moves). We shall discuss this point subsequently.

4. The Stream Function. It follows from equation (2.4) that there exists a function ψ called the stream function such that

$$u = -\psi_y, \qquad v = \psi_x \tag{4.1}$$

On an exterior boundary, equation (2.5) obtains and this may be written as

$$-\psi_y \cos(x, n) + \psi_x \cos(y, n) = 0 \tag{4.2}$$

If the equation of the boundary is given parametrically by

$$x = x(s), \qquad y = y(s)$$

with

$$\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 = 1$$

then the direction cosines of the normal are

$$\cos(x, n) = -\frac{dy}{ds}$$

$$\cos(y, n) = \frac{dx}{ds}$$

and equation (4.2) becomes

$$\psi_x \frac{dx}{ds} + \psi_y \frac{dy}{ds} = 0$$

That is

$$\psi = \text{constant}$$

on the exterior boundary. This constant may be chosen to be zero and the exterior boundary condition becomes

$$\psi = 0 \tag{4.3}$$

5. Equations Determining the Stream Function and Density. Substituting from equation (4.1) into equations (2.1) and (2.2), we obtain

$$\rho(\psi_{ty} - \psi_y\psi_{xy} + \psi_x\psi_{yy}) = p_x$$

$$\rho(\psi_{tx} - \psi_y\psi_{xx} + \psi_x\psi_{xy}) = -p_y + g\rho$$

where the subscript denotes a derivative with respect to the variable indicated. Differentiating the first of these with respect to y and the second with respect to x and adding, we have

$$(\rho\psi_{tx})_x + (\rho\psi_{ty})_y = \rho_x(g - {}^1I) - \rho_y{}^2I - \rho^3I = -\omega \tag{5.1}$$

where

$${}^1I = \psi_x\psi_{xy} - \psi_y\psi_{xx} \tag{5.2}$$

$${}^2I = \psi_x\psi_{yy} - \psi_y\psi_{xy} \tag{5.3}$$

$${}^3I = \psi_x\lambda_y - \lambda_x\psi_y \tag{5.4}$$

with

$$\lambda = \psi_{xx} + \psi_{yy} \tag{5.5}$$

If we now set

$$\chi = -\psi_t \tag{5.6}$$

then

$$-(\rho\chi_x)_x - (\rho\chi_y)_y = -\omega \tag{5.7}$$

where ω is given in terms of ψ and ρ by the right-hand side of equation (5.1), and on the boundary

$$\chi = 0 \tag{5.8}$$

We may regard equation (5.1) and the boundary condition $\psi = 0$ as equations for determining ψ and use equation (4.1) to define u. The pressure may then be calculated from equations (2.1) or (2.2).

The density may then be determined from equation (2.3), which may be written as

$$\rho_t = \psi_y\rho_x - \psi_x\rho_y \tag{5.9}$$

The system of equations (5.6) through (5.9) defines the problem to be approximated by difference equations and to be solved numerically on the computer.

6. The Difference Equations. For any function $a(x, y, t)$ let

$$a^h_{i,j} = a(x_i, y_j, t^h) \tag{6.1}$$

where

$$\begin{aligned} x_i &\equiv i\Delta x & i &= 0, 1, \cdots I - 1, I \\ y_j &\equiv j\Delta y & j &= 0, 1, \cdots J - 1, J \\ t^h &\equiv h\Delta t & h &= 0, 1, 2, \cdots \end{aligned} \tag{6.2}$$

and I and J are integers. Occasionally indices $i \pm \frac{1}{2}$, $j \pm \frac{1}{2}$, and $h + \frac{1}{2}$ are used.

The difference equations we shall consider will involve quanties $\psi^h_{i,j}$, $\chi^h_{i,j}$, and $\rho^h_{i,j}$ for h, i, and j given by equation (6.2). The mesh points with $i = 0$ or I ($j = 0, 1, \cdots J$) or $j = 0$ or J ($i = 0, 1, \cdots I$) are said to be boundary points.

The remaining mesh points are called interior points. Equations (4.3) and (5.8) define $\psi^h_{i,j} = \chi^h_{i,j} = 0$ for boundary points. Hence, we must give an algorithm for determining $\psi^h_{i,j}$ for interior points and $\rho^h_{i,j}$ for interior and boundary points. This algorithm involves the finite difference representation of equation (5.7) for interior points and equation (5.9) for all points.

Equation (5.7) is replaced by

$$\begin{aligned}(6.3)\qquad &\frac{1}{(\Delta x)^2}\left[-\rho^h_{i+\frac{1}{2},j}\,(\chi^h_{i+1,j}-\chi^h_{i,j})+\rho^h_{i-\frac{1}{2},j}\,(\chi^h_{i,j}-\chi^h_{i-1,j})\right]\\ &+\frac{1}{(\Delta y)^2}\left[-\rho^h_{i,j+\frac{1}{2}}\,(\chi^h_{i,j+1}-\chi^h_{i,j})+\rho^h_{i,j-\frac{1}{2}}\,(\chi^h_{i,j}-\chi^h_{i,j-1})\right]=-\omega^h_{i,j}\end{aligned}$$

for interior points, that is, for

$$1 \leqq i \leqq I-1$$
$$1 \leqq j \leqq J-1$$

where

$$\begin{aligned}(6.4)\qquad 2\rho^h_{i\pm\frac{1}{2},j} &= \rho^h_{i\pm1,j}+\rho^h_{i,j}\\ 2\rho^h_{i,j\pm\frac{1}{2}} &= \rho^h_{i,j\pm1}+\rho^h_{i,j}\end{aligned}$$

and $\omega^h_{i,j}$ is formed from $\psi^h_{i,j}$ and $\rho^h_{i,j}$ as indicated in equation (A.14) of Appendix I.

Equation (5.6) is replaced by the equation

$$(6.5)\qquad \psi^{h+1}_{i,j} = \psi^{h-1}_{i,j} - 2\Delta t\chi^h_{i,j}$$

for

$$h > 0$$

and by

$$(6.6)\qquad \psi^1 = \psi^0_{i,j} - \Delta t\chi^0_{i,j}$$

when

$$h = 0$$

Equation (5.9) is replaced by the equation

$$\begin{aligned}(6.7)\qquad \rho^{h+1}_{i,j} = \rho^h_{i,j} &+ \left(\frac{\Delta t}{\Delta x}\right)\begin{cases}(\rho^h_{i,j}-\rho^h_{i-1,j}) & \text{if } (\psi_y)^{h+\frac{1}{2}}_{i,j}<0\\ (\rho^h_{i+1,j}-\rho^h_{i,j}) & \text{if } (\psi_y)^{h+\frac{1}{2}}_{i,j}\geqq 0\end{cases}(\psi_y)^{h+\frac{1}{2}}_{i,j}\\ &-\left(\frac{\Delta t}{\Delta y}\right)\begin{cases}(\rho^h_{i,j}-\rho^h_{i,j-1}) & \text{if } (\psi_x)^{h+\frac{1}{2}}_{i,j}\geqq 0\\ (\rho^h_{i,j+1}-\rho^h_{i,j}) & \text{if } (\psi_x)^{h+\frac{1}{2}}_{i,j}<0\end{cases}(\psi_x)^{h+\frac{1}{2}}_{i,j}\end{aligned}$$

for interior points, where

$$2(\psi_x)^{h+\frac{1}{2}}_{i,j} = (\psi_x)^{h+1}_{i,j}+(\psi_x)^h_{i,j}$$

and a similar equation defines $(\psi_y)^{h+\frac{1}{2}}_{i,j}$.

On a boundary such as $j = J$, $(\psi_x)^h_{i,j} = 0$ from the boundary conditions, and if initially $\rho^0_{i,J}$ = constant, it will follow from equation (6.7) that $\rho^h_{i,J} = \rho^0_{i,J}$.

That is, the density discontinuity will not be able to reach the boundary $j = J$. For this reason equation (6.7) does not seem to be a suitable one for determining the time behavior of the density at the boundaries. It was replaced by

$$\rho_{i,j}^{h+1} = \rho_{i,j}^{h} + \frac{\Delta t}{\Delta x}\begin{cases}\rho_{i,j}^{h}\,(\psi_y)_{i,j}^{h+\frac{1}{2}} - \rho_{i-1,j}^{h}\,(\psi_y)_{i-1,j}^{h+\frac{1}{2}} & \text{if } (\psi_y)_{i,j}^{h+\frac{1}{2}} < 0 \\ \rho_{i+1}^{h}\,(\psi_y)_{i+1,j}^{h+\frac{1}{2}} - \rho_{i,j}^{h}\,(\psi_y)_{i,j}^{h+\frac{1}{2}} & \text{if } (\psi_y)_{i,j}^{h+\frac{1}{2}} \geqq 0\end{cases} \tag{6.8}$$
$$- \frac{\Delta t}{\Delta y}\begin{cases}\rho_{i,j}^{h}\,(\psi_x)_{i,j}^{h+\frac{1}{2}} - \rho_{i,j-1}^{h}\,(\psi_x)_{i,j-1}^{h+\frac{1}{2}} & \text{if } (\psi_x)_{i,j}^{h+\frac{1}{2}} \geqq 0 \\ \rho_{i,j+1}^{h}\,(\psi_x)_{i,j+1}^{h+\frac{1}{2}} - \rho_{i,j}^{h}\,(\psi_x)_{i,j}^{h+\frac{1}{2}} & \text{if } (\psi_x)_{i,j}^{h+\frac{1}{2}} < 0\end{cases}$$

for boundary points. This equation is a finite difference representation of

$$\rho_t = -(\rho u)_x - (\rho v)_y$$

just as equation (6.7) is of equation (5.9).

Equation (6.7) differs from the finite difference form of equation (5.9) proposed by von Neumann, namely

$$\rho_{i,j}^{h+1} = \rho_{i,j}^{h} + \Delta t(\rho_t)_{i,j}^{h} + \frac{(\Delta t)^2}{2}(\rho_{tt})_{i,j}^{h} \tag{6.9}$$

where $(\rho_t)_{i,j}^{h}$ and $(\rho_{tt})_{i,j}^{h}$ are evaluated from the values of $\rho_{i,j}^{h}$, $\psi_{i,j}^{h}$ and $\chi_{i,j}^{h}$ by substituting into the centered finite difference representation of equation (5.9) and the equation obtained by differentiating this equation with respect to t and substituting for ρ_t from (5.9) and χ for $-\psi_t$.

In the early calculations the finite difference form of equation (6.9) was used. However, it was found that near a discontinuity in the density ρ the values of ρ increased on the high side of the discontinuity and decreased on the low side, thus steadily increasing the size of the discontinuity. This unstable behavior did not occur when equation (6.7) was used.

7. The solution of Equation (6.3). This equation is of the form of a set of linear equations which may be written as

$$A\chi = \omega \tag{7.1}$$

where χ is an unknown $M[= (I-1)(J-1)]$ dimensional vector, ω is a known vector of this many dimensions, and A is a known $M \times M$ matrix.

In Appendix II von Neumann discusses iteration schemes for solving these equations. He concludes that if A is a positive (or negative) definite matrix [the matrix A of equation (6.3) is negative definite] with a largest proper value less than or equal to b and a smallest one greater than or equal to a, then the "best" (in the sense defined in Appendix II) iteration scheme is given by the equation

$$\eta^{k+1} = 2b_{k+1}\left[\eta^k - \eta^{k-1} + \frac{2}{a+b}(\omega - A\eta^k)\right] + \eta^{k-1} \tag{7.2}$$

where

$$b_1 = 1$$
$$b_{k+1} = \frac{1}{2 - (1-\epsilon)^2 b_k} \tag{7.3}$$

and

$$\epsilon = \frac{2a}{a+b} \tag{7.4}$$

In order to apply this scheme, bounds for the lowest and highest proper values, the numbers a and b, must be determined.

Using the equations given by von Neumann in Section 15 of Appendix II, we may set

$$\begin{aligned} a &= 4\bar{a}\left(\frac{1}{(\Delta x)^2}\sin^2\frac{\pi}{2I} + \frac{1}{(\Delta y)^2}\sin^2\frac{\pi}{2J}\right) \\ b &= 4\bar{b}\left(\frac{1}{(\Delta x)^2}\cos^2\frac{\pi}{2I} + \frac{1}{(\Delta y)^2}\cos^2\frac{\pi}{2J}\right) \end{aligned} \tag{7.5}$$

where $\bar{a}$ and $\bar{b}$ are lower and upper bounds, respectively, of the density. That is

$$0 < \bar{a} \leqq \rho_{i,j} \leqq \bar{b} \tag{7.6}$$

for all i and j.

8. The Flow Diagram. A condensed flow diagram is reproduced as Fig. 2. Before operation, initial values of $\psi_{i,j}$ and $\rho_{i,j}$ at time $h = 0$ are stored. The values of $\rho_{i,j}$ were prepared as follows: If i, j labels a mesh point which does not lie on the density discontinuity, the value $\rho_{i,j} = \rho(x_i, y_j)$. If x_i, y_j lies on the density discontinuity, we define

$$\rho_{i,j} = \tfrac{1}{2}[\rho(x_i, y_{j+1}) + \rho(x_i, y_{j-1})]$$

$\psi_{i,j}$ was taken to be zero initially. When the routine is started, part A is traversed, which sets $h = 0$ and optionally prints out the initial values of ψ and ρ in a format uniform with subsequent results.

Part B computes the values of $\omega^h_{i,j}$, the right side of equation (5.1), for all values of i, j corresponding to interior points, as needed in equation (7.1). The box with the sole notation i, j indicates an induction loop repeatedly using the program of the box to right of it for all appropriate values of i, j. All derivatives in the formula for $\omega^h_{i,j}$ are computed by taking the difference of the values of the function at lattice points on each side, for example

$$(\psi_x)_{i,j} = \frac{1}{2\Delta x}(\psi_{i+1,j} - \psi_{i-1,j})$$

except in certain cases near the boundary where one of these quantities does not exist, and then a one-sided derivative, e.g.,

$$\frac{1}{\Delta x}(\psi_{i+1,j} - \psi_{i,j}) \qquad \text{for} \quad i = 0, j = 0, 1, 2, \cdots, J$$

is used.

Part C computes η^0, the first term in the sequence of vectors η^k to be constructed converging to χ. If $h = 0$, then η^0 is made zero, which is a reasonable estimate in the case where the liquid starts moving from rest. After the motion has proceeded

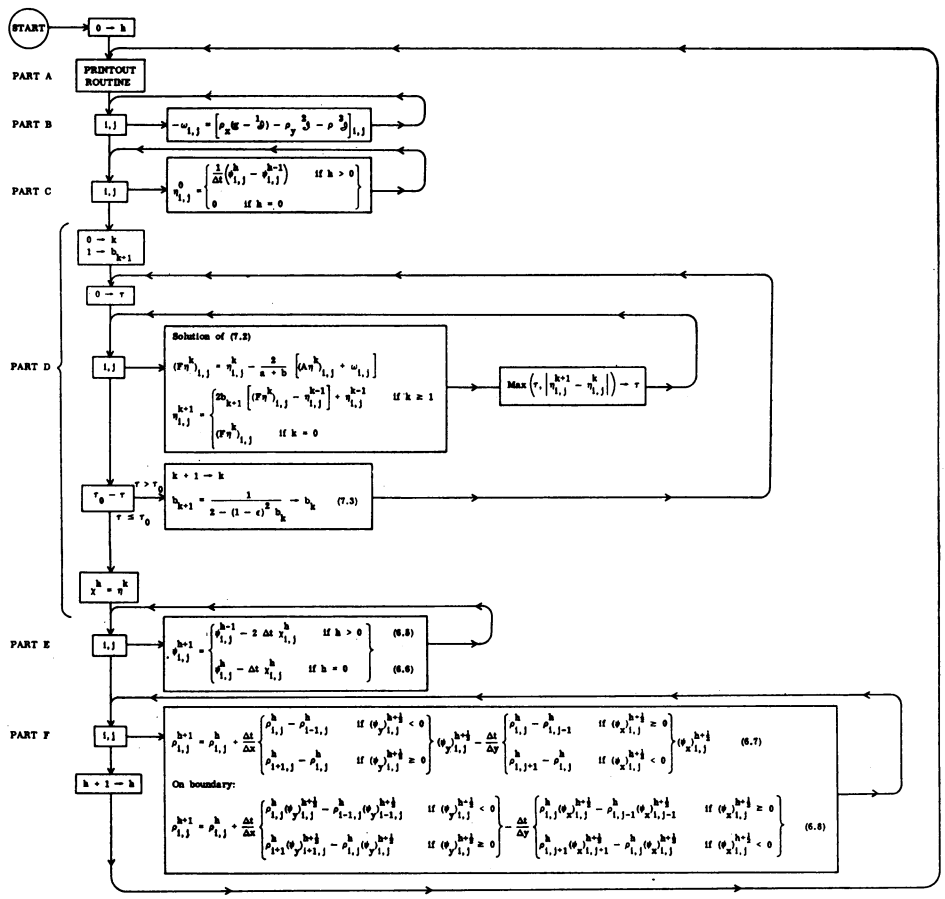

FIG. 2.—Condensed flow diagram.

one or more time intervals, and so $h > 0$, then η^0 is given the value

$$(1/\Delta t)(\psi^h - \psi^{h-1})$$

which is approximately $\chi^{h-\frac{1}{2}}$, a good start on a sequence to approach χ^h.

Part D solves equation (6.3) by means of the iteration and mean procedure characterized by equations (7.2), (7.3), and (7.4). There is an i, j induction loop inside a larger k loop. For each value of k the vector η^{k+1} with components $\eta^{k+1}_{i,j}$ is computed by equation (7.2), and $\max(\eta^{k+1}_{i,j} - \eta^k_{i,j})$ is computed. If this maximum is above a predetermined constant τ_0, then k is increased by 1, b_{k+1} is computed from equation (7.3), and the i, j induction loop is entered again to compute the next term in the sequence of η^k values. When a vector η^k is obtained which is sufficiently close to η^{k-1}, then it is considered to be χ^h, and the control passes to part E.

Part E computes the new values ψ^{h+1} by equations (6.5) and (6.6), and Part F computes the new values ρ^{h+1} by equations (6.7) and (6.8). The time index h is then increased by 1, the results are optionally printed, and the control passes again to Part B.

The finite difference equations used in the numerical computation according to the flow diagram are:

PART A. Part A has no formulae.

PART B. Part B is an induction loop for computing all interior points of $\omega^h_{i,j}$ as a function of $\psi^h_{i,j}$ and $\rho^h_{i,j}$. The formulae used are

$$2(\Delta x)(\psi_x)^h_{i,j} = \psi^h_{i+1,j} - \psi^h_{i-1,j} \tag{B.1}$$

$$2(\Delta y)(\psi_y)^h_{i,j} = \psi^h_{i,j+1} - \psi^h_{i,j-1} \tag{B.2}$$

$$4(\Delta x)(\Delta y)(\psi_{xy})^h_{i,j} = \psi^h_{i+1,j+1} - \psi^h_{i-1,j+1} - \psi^h_{i+1,j-1} + \psi^h_{i-1,j-1} \tag{B.3}$$

The next three formulae are computed as a five-cycle induction loop for $(i', j') = (i, j), (i + 1, j), (i - 1, j), (i, j + 1), (i, j - 1)$

$$(\Delta x)^2(\psi_{xx})^h_{i',j'} = \psi^h_{i'+1,j'} - 2\psi^h_{i',j'} + \psi^h_{i'-1,j'} \tag{B.4}$$

$$(\Delta y)^2(\psi_{yy})^h_{i',j'} = \psi_{i',j'+1} - 2\psi^h_{i',j'} + \psi^h_{i',j'-1} \tag{B.5}$$

$$(\Delta x)^2\lambda^h_{i',j'} = (\Delta x)^2(\psi_{xx})^h_{i',j'} + a(\Delta y)^2(\psi_{yy})^h_{i',j'} \tag{B.6}$$

where $a = (\Delta x/\Delta y)^2$.

$$2(\Delta x)^3(\lambda_x)^h_{i,j} = (\Delta x)^2\lambda^h_{i+1,j} - (\Delta x)^2\lambda^h_{i-1,j} \quad \text{if} \quad i \neq 1, I - 1 \tag{B.7a}$$

$$(\Delta x)^3(\lambda_x)^h_{i,j} = (\Delta x)^2\lambda^h_{i+1,j} - (\Delta x)^2\lambda^h_{i,j} \quad \text{if} \quad i = 1 \tag{B.7b}$$

$$(\Delta x)^3(\lambda_x)^h_{i,j} = (\Delta x)^2\lambda^h_{i,j} - (\Delta x)^2\lambda^h_{i-1,j} \quad \text{if} \quad i = I - 1 \tag{B.7c}$$

$$2(\Delta x)^2(\Delta y)(\lambda_y)^h_{i,j} = (\Delta x)^2\lambda^h_{i,j+1} - (\Delta x)^2\lambda^h_{i,j-1} \quad \text{if} \quad j \neq 1, J - 1 \tag{B.8a}$$

$$(\Delta x)^2(\Delta y)(\lambda_y)^h_{i,j} = (\Delta x)^2\lambda^h_{i,j+1} - (\Delta x)^2\lambda^h_{i,j} \quad \text{if} \quad j = 1 \tag{B.8b}$$

$$(\Delta x)^2(\Delta y)(\lambda_y)^h_{i,j} = (\Delta x)^2\lambda^h_{i,j} - (\Delta x)^2\lambda^h_{i,j-1} \quad \text{if} \quad j = J - 1 \tag{B.8c}$$

$$2(\Delta x)(\rho_x)^h_{i,j} = \rho^h_{i+1,j} - \rho^h_{i-1,j} \tag{B.9}$$

$$2(\Delta y)(\rho_y)^h_{i,j} = \rho^h_{i,j+1} - \rho^h_{i,j-1} \tag{B.10}$$

$$8(\Delta x)^2(\Delta y)^1I^h_{i,j} = [2(\Delta x)(\psi_x)^h_{i,j}][4(\Delta x)(\Delta y)(\psi_{xy})^h_{i,j}] - 4[2(\Delta y)(\psi_y)^h_{i,j}][(\Delta x)^2(\psi_{xx})^h_{i,j}] \tag{B.11}$$

$$8(\Delta x)(\Delta y)^{2\,2}I^h_{i,j} = 4[2(\Delta x)(\psi_x)^h_{i,j}][(\Delta y)^2(\psi_{yy})^h_{i,j}] - [2(\Delta y)(\psi_y)^h_{i,j}][4(\Delta x)(\Delta y)(\psi_{xy})^h_{i,j}] \tag{B.12}$$

$$4(\Delta x)^3(\Delta y)^3I^h_{i,j} = [2(\Delta x)(\psi_x)^h_{i,j}][2(\Delta x)^2(\Delta y)(\lambda_y)^h_{i,j}] - [2(\Delta y)(\psi_y)^h_{i,j}][2(\Delta x)^3(\lambda_x)^h_{i,j}] \tag{B.13}$$

$$16(\Delta x)^3(\Delta y)\omega^h_{i,j} = [2(\Delta x)(\rho_x)^h_{i,j}][8(\Delta x)^2(\Delta y)^1I^h_{i,j} - 8(\Delta x)^2(\Delta y)g] + a[2(\Delta y)(\rho_y)^h_{i,j}][8(\Delta x)(\Delta y)^{2\,2}I^h_{i,j}] + [4\rho^h_{i,j}][4(\Delta x)^3(\Delta y)^3I^h_{i,j}] \tag{B.14}$$

where

$$a = \left(\frac{\Delta x}{\Delta y}\right)^2$$

PART C. Part C is an induction loop, with respect to i, j of all lattice points to compute the first trial value, $\eta^0_{i,j}$, at cycle h for use in the iteration process. The formula is

$$\text{(C.1)} \qquad \Delta x \Delta y\, \eta^0_{i,j} = \begin{cases} \dfrac{(\Delta x)(\Delta y)}{\Delta t}\,(\psi^{h-1}_{i,j} - \psi^h_{i,j}) & \text{if } h > 0 \\ 0 & \text{if } h = 0 \end{cases}$$

PART D. Part D is an induction loop with respect to k, with a smaller i, j induction loop for each k. This is the solution of the difference equation by the iterative and mean method. The actual formulae are

$$\text{(D.1)} \qquad 2\rho^h_{i+\frac{1}{2},j} = \rho^h_{i,j} + \rho^h_{i+1,j}$$

$$\text{(D.2)} \qquad 2\rho^h_{i-\frac{1}{2},j} = \rho^h_{i,j} + \rho^h_{i-1,j}$$

$$\text{(D.3)} \qquad 2a\rho^h_{i,j+\frac{1}{2}} = a(\rho^h_{i,j} + \rho^h_{i,j+1})$$

As before

$$a = \left(\frac{\Delta x}{\Delta y}\right)^2$$

$$\text{(D.4)} \qquad 2a\rho^h_{i,j-\frac{1}{2}} = a(\rho^h_{i,j} + \rho^h_{i,j-1})$$

$$\text{(D.5)} \qquad \sigma^h_{i,j} = 2\rho^h_{i+\frac{1}{2},j} + 2\rho^h_{i-\frac{1}{2},j} + 2a\rho^h_{i,j+\frac{1}{2}} + 2a\rho^h_{i,j-\frac{1}{2}}$$

$$\begin{aligned} -2(\Delta x)^3(\Delta y)(A\eta^k)_{i,j} &= [2\rho^h_{i+\frac{1}{2},j}][(\Delta x)(\Delta y)\eta^k_{i+1,j}] \\ &+ [2\rho^h_{i-\frac{1}{2},j}][(\Delta x)(\Delta y)\eta^k_{i-1,j}] + [2a\rho^h_{i,j+\frac{1}{2}}][(\Delta x)(\Delta y)\eta^k_{i,j+1}] \\ &+ [2a\rho^h_{i,j-\frac{1}{2}}][(\Delta x)(\Delta y)\eta^k_{i,j-1}] - \sigma_{i,j}[(\Delta x)(\Delta y)\eta^k_{i,j}] \end{aligned} \qquad \text{(D.6)}$$

$$\begin{aligned} (\Delta x)(\Delta y)(F\eta^k)_{i,j} = [(\Delta x)(\Delta y)\eta^k_{i,j}] + \frac{1}{2}\left[\frac{a}{(\Delta x)^2}\right]([-2(\Delta x)^3(\Delta y)(A\eta^k)_{i,j}] \\ - \frac{1}{8}[16(\Delta x)^3(\Delta y)\omega^h_{i,j}]) \end{aligned} \qquad \text{(D.7)}$$

where $d \equiv \dfrac{\epsilon}{a}$, cf. Eq. (7.4).

$$\text{(D.8)} \qquad (\Delta x)(\Delta y)\eta^{k+1}_{i,j} = \begin{cases} 2b_{k+1}\{[(\Delta x)(\Delta y)(F\eta^k)_{i,j}] - [(\Delta x)(\Delta y)\eta^{k-1}_{i,j}]\} \\ \qquad + [(\Delta x)(\Delta y)\eta^{k-1}_{i,j}] \quad \text{if } k = 0 \\ (\Delta x)(\Delta y)(F\eta^k)_{i,j} \quad \text{if } k = 0 \end{cases}$$

$$\text{(D.9)} \qquad b_1 = 1 \qquad \text{for } k = 0$$

$$\text{(D.10)} \qquad b_{k+1} = \frac{1}{2 - (1-\epsilon)^2 b_k} \qquad \text{for } k > 0$$

PART E. Part E is an induction loop with respect to i, j and is used to compute $\psi^{h+1}_{i,j}$ for all points as a function of $\psi^{h-1}_{i,j}$ and $\chi^h_{i,j}$

$$\text{(E.1a)} \qquad \psi^{h+1}_{i,j} = \psi^{h-1}_{i,j} - \frac{2\Delta t}{(\Delta x)(\Delta y)}[(\Delta x)(\Delta y)\chi^h_{i,j}] \qquad \text{if } h > 0$$

(E.1b) $$\psi_{i,j}^{h+1} = \psi_{i,j}^{h} - \frac{\Delta t}{(\Delta x)(\Delta y)}\,[(\Delta x)(\Delta y)\chi_{i,j}^{h}] \qquad \text{if} \quad h = 0$$

PART F. Part F is an induction loop with respect to i, j and is used to compute $\rho_{i,j}^{h+1}$ as a function of $\psi_{i,j}^{h}$, $\psi_{i,j}^{h+1}$, and $\rho_{i,j}^{h}$.

(F.1a)
$$\rho_{i,j}^{h+1} = \rho_{i,j}^{h} + \frac{\Delta t}{\Delta x}\left\{\begin{array}{ll}(\rho_{i,j}^{h} - \rho_{i-1,j}^{h}) & \text{if} \quad (\psi_y)_{i,j}^{h+\frac{1}{2}} < 0 \\ (\rho_{i+1,j}^{h} - \rho_{i,j}^{h}) & \text{if} \quad (\psi_y)_{i,j}^{h+\frac{1}{2}} \geqq 0\end{array}\right\}(\psi_y)_{i,j}^{h+\frac{1}{2}}$$
$$- \frac{\Delta t}{\Delta y}\left\{\begin{array}{ll}(\rho_{i,j}^{h} - \rho_{i,j-1}^{h}) & \text{if} \quad (\psi_x)_{i,j}^{h+\frac{1}{2}} \geqq 0 \\ (\rho_{i,j+1}^{h} - \rho_{i,j}^{h}) & \text{if} \quad (\psi_x)_{i,j}^{h+\frac{1}{2}} < 0\end{array}\right\}(\psi_x)_{i,j}^{h+\frac{1}{2}}$$
$$\text{for} \quad \begin{array}{l} i = 0, 1, \cdots, I \\ j = 1, 2, \cdots, J-1 \end{array}$$

(F.1b)
$$\rho_{i,j}^{h+1} = \rho_{i,j}^{h} + \frac{\Delta t}{\Delta x}\left\{\begin{array}{ll}\rho_{i,j}^{h}(\psi_y)_{i,j}^{h+\frac{1}{2}} - \rho_{i-1,j}^{h}(\psi_y)_{i-1,j}^{h+\frac{1}{2}} & \text{if} \quad (\psi_y)_{i,j}^{h+\frac{1}{2}} < 0 \\ \rho_{i+1,j}^{h}(\psi_y)_{i+1,j}^{h+\frac{1}{2}} - \rho_{i,j}^{h}(\psi_y)_{i,j}^{h+\frac{1}{2}} & \text{if} \quad (\psi_y)_{i,j}^{h+\frac{1}{2}} \geqq 0\end{array}\right\}$$
$$- \frac{\Delta t}{\Delta y}\left\{\begin{array}{ll}\rho_{i,j}^{h}(\psi_x)_{i,j}^{h+\frac{1}{2}} - \rho_{i,j-1}^{h}(\psi_x)_{i,j-1}^{h+\frac{1}{2}} & \text{if} \quad (\psi_x)_{i,j}^{h+\frac{1}{2}} \geqq 0 \\ \rho_{i,j+1}^{h}(\psi_x)_{i\,j+1}^{h+\frac{1}{2}} - \rho_{i,j}^{h}(\psi_x)_{i,j}^{h+\frac{1}{2}} & \text{if} \quad (\psi_x)_{i,j}^{h+\frac{1}{2}} < 0\end{array}\right\}$$
$$\text{for} \quad \begin{array}{l} i = 1, 2, \cdots, I-1 \\ j = 0, J \end{array}$$

9. Stability and the Choice of Δt. The behavior of the solutions of equation (6.7) with regard to stability is similar to that of the corresponding equation in one dimension with a constant velocity of propagation which will be taken to be positive. Then the equation of conservation of mass becomes

$$\rho_t = -u\rho_x$$

which has the finite difference representation

$$\rho_j^{h+1} = \rho_j^{h} - \frac{\Delta t}{\Delta x}\,u(\rho_j^{h} - \rho_{j-1}^{h})$$

or

(9.1) $$\rho_j^{h+1} = (1 - \alpha)\rho_j^{h} + \alpha\rho_{j-1}^{h}$$

where

(9.2) $$\alpha = \frac{u\Delta t}{\Delta x}$$

Equation (9.1) has solutions of the form

(9.3) $$\rho_j^{h} = \exp\left[\frac{i\pi\eta}{J}\,(j\Delta x - \beta h\Delta t)\right]$$

where

$$\exp\left[-\frac{i\pi\eta}{J}\,\beta\Delta t\right] = 1 + \alpha\exp\left[-\frac{i\pi\eta}{J}\,\Delta x - 1\right]$$

and hence

$$\left|\exp\left(-\frac{i\pi\eta}{J}\beta\Delta t\right)\right|^2 = 1 - 4\alpha(1-\alpha)\cos^2\pi\frac{\eta\Delta x}{2J} \tag{9.4}$$

Thus β will be real, and the equation will be stable if and only if

$$\alpha \leqq 1$$

That is, if

$$\Delta t \leqq \frac{\Delta x}{u}$$

The value of Δt in the two-dimensional calculations was chosen so that

$$|\psi_y| \leqq \frac{\Delta x}{\Delta t}$$

and

$$|\psi_x| \leqq \frac{\Delta y}{\Delta t}$$

The units used were such that a change in Δt could be accomplished by changing the value of the constant representing g the acceleration of gravity, and scaling ψ.

It follows from equation (9.4) that up to second-order terms in Δx

$$\beta = u\left[1 - i\left(1 - u\frac{\Delta t}{\Delta x}\right)\frac{\pi\eta}{2J}\Delta x\right] \tag{9.5}$$

Hence the elementary solutions of equation (9.1) of the form of (9.3) may be written as

$$\rho_j = \exp\left[\frac{i\pi\eta}{J}(j\Delta x - h\Delta t u)\right]\exp\left[-j\left(\frac{\pi\eta}{J}\right)^2\Delta x^2 h\alpha(1-\alpha)\right] \tag{9.6}$$

The first factor of this expression may be written as

$$\exp\left[\frac{i\pi\eta}{J}(x - ut)\right]$$

with $x = j\Delta x$ and $t = h\Delta t$. This is a solution of the differential equation. Thus the second factor on the right-hand side of equation (9.6) shows how each of these elementary solutions of the differential equation is distorted when that equation is replaced by the finite difference equation (9.1).

Von Neumann in a personal communication to S. Ulam (cf. Appendix III of reference 1) has used the term "pseudo-diffusion" for the distortion associated with an initial distribution of density

$$\rho_0(x) = \begin{cases} 1 & \text{for } x \geqq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

He has shown that if this function is taken as an initial condition for the difference equation (9.1), then the 0 values of ρ advance and the 1 values of ρ recede to the

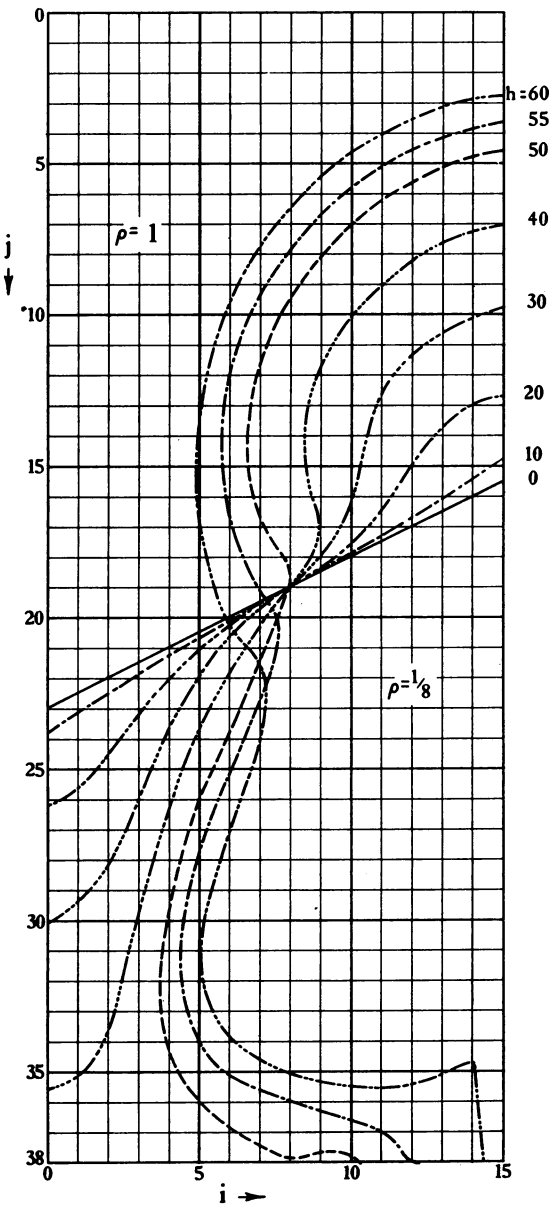

FIG. 3.—Gravity flow of two incompressible inviscid fluids at various cycles h.

right with a velocity

$$\frac{\delta x}{\delta t} = u$$

and at the same time a pseudo-diffusion-mixing region forms around the interface of advance whose width is essentially measured by

$$\delta x \sim \sqrt{(1 - \alpha)x \cdot \Delta x}$$

Since α was made close to one by the choice of Δt, the region of pseudo-diffusion was small for the calculations reported here.

10. Computation Time. The results reported in Section 11 of this paper were run on Maniac I with $I = 15$ and $J = 38$, that is, with 624 lattice points (518 interior points). The program required 300 words (600 orders of code exclusive of print routines and exclusive of orders necessary for moving information to and from the magnetic drum because of the limited electrostatic storage capacity of Maniac I). About 3750 words of dynamic storage were required.

The time required for running one time cycle of the program on Maniac I was 18 seconds for each iteration cycle plus 100 seconds for all the rest of the program. The iteration process converges so as to give accuracy in an additional decimal place every 10 minutes, so that one time cycle requires about an hour for six-place accuracy (about 200 iterations) or a half hour for three-place accuracy (about 100 iterations). About 40 per cent of this time, however, is used in transfers to and from the magnetic drum, so that this much time is to be charged to the fact that a 4000-word problem was being run on a machine with 1000-word random-access memory capacity.

11. The Results. The results of the computations done on Maniac I are summarized in Figs. 3, 4, and 5, where the lattice points occur at integer values of

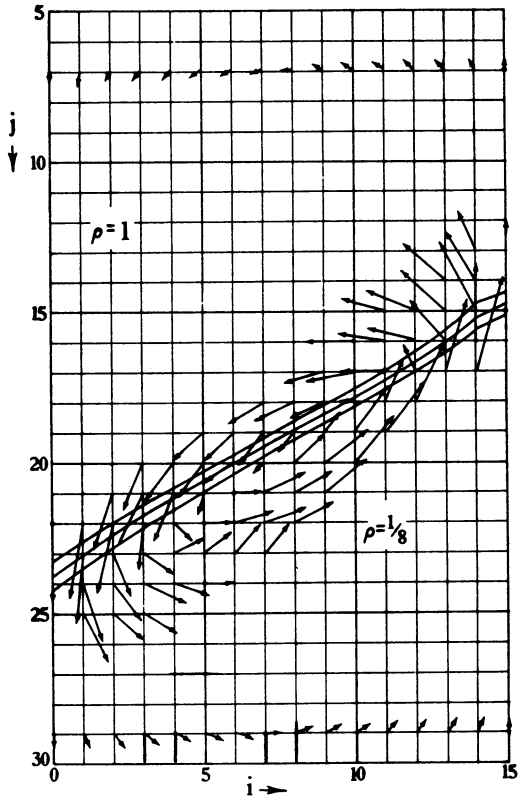

FIG. 4.—Velocity field at $h = 10$.

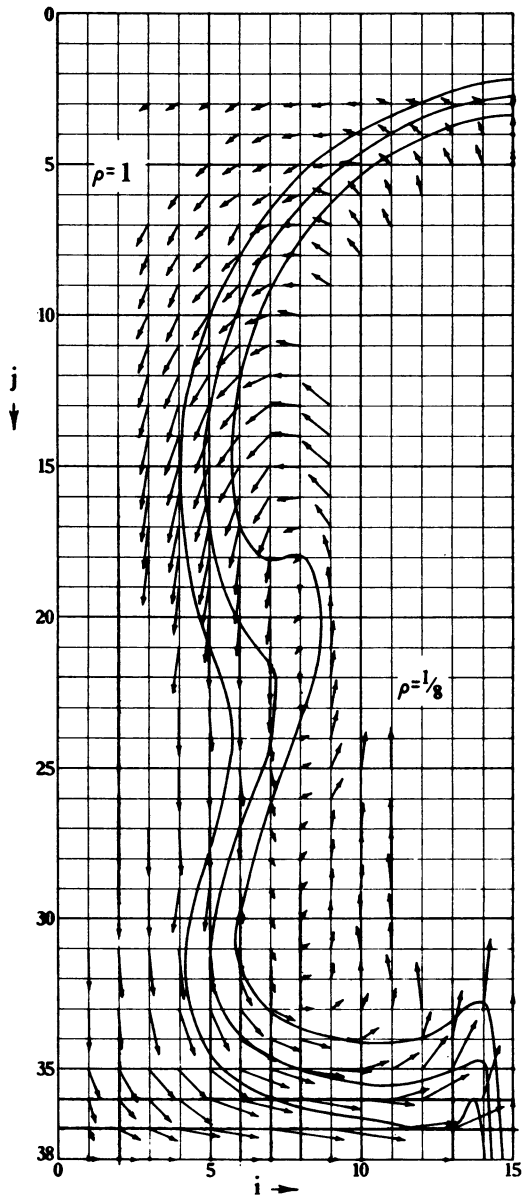

FIG. 5.—Velocity field at $h = 60$.

the abscissae (i) and ordinates (j). In the first of these the locus of $\rho = \frac{1}{2}(\rho_{\text{upper}} + \rho_{\text{lower}}) = 0.5625$ is shown for various values of h, where $h\Delta t$ is the time elapsed since the start of the calculations. Curves are shown for $h = 0$, 10, 20, 30, 40, 50, 55, and 60.

Figure 4 shows the velocity field at $h = 10$, as well as the loci $\rho = 0.5625$, and $\rho = 0.3875$, and $\rho = 0.7375$. The latter two curves are the loci at which 30 and 70 per cent, respectively, of the initial difference in density are achieved. The band

covered by these two curves gives a measure of the pseudo-diffusion phenomenon. It is evident from Fig. 4 that the zone of pseudo-diffusion is less than one mesh length in thickness.

Figure 5 is similar to Fig. 4 but in this case $h = 60$. Now the pseudo-diffusion phenomenon has increased but on the whole is still confined to a region of the order of two mesh lengths.

If $\Delta x = \Delta y$ is taken to be 1 centimeter, then the total time covered by the calculations (60 Δt) would be 0.339 second. The maximum speed attained by the fluid is 0.0184 centimeter/second.

In Fig. 5 the density distribution in the lower right-hand corner seems to have a somewhat anomalous behavior. This may be due to the fact that equation (6.8) was only applied to the boundaries $j = 0$ and $j = J$ and not to the boundaries $i = 0$ and $i = I$.

12. Concluding Remarks. The most time-consuming part of the calculations performed was that devoted to the computation of $\chi^h_{i,j}$. The subsequent computation of the velocities $u^h_{i,j} = -(\psi_y)^h_{i,j}$ and $v^h_{i,j} = (\psi_x)^h_{i,j}$, and the density $\rho^h_{i,j}$ took relatively little time. However, the behavior of the density was most sensitive to the type of integration formula used. It is not yet clear that the formulas actually used were the best ones from the point of view of minimizing the zone of pseudo-diffusion.

It is expected that the use of equation (6.8) for interior points as well as boundary points or other devices will keep the region of pseudo-diffusion small enough so that calculations on moving incompressible fluids with moving interfaces can be made in Eulerian coordinates. If this conjecture would prove to be correct it would be possible to use Eulerian coordinates and avoid the main difficulty of working in Lagrangian ones; namely the necessity of introducing a new Lagrangian mesh periodically because neighboring particles do not remain neighboring particles.

APPENDIX I*

The differential equations are:

Interior:

$$u_t + uu_x + vu_y = -\frac{1}{\rho} p_x, \tag{1}$$

$$v_t + uv_x + vv_y = -\frac{1}{\rho} p_y + g, \tag{2}$$

$$\rho_t + u\rho_x + v\rho_y = 0, \tag{3}$$

$$u_x + v_y = 0. \tag{4}$$

Boundary:

$$\cos(x, n)\, u + \cos(y, n)\, v = 0. \tag{5}$$

* Although the material in Appendix I and Appendix II was left by von Neumann in the form of handwritten notes and not in a form intended for wide distribution, the value of its content is thought to justify its inclusion in this paper.

From (4):

$$u = -\psi_y, \qquad v = \psi_x. \tag{6}$$

(6) replaces (4).

Now (5) becomes:

$$-\cos(x, n)\,\psi_y + \cos(y, n)\,\psi_x = 0,$$

i.e.,

$$\psi_x : \psi_y = \cos(x, n) : \cos(y, n),$$

i.e.,

$$\psi_x, \psi_y \text{ is purely normal,}$$

i.e.,

$$\psi_{\tan} = 0 \quad \text{(the tangential derivative of } \psi \text{ vanishes).}$$

This means that

$$\psi = C \ (= \text{constant})$$

along the boundary. Now replacing ψ by $\psi - C$ does not interfere with (6) (ψ's defining relation). Hence $C = 0$ may be assumed, i.e.:

$$\psi = 0 \tag{7}$$

(on the boundary). (7) replaces (5).

Now only (1), (2), (3) are left. These become:

$$-\psi_{yt} + \psi_y \psi_{xy} - \psi_x \psi_{yy} = -\frac{1}{\rho} p_x,$$

$$\psi_{xt} - \psi_y \psi_{xx} + \psi_x \psi_{xy} = -\frac{1}{\rho} p_y + g,$$

$$\rho_t - \psi_y \rho_x + \psi_x \rho_y = 0,$$

i.e.,

$$\rho[\psi_{ty} + I(\psi, \psi_y)] = p_x, \tag{8}$$

$$\rho[\psi_{tx} + I(\psi, \psi_x) - g] = -p_y, \tag{9}$$

$$\rho_t = -I(\psi, \rho). \tag{10}$$

(8), (9), (10) replace (1), (2), (3). p can be eliminated between (8), (9), by forming $(8)_y + (9)_x$. This gives

$$\{\rho[\psi_{ty} + I(\psi, \psi_y)]\}_y + \{\rho[\psi_{tx} + I(\psi, \psi_x) - g]\}_x = 0,$$

i.e.,

$$(\rho\psi_{tx})_x + (\rho\psi_{ty})_y = -\{-g\rho_x + [\rho I(\psi, \psi_x)]_x + [\rho I(\psi, \psi_y)]_y\}. \tag{11}$$

(11) replaces (8), (9) (with p eliminated).

Thus the entire system now consists of (10), (11) (interior) and (7) (boundary), while (6) is merely a definition.

Rewriting (10), (11) and (6):

Interior:

$$L\psi_t = -\{[\rho(-g + I[\psi, \psi_x])]_x + [\rho I(\psi, \psi_y)]_y\}, \tag{12}$$

where

$$L\chi \equiv (\rho\chi_x)_x + (\rho\chi_y)_y\,, \qquad \rho_t = -I(\psi, \rho). \tag{13}$$

Boundary:

$$\psi = 0. \tag{14}$$

This suggests the following procedure: Put

$$\chi = -\psi_t\,. \tag{15}$$

Then:

Let ψ, ρ be known. Then ψ_t, ρ_t are obtained by this procedure: Put

$$\omega = \{\rho[-g + I(\psi, \psi_x)]\}_x + [\rho I(\psi, \psi_y)]_y\,. \tag{16}$$

Determine χ from this elliptic boundary value problem:

$$L\chi = \omega, \tag{17}$$

where

$$L\chi \equiv (\rho\chi_x)_x + (\rho\chi_y)_y\,,$$

and on the boundary

$$\chi = 0. \tag{18}$$

Then:

$$\psi_t = -\chi, \tag{19}$$

$$\rho_t = -I(\psi, \rho). \tag{20}$$

The expression (16) for ω may be rewritten:

$$\omega = \rho_x[-g + I(\psi, \psi_x)] + \rho[I(\psi, \psi_x)]_x + \rho_y I(\psi, \psi_y) + \rho[I(\psi, \psi_y)]_y\,.$$

Now

$$[I(\psi, \psi_x)]_x = I(\psi_x\,, \psi_x) + I(\psi, \psi_{xx}) = I(\psi, \psi_{xx}),$$

$$[I(\psi, \psi_y)]_y = I(\psi_y\,, \psi_y) + I(\psi, \psi_{yy}) = I(\psi, \psi_{yy}).$$

Hence

$$\omega = \rho_x[-g + I(\psi, \psi_x)] + \rho_y I(\psi, \psi_y) + \rho I(\psi, \psi_{xx} + \psi_{yy}). \tag{21}$$

Equation (21) replaces (16); it is more convenient for numerical calculation.

This is a more detailed expression for ω:

$$\lambda = \psi_{xx} + \psi_{yy}\,, \tag{22.1}$$

$${}^1I = \psi_x\psi_{xy} - \psi_y\psi_{xx}\,, \tag{22.2}$$

$${}^2I = \psi_x\psi_{yy} - \psi_y\psi_{xy}\,, \tag{22.3}$$

$${}^3I = \psi_x\lambda_y - \psi_y\lambda_x\,, \tag{22.4}$$

$$\omega = \rho_x(g + {}^1I) + \rho_y{}^2I + \rho{}^3I. \tag{22.5}$$

In addition to this the right hand side of (20) is the negative of

$$^4I = I(\psi, \rho), \tag{23}$$

where in detail

$$^4I = \psi_x\rho_y - \psi_y\rho_x . \tag{24}$$

Thus the relevant equations are these: (17) with (22.1)–(22.5) and (on the boundary) (18), and then (19) and (20) with (24).

Now introduce a finite lattice for x, y, t:

$$x = x_i \equiv i\,\Delta x, \qquad i = 0, 1, \cdots, I-1, I, \tag{25.1}$$

$$y = y_j \equiv j\,\Delta y, \qquad j = 0, 1, \cdots, J-1, J, \tag{25.2}$$

$$t = t^h \equiv h\,\Delta t, \qquad h = 0, 1, 2, \cdots . \tag{25.3}$$

Occasionally indices $i \pm \frac{1}{2}$, $j \pm \frac{1}{2}$, $h \pm \frac{1}{2}$ are also used. For any quantity

$$\alpha = \alpha(x, y, t). \tag{26.1}$$

The following convention is used:

$$\alpha^h_{ij} = \alpha(x_i, y_j, t^h). \tag{26.2}$$

(17) and (22.1)–(22.5) are needed for $i = 1, \cdots, I-1$; $j = 1, \cdots, J-1$ only. (18) is needed for the other i, j only: $i = 0, I$; $j = 0, 1, \cdots, J-1, J$ or $i = 0, 1, \cdots, I-1, I$; $j = 0, J$. (19), (20) are needed for all i, j: $i = 0, 1, \cdots, I-1, I$; $j = 0, 1, \cdots, J-1, J$.

The entire system of equations can now be rewritten as follows:

ψ^h_{ij} is defined for $i = 1, \cdots, I-1$; $j = 1, \cdots, J-1$.

ρ^h_{ij} is defined for $i = 0, 1, \cdots, I-1, I$; $j = 0, 1, \cdots, J-1, J$.

In (A.1)–(A.14):*

$$i = 1, \cdots, I-1; \qquad j = 1, \cdots, J-1.$$

$$(\bar{\psi}_x)^h_{ij} = \psi^h_{i+1j} - \psi^h_{i-1j} \tag{A.1}$$

for $i \neq 1, I-1$;
for $i = 1$ omit term ——;
for $i = I-1$ omit term ~~~.

$$(\bar{\psi}_y)^h_{ij} = \psi^h_{ij+1} - \psi^h_{ij-1} \tag{A.2}$$

for $j \neq 1, J-1$;
for $j = 1$ omit term ——;
for $j = J-1$ omit term ~~~.

$$(\bar{\psi}_{xy})^h_{ij} = \psi^h_{i+1j+1} - \psi^h_{i-1j+1} - \psi^h_{i+1j-1} + \psi^h_{i-1j-1} \tag{A.3}$$

* The bar notation includes the required constant times Δx or Δy combinations.

for $i \neq 1, I-1; j \neq 1, J-1$;
for $i = 1$ omit terms____;
for $i = I-1$ omit terms~~~~;
for $j = 1$ omit terms_ _ _;
for $j = J-1$ omit terms....

In (A.4)–(A.6):

$$(i', j') = (i, j), \underbrace{(i+1, j)}, \underline{(i-1, j)}, \overset{\cdots\cdots}{(i, j+1)}, \underline{\underline{(i, j-1)}}$$

for $i \neq 1, I-1; j \neq 1, J-1$;
for $i = 1$ omit term____;
for $i = I-1$ omit term~~~~;
for $j = 1$ omit term_ _ _;
for $j = J-1$ omit term....
(Hence $i' = 1, \cdots, I-1; j' = 1, \cdots, J-1$.)

$$(\bar{\psi}_{xx})^h_{i'j'} = \underbrace{\psi^h_{i'+1j'}} - 2\psi^h_{i'j'} + \underline{\psi^h_{i'-1j'}} \tag{A.4}$$

for $i' \neq 1, I-1$;
for $i' = 1$ omit term____;
for $i' = I-1$ omit term~~~~.

$$(\bar{\psi}_{yy})^h_{i'j'} = \underbrace{\psi^h_{i'j'+1}} - 2\psi^h_{i'j'} + \underline{\psi^h_{i'j'-1}} \tag{A.5}$$

for $j \neq 1, J-1$;
for $j' = 1$ omit term____;
for $j' = J-1$ omit term~~~~.

$$\bar{\lambda}^h_{i'j'} = (\bar{\psi}_{xx})^h_{i'j'} + a(\bar{\psi}_{yy})^h_{i'j'}, \tag{A.6}$$

where

$$a = \left(\frac{\Delta x}{\Delta y}\right)^2.$$

$$(\bar{\lambda}_x)^h_{ij} = \bar{\lambda}^h_{i+1j} - \bar{\lambda}^h_{i-1j} \tag{A.7}$$

for $i \neq 1, I-1$;
for $i = 1$ replace the index $i-1$ by 1 and double the entire expression;
for $i = I-1$ replace the index $i+1$ by $I-1$ and double the entire expression.

$$(\bar{\lambda}_y)^h_{ij} = \bar{\lambda}^h_{ij+1} - \bar{\lambda}_{ij-1} \tag{A.8}$$

for $j \neq 1, J-1$;
for $j = 1$ replace the index $j-1$ by 1 and double the entire expression;
for $j = J-1$ replace the index $j+1$ by $J-1$ and double the entire expression.

$$(\bar{p}_x)^h_{ij} = \rho^h_{i+1j} - \rho^h_{i-1j}. \tag{A.9}$$

$$(\bar{p}_y)^h_{ij} = \rho^h_{ij+1} - \rho^h_{ij-1}. \tag{A.10}$$

$${}^1\bar{I}^h_{ij} = (\bar{\psi}_x)^h_{ij}(\bar{\psi}_{xy})^h_{ij} - (\bar{\psi}_y)^h_{ij}(\bar{\psi}_{xx})^h_{ij}. \tag{A.11}$$

$${}^2\bar{I}^h_{ij} = (\bar{\psi}_x)^h_{ij}(\bar{\psi}_{yy})^h_{ij} - (\bar{\psi}_y)^h_{ij}(\bar{\psi}_{xy})^h_{ij}. \tag{A.12}$$

(A.13) $$ {}^{3}\bar{I}^{h}_{ij} = (\bar{\psi}_x)^{h}_{ij}(\bar{\lambda}_y)^{h}_{ij} - (\bar{\psi}_y)^{h}_{ij}(\bar{\lambda}_x)^{h}_{ij} . $$

(A.14) $$ \bar{\omega}^{h}_{ij} = (\bar{p}_x)^{h}_{ij}\,(-\bar{g} + {}^{1}\bar{I}^{h}_{ij}) + a(\bar{p}_y)^{h}_{ij}\,{}^{2}\bar{I}^{h}_{ij} + \bar{p}^{h}_{ij}\,{}^{3}\bar{I}^{h}_{ij} , $$

where $\bar{g} = (\Delta x)^2\,\Delta y g$ [for a, cf. (A.6)].

In (B.1)–(B.6):

$$ i = 1, \cdots, I-1; \qquad j = 1, \cdots, J-1. $$

This definition of $\bar{\chi}^{h}_{ij}$ [i.e., (B.6)] is, however, implicit.

(B.1) $$ {}^{1}\bar{\sigma}^{h}_{ij} = \bar{p}^{h}_{ij} + \bar{p}^{h}_{i+1j} . $$

(B.2) $$ {}^{2}\bar{\sigma}^{h}_{ij} = \bar{p}^{h}_{ij} + \bar{p}^{h}_{i-1j} . $$

(B.3) $$ {}^{3}\bar{\sigma}^{h}_{ij} = a(\bar{p}^{h}_{ij} + \bar{p}^{h}_{ij+1}). $$

(B.4) $$ {}^{4}\bar{\sigma}^{h}_{ij} = a(\bar{p}^{h}_{ij} + \bar{p}^{h}_{ij-1}). $$

(B.5) $$ {}^{5}\bar{\sigma}^{h}_{ij} = ({}^{1}\bar{\sigma}^{h}_{ij} + {}^{2}\bar{\sigma}^{h}_{ij}) + a({}^{3}\bar{\sigma}^{h}_{ij} + {}^{4}\bar{\sigma}^{h}_{ij}) \qquad \text{[for } a\text{, cf. (A.6)]}. $$

(B.6) $$ {}^{1}\bar{\sigma}^{h}_{ij}\,\bar{\chi}^{h}_{i+1j} + {}^{2}\bar{\sigma}^{h}_{ij}\,\bar{\chi}^{h}_{i-1j} + {}^{3}\bar{\sigma}^{h}_{ij}\,\bar{\chi}^{h}_{ij+1} + {}^{4}\bar{\sigma}^{h}_{ij}\,\bar{\chi}^{h}_{ij-1} - {}^{5}\bar{\sigma}^{h}_{ij}\,\bar{\chi}^{h}_{ij} = \bar{\omega}^{h}_{ij} $$

for $i \neq 1, I-1; j \neq 1, J-1$;

for $i = 1$ omit the term ——;

for $i = I-1$ omit the term ~~~;

for $j = 1$ omit the term – – –;

for $j = J-1$ omit the term ....

The term with the double underscore should be $C\,\bar{\chi}^{h}_{ij}$, the corrected $\bar{\chi}^{h}_{ij}$.

In (C):

$$ i = 1, \cdots, I-1; \qquad j = 1, \cdots, J-1. $$

(C) $$ \psi^{h+1}_{ij} = \psi^{h-1}_{ij} - 4b\,\bar{\chi}^{h}_{ij} $$

where

$$ b = \frac{\Delta t}{\Delta x\,\Delta y} . $$

In (D.1–(D.3):

$$ i = 0, 1, \cdots, I-1, I; \qquad j = 0, 1, \cdots, J-1, J. $$

As in (C):

(D.1) $$ b = \frac{\Delta t}{\Delta x\,\Delta y} . $$

$$ \alpha = \tfrac{1}{2}b(\psi^{h}_{ij+1} + \psi^{h+1}_{ij+1} - \psi^{h}_{ij-1} - \psi^{h+1}_{ij-1}), $$

for $i \neq 0, I$ and $j \neq 0, J$;

for $i = 0, I$ omit the entire expression;

for $i \neq 0, I$ and $J = 0$ omit the terms and factor____;

for $i \neq 0, I$ and $j = J$ omit the terms and factor~~~~.

$$\beta = \tfrac{1}{2}b(-\psi^{h}_{i+1j} - \psi^{h+1}_{i+1j} + \psi^{h}_{i-1j} + \psi^{h+1}_{i-1j}) \tag{D.2}$$

for $i \neq 0, I$ and $j \neq 0, J$;

for $j = 0, J$ omit the entire expression;

for $i = 0$ omit the terms and factor____;

for $j \neq 0$, J and $i = I$ omit the terms and factor~~~~.

$$\begin{aligned}\rho^{h+1}_{ij} = \rho^{h}_{ij}(1 - \alpha^2 - \beta^2) + \tfrac{1}{2}\rho^{h}_{i+1j}(\alpha + \alpha^2) + \tfrac{1}{2}\rho^{h}_{i-1j}(-\alpha + \alpha^2) \\ + \tfrac{1}{2}\rho^{h}_{ij+1}(\beta + \beta^2) + \tfrac{1}{2}\rho^{h}_{ij-1}(-\beta + \beta^2) \\ + \tfrac{1}{4}(\rho^{h}_{i+1j+1} - \rho^{h}_{i+1j-1} - \rho^{h}_{i-1j+1} + \rho^{h}_{i-1j-1}) - \alpha\beta\end{aligned} \tag{D.3}$$

for $i \neq 0, I$ and $j \neq 0, J$;

for $i = 0, I$ omit the terms____;

for $j = 0, J$ omit the terms~~~~.

[Note: The points with $i = 0, I$ and $j = 0, J$ (together!) may be bypassed.]

Alternatively, in place of (D.3):

$$\epsilon = \operatorname{Sgn}\alpha,\ \eta = \operatorname{Sgn}\beta$$

$$\begin{aligned}\rho^{h+1}_{ij} = \rho^{h}_{ij} + (\rho^{h}_{i+\epsilon j} - \rho^{h}_{ij})\,|\alpha| + (\rho^{h}_{ij+\eta}\,\rho^{h}_{ij})\,|\beta| \\ + (\rho^{h}_{i+\epsilon j+\eta} - \rho^{h}_{i+\epsilon j} - \rho^{h}_{ij+\eta} + \rho^{h}_{ij})\,|\alpha|\,|\beta|\end{aligned} \tag{D'.3}$$

for $i \neq 0, I$ and $J \neq 0, J$;

for $i = 0, J$ omit the terms____;

for $j = 0, J$ omit the terms~~~~.

[Note: The points with $i = 0, I$ and $j = 0, I$ (together!) may be bypassed.]

APPENDIX II

1. The purpose of this paper is to find a rapidly converging iterative method for the solution of linear equation systems, and quite particularly of those which arise from the difference equation treatment of partial differential equations of the elliptic type [2nd order, s $(= 2, 3, \cdots)$ variables]. Sections **2–6** are introductory. The method will be described and discussed in Sections **7–13**. The application to the (elliptic) differential equation case will be made in Sections **14–15**. Some comparisons will be made. The results are summarized in Sections **11, 13, 15.**

2. Consider a system of n linear equations in n variables, written vectorially:

$$A\xi = \alpha. \tag{1}$$

Here α is a known nth order vector, A a known nth order matrix, ξ the unknown nth order vector. In order that the problem be meaningful, A must be non-singular. This will be assumed.

An iterative method is based on a correction step, which replaces a ξ, that may not solve (1), by a ξ^1, that, in some suitable sense, should more nearly solve (1). This correction step should be a linear operation F applied to the two nth order vectors ξ, α, i.e., to the $2n$th order vector $\{\xi, \alpha\}$. It then produces the nth order vector ξ^1:

$$\xi^1 = F\{\xi, \alpha\}. \tag{2}$$

It is convenient, to put in place of the nth order vector ξ^1 again a $2n$th order vector, namely $\{\xi^1, \alpha\}$. In this case let us write E in place of F:

$$\{\xi^1, \alpha\} = E\{\xi, \alpha\}. \tag{3}$$

Thus E is a $2n$th order matrix.

Note that the linearity of F means that it can be written as follows:

$$F\{\xi, \alpha\} = G\xi + H\alpha, \tag{4}$$

where G, H are nth order matrices.

(2), (3), (4) mean that the $2n$th order matrix E can be written as a 2nd order hypermatrix of nth order matrices, as follows:

$$E = \left(\begin{array}{c|c} G & H \\ \hline O & I \end{array}\right). \tag{5}$$

Here, O, I are the (nth order) zero and unit matrix, as usual.

3. A minimum requirement to be imposed on a correction step in the sense of **2** is this: If ξ^* is the solution of (1), then the correction should leave $\xi = \xi^*$ unchanged, i.e., produce $\xi^1 = \xi (= \xi^*)$. This is the "weak" condition. A reasonable further requirement is that if ξ is not a solution of (1) ($\xi^1 \neq \xi^*$), then the correction should change ξ, i.e., produce a $\xi^1 \neq \xi$. This is the "strong" condition. That is, the weak (strong) condition requires that $\xi = \xi^*$ be sufficient (necessary and sufficient) for $\xi^1 = \xi$.

By (2), (4) $\xi^1 = \xi$ means

$$(I - G)\xi = H\alpha. \tag{6}$$

The weak condition requires, that (1) imply (6), i.e., that always

$$(I - G)\xi = HA\xi,$$

i.e.,

$$\begin{aligned} I - G &= HA, \\ G &= I - HA. \end{aligned} \tag{7}$$

The strong condition requires, in addition to this, that (6) imply (1), i.e., in view of (7), that

$$HA\xi = H\alpha$$

imply

$$A\xi = \alpha.$$

This means obviously that

(8a) H is non-singular.

Equivalently:

(8b) 0 is not a characteristic root of H.

Since A is non-singular, non-singularity of H is equivalent to that of HA, i.e., [by (7)] of $I - G$; i.e., equivalent to this: 0 is not a characteristic root of $I - G$, or equivalently:

(8c) 1 is not a characteristic root of G.

To begin with, we will only stipulate the weak condition, i.e., (7).

4. The ordinary iterative procedure consists of repeating the basic step (2) successively, and to expect that the sequence so generated will converge to the solution ξ^* of (1), irrespective of the starting point ξ.

Disregarding for the moment the question of convergence, the sequence in question, $\xi^0, \xi^1, \xi^2, \cdots$, is defined by

$$\left.\begin{aligned} \xi^0 &= \xi, \\ \xi^{k+1} &= F\{\xi^k, \alpha\} \qquad (k = 0, 1, 2, \cdots). \end{aligned}\right\} \tag{9}$$

In view of (2), (3), the second equation of (9) can be written

$$\{\xi^{k+1}, \alpha\} = E\{\xi^k, \alpha\} \qquad (k = 0, 1, 2, \cdots),$$

and hence (9) is equivalent to

$$\{\xi^{k+1}, \alpha\} = E^{k+1}\{\xi, \alpha\} \qquad (k = 0, 1, 2, \cdots). \tag{10}$$

We know that for $\xi = \xi^*$ [ξ^* the solution of (1), cf. 3] all $\xi^k = \xi^*$, hence (10) gives

$$\{\xi^*, \alpha\} = E^k\{\xi^*, \alpha\}.$$

Hence (10) is equivalent to what is obtained by subtracting this equation from it, i.e., to

$$\{\xi^k - \xi^*, 0\} = E^k\{\xi - \xi^*, 0\},$$

i.e., in view of (5) to

$$\xi^k - \xi^* = G^k(\xi - \xi^*) \qquad (k = 0, 1, 2, \cdots). \tag{11}$$

Note that (10) is an effective calculational procedure, while (11) is not, since it contains the unknown ξ^*; however, some proofs and evaluations can be more advantageously based on (11).

It is well known that frequently the convergence properties of a sequence can be significantly improved by replacing each element of the sequence by a suitable mean of itself and the preceding elements of the sequence. In this sense, one might replace the sequence $\xi^0, \xi^1, \xi^2, \cdots$ by a sequence $\mathfrak{n}^0, \mathfrak{n}^1, \mathfrak{n}^2, \cdots$, where

$$\mathfrak{n}^k = \sum_{l=0}^{k} a_{kl}\xi^l \qquad (k = 0, 1, 2, \cdots), \tag{12}$$

with a suitable set of coefficients a_{kl}. The characterization of the $\mathfrak{n}^k$ as means (of the ξ^k) makes it natural to require

$$\sum_{l=0}^{k} a_{kl} = 1 \qquad (k = 0, 1, 2, \cdots). \tag{13}$$

This condition can also be obtained from the natural requirement that for $\xi = \xi^*$, when all $\xi^k = \xi^*$, there shall also be all $\mathfrak{n}^k = \xi^*$. At any rate, we stipulate (13). The characterization of the $\mathfrak{n}^k$ as means might also suggest the requirement that all $a_{kl} \geqq 0$, but we will not impose it; indeed, the choice that we will later make, and that seems to be particularly favorable, will violate this condition [cf. Sections **7–9**, in particular (53)].

Instead of working with the coefficients a_{kl} themselves, we can also work with the corresponding polynomials

$$P_k(Z) = \sum_{l=0}^{k} a_{kl} Z^l \qquad (k = 0, 1, 2, \cdots). \tag{14}$$

Then (13) becomes

$$P_k(1) = 1. \tag{15}$$

Thus $P_k(Z)$ is a kth order polynomial fulfilling (15), and (so far) subject to no other restrictions.

Now (12) becomes, using (10) [and (15)],

$$\{\mathfrak{n}^k, \alpha\} = P_k(E)\ \{\xi, \alpha\}, \tag{16}$$

or equivalently, using (11),

$$\mathfrak{n}^k - \xi^* = P_k(G)\ (\xi - \xi^*). \tag{17}$$

The relationship between (16), (17) is similar to that between (10), (11), as discussed immediately after (11).

The broader convergence problem for the iterative-and-mean procedure is this: Choose the a_{kl}, i.e., the sequence $[P_0(Z), P_1(Z), P_2(Z), \cdots]$, [$P_k(Z)$ a kth order polynomial fulfilling (15), cf. above], so that

$$\lim_{k\to\infty} \mathfrak{n}^k = \xi^* \tag{18}$$

for all choices of the starting point ξ. (18) can be also be written like this:

$$\lim_{k\to\infty} D(\mathfrak{n}^k - \xi^*) = 0, \tag{19}$$

where $D(\xi)$ is any norm in the space of all nth order vectors ξ (i.e., in n-dimensional Euclidean space), which is equivalent to the ordinary (Euclidean) topology of that space. We will make a specific choice of $D(\xi)$ soon: Section **7** (30).

The ordinary iteration convergence problem (without means, cf. the beginning of **4**) corresponds to the choice $P_k(Z) \equiv Z^k$ $(k = 0, 1, 2, \cdots)$ for the sequence $[P_0(Z), P_1(Z), P_2(Z), \cdots]$.

5. We will now consider the broad convergence problem (iterative-and-mean procedure, cf. **4** above) in more specific detail.

(19) works with

$$d_k = D(\mathfrak{n}^k - \xi^*) \qquad (k = 0, 1, 2, \cdots), \tag{20}$$

and its requirement is

$$\lim_{k\to\infty} d_k = 0 \qquad \text{(for all } \xi\text{)}. \tag{21}$$

Hence the relevant quantity is d_k, and we must concentrate on estimating its size (for all ξ).

Combining (20) and (17) gives

$$d_k = D[P_k(G)\omega], \tag{22}$$

where

$$\omega = \xi - \xi^*.$$

(22), (21) show that the convergence problem is actually one of the convergence of the matrices $P_k(G)$ $(k \to \infty)$ to zero.

Thus the problem presents itself in this form: Given a matrix G, what conditions must a sequence of polynomials $[P_0(Z), P_1(Z), P_2(Z), \cdots]$ fulfill so as to have

$$\lim_{k\to\infty} P_k(G) = 0. \tag{23}$$

The answer is well known: Let λ_i $(i = 1, \cdots, \mu$, of course $\mu \leqq n)$ be the characteristic roots of G, and let e_i $(= 1, 2, \cdots)$ be the order of the elementary divisor of G that corresponds to λ_i. Denote the ρth derivative of $P_k(Z)$ by $P_k^{(\rho)}(Z)$. Then the necessary and sufficient condition for the validity of (23) is this:

$$\lim_{k\to\infty} P_k^{(\rho)}(\lambda_i) = 0 \tag{24}$$

for all those combinations $i(= 1, \cdots, \mu)$, $\rho(= 0,1, \cdots)$ for which $e_i > \rho$. When all $e_i = 1$, then (24) becomes simply

$$\lim_{k\to\infty} P_k(\lambda_i) = 0 \qquad (i = 1, \cdots, \mu). \tag{25}$$

For a Hermitian G, in particular, this is always the case.

We saw in **4**, that the $P_k(Z)$ are subject to the condition (15): $P_k(1) = 1$. Hence (24), i.e., (23), is unfulfillable if some $\lambda_i = 1$, i.e., if 1 is a characteristic root of G. In other words: The condition (8c), i.e., the strong condition of **3**, is reimposed for this reason. [This could have been seen directly too: If that condition fails, then for some $\xi \neq \xi^*$ there is $\xi^1 = \xi$, hence all $\xi^k = \xi$, hence all $\mathfrak{n}^k = \xi$, and so $\lim_{k\to\infty} \mathfrak{n}^k = \xi \neq \xi^*$, contradicting (18).]

If, on the other hand, (8c), i.e., the strong condition in **3**, holds, i.e., if all $\lambda_i \neq 1$, then it is not difficult to see that (24) and (15) are compatible. Indeed, even a fixed $P(Z)$ [for all $P_k(Z)$ with $k \geqq$ the precise order of $P(Z)$, which is $\sum_i e^i$, cf. below] will do:

$$P(Z) \equiv c\prod_i (Z - \lambda_i)^{e_i},$$

with c determined from (15), meets all requirements. This, however, is of small practical importance, since the λ_i may not be known, and the above expression for $P(Z)$ may in any case be too complicated for actual evaluation. If it is only known that the λ_i lie in the interior of a certain (bounded and closed) domain Λ, then a sequence $[P_0(Z), P_1(Z), P_2(Z), \cdots]$ of the desired kind can still be specified, if (and only if) Λ does not separate 1 from ∞. We will, however, not go here into this matter any further.

6. The ordinary iterative procedure corresponds, as we observed at the end of **4**, to the choice $P_k(Z) \equiv Z^k$. Hence $P_k^{(\rho)}(Z) \equiv k(k-1) \cdots (k-\rho+1)Z^{k-\rho}$. Therefore the convergence criterion (24) requires precisely, that all $|\lambda_i| < 1$. We state this explicitly:

(26) The ordinary iterative procedure converges (cf. the beginning of 4) if and only if $|\lambda| < 1$ for all characteristic values λ of G.

As we saw in the last part of **5**, this condition is by no means necessary for the convergence of some suitable iterative-and-mean procedure. We will nevertheless limit ourselves to this case:

(27) $$|\lambda| < 1 \quad \text{for all characteristic roots } \lambda \text{ of } G.$$

In addition, we will assume that G is Hermitian, because this covers certain important applications, and permits the employment of some rather effective methods. We restate this:

(28) $$G \text{ is Hermitian.}$$

(28) implies that all characteristic roots (or characteristic values) of G are real. Hence (27) becomes this:

(29) $$-1 < \lambda < 1 \quad \text{for all characteristic values } \lambda \text{ of } G.$$

Under these conditions the ordinary iterative procedure, i.e., the choice (20), $P_k(Z) \equiv Z^k$ (cf. above), is adequate, i.e., it guarantees convergence. We wish, however, to determine that iterative-and-mean procedure, i.e., that sequence $[P_0(Z), P_1(Z), P_2(Z), \cdots]$ for which this convergence is (uniformly) fastest. We will, therefore reconsider the convergence problem under this aspect, subject to the restrictions (28), (29).

7. We want to choose the sequence $[P_0(Z), P_1(Z), P_2(Z), \cdots]$ so as to obtain the uniformly fastest possible convergence. This convergence is to be taken in the sense of (21), i.e., we want to make for each k the d_k of (20) as small as possible. {By (22) this means that we want to make $D[P_k(G)\omega]$ as small as possible.} This should be true, in some suitable sense, uniformly—i.e., uniformly in the variables of (22). These variables are [since k is given and $P_k(Z)$ is being looked for] G and $\omega (= \xi - \xi^*)$. Let us therefore examine the meaning of uniformity with respect to G and ω.

First, since we are now dealing with a situation in which a Hermitian matrix, G, occupies a central role, it is reasonable to prescribe that the norm $D(\xi)$ be the Euclidean norm

(30) $$D(\xi) = \sqrt{\sum_{i=1}^{n} |\xi_i|^2}$$

($\xi_1, \cdots, \xi_n$ are the [complex, numerical] components of the [nth order] vector ξ). (Cf. the remark following (19) in 4)

It is also useful to introduce, for a general (nth order) matrix K, the concepts of the "upper bound" $| K |_u$ and of the "lower bound" $| K |_l$:

$$| K |_u = \operatorname*{Max}_{\xi \neq 0} \frac{D(K\xi)}{D(\xi)} = \text{Min [C such that } D(K\xi) \leqq CD(\xi)], \tag{31a}$$

$$| K |_l = \operatorname*{Min}_{\xi \neq 0} \frac{D(K\xi)}{D(\xi)} = \text{Max [C such that } D(K\xi) \geqq CD(\xi)]. \tag{31b}$$

Second, let us consider the variability of G. G is subject to the requirements (28), (29), i.e., it must be Hermitian and fulfill (29). It is clearly logical to make the requirement (29) with uniformity, i.e., to require for a suitable $\epsilon > 0$ that

$$-(1 - \epsilon) \leqq \lambda \leqq 1 - \epsilon \tag{32a}$$

for all characteristic values λ of G, or equivalently

$$| G |_u \leqq 1 - \epsilon. \tag{32b}$$

Third, let us consider the variability of ω. We want to make $D[P_k(G)\omega]$ as small as possible (cf. above). $\omega = 0$ is uninteresting, and it is plausible that we should want to make the ratio $D[P_k(G)\omega]/D(\omega)$ uniformly small for all ω, i.e., to make

$$\operatorname*{Max}_{\omega \neq 0} \frac{D[P_k(G)\omega]}{D(\omega)}$$

small. This means, by (31a), that we want to make $| P_k(G) |_u$ small.

Combining the second and the third remarks, we see that we want to make $| P_k(G) |_u$ uniformly small for all Hermitian G that fulfill (32a) [i.e., (32b)]. That is, we want to minimize

$$\bar{d}_k = \operatorname*{Max}_{\substack{G \text{ Hermitian} \\ |G|_u \leqq 1-\epsilon}} (| P_k(G) |_u) = \text{Min } \{C \text{ such that for all } \omega \text{ and all Hermitian } G \text{ with } | G |_u \leqq 1 - \epsilon, D[P_k(G)\omega] \leqq CD(\omega)\}. \tag{33}$$

Now $| P_k(G) |_u$ is the maximum $P_k(\lambda)$, where λ runs over all characteristic values of G. In view of the equivalence of (32a) and (32b), the precise limitation on these λ is $-(1 - \epsilon) \leqq \lambda \leqq 1 - \epsilon$. Hence the first part of (33) can be rewritten

$$\bar{d}_k = \operatorname*{Max}_{-(1-\epsilon) \leqq \lambda \leqq (1-\epsilon)} (| P_k(\lambda) |). \tag{34}$$

Thus we are looking for that kth order polynomial $P_k(Z)$, fulfilling (15), $P_k(1) = 1$, for which

$$\operatorname*{Max}_{-(1-\epsilon) \leqq Z \leqq 1-\epsilon} (| P_k(Z) |)$$

is minimal. Equivalently: We are looking for that kth order polynomial $Q_k(Z)$, fulfilling $| Q_k(Z) | \leqq 1$ for all Z in $-(1 - \epsilon) \leqq Z \leqq (1 - \epsilon)$, for which $Q_k(1)$ is maximal. Indeed, for this $Q_k(1)$ clearly

$$\operatorname*{Max}_{-(1-\epsilon) \leqq Z \leqq 1-\epsilon} (| Q_k(Z) |) = 1,$$

and

$$P_k(Z) \equiv \frac{Q_k(Z)}{Q_k(1)}, \tag{35}$$

$$\operatorname*{Max}_{-(1-\epsilon)\leqq Z\leqq(1-\epsilon)} (|P_k(Z)|) = \frac{1}{Q_k(1)}. \tag{36}$$

Again equivalently: We are looking for that kth order polynomial $R_k(Z)$, fulfilling $|R(Z)| \leqq 1$ for all Z in $-1 \leqq Z \leqq 1$, for which $R_k[1/(1-\epsilon)]$ is maximal. Clearly

$$Q_k(Z) \equiv R_k\left(\frac{Z}{1-\epsilon}\right). \tag{37}$$

8. The last problem in **7** [the one relative to $R_k(Z)$] is classical. It has been solved by Chebyshev [2]. The $R_k(Z)$ in question is the kth Chebyshev-polynomial, defined by

$$R_k(\cos u) \equiv \cos(ku). \tag{38}$$

It is clear from (38) that $R_k(Z)$ is the kth order polynomial, and that $-1 \leqq Z \leqq 1$ implies $|R_k(Z)| \leqq 1$, as desired. Putting $u = iv$ gives $R_k(Ch\ v) = Ch(kv)$, putting $e^v = x$ gives $R_k[\frac{1}{2}(x + x^{-1})] = \frac{1}{2}(x^k + x^{-k})$, and putting $x = Z + \sqrt{Z^2 - 1}$ gives

$$R_k(Z) \equiv \tfrac{1}{2}[(Z + \sqrt{Z^2-1})^k + (Z + \sqrt{Z^2-1})^{-k}]. \tag{39}$$

Now putting $Z = 1/(1-\epsilon)$ gives

$$R_k\left(\frac{1}{1-\epsilon}\right) = \frac{1}{2}\left\{\left[\frac{1+\sqrt{(2-\epsilon)\epsilon}}{1-\epsilon}\right]^k + \left[\frac{1+\sqrt{(2-\epsilon)\epsilon}}{1-\epsilon}\right]^{-k}\right\}. \tag{40}$$

Combining (37) with (35), (36) gives:

$$P_k(Z) \equiv \frac{R_k\left(\frac{Z}{1-\epsilon}\right)}{R_k\left(\frac{1}{1-\epsilon}\right)}, \tag{41}$$

$$\operatorname*{Max}_{-(1-\epsilon)\leqq Z\leqq 1-\epsilon} (|P_k(Z)|) = \frac{1}{R_k\left(\frac{1}{1-\epsilon}\right)}. \tag{42}$$

It is worthwhile to compare the efficiency of this scheme with that of the ordinary iterative procedure (without means), i.e., with the choice $P_k(Z) \equiv Z^k$ (cf. the end of **4** and the beginning of **6**).

Consider first the present choice for $P_k(Z)$ [i.e., (41)]. The logarithm of the first term in the bracket on the right hand side of (40) is

$$\ln\left(\frac{1+\sqrt{(2-\epsilon)\epsilon}}{1-\epsilon}\right)\cdot k,$$

i.e., for $\epsilon \ll 1$ it is $\sim \sqrt{2\epsilon}\cdot k$. The logarithm of the second term is correspondingly $\sim -\sqrt{2\epsilon}\cdot k$. Assume furthermore $\sqrt{2\epsilon}\cdot k \gg 1$, then the first term is dominant,

i.e.,

$$R_k\left(\frac{1}{1-\epsilon}\right) \sim \tfrac{1}{2}e^{h_1}$$

with $h_1 \sim \sqrt{2\epsilon}\cdot k$, and so by (42)

$$\text{(43a)} \qquad \underset{-(1-\epsilon)\leqq Z\leqq 1-\epsilon}{\text{Max}} (|\, P_k(Z)\,|) = 2e^{-h_1}$$

with $h_1 \sim \sqrt{2\epsilon}\cdot k$, if $\epsilon \ll 1$, $\sqrt{2\epsilon}\cdot k \gg 1$.

Consider next the choice $P_k(Z) \equiv Z^k$. Then clearly

$$\underset{-(1-\epsilon)\leqq Z\leqq 1-\epsilon}{\text{Max}} (|\, P_k(Z)\,|) = (1-\epsilon)^k.$$

The logarithm of the right hand side is $\ln(1-\epsilon)\cdot k$, i.e., for $\epsilon \ll 1$ it is $\sim \epsilon\cdot k$. Hence in this case

$$\text{(43b)} \qquad \underset{-(1-\epsilon)\leqq Z\leqq 1-\epsilon}{\text{Max}} (|\, P_k(Z)\,|) = e^{-h_2}$$

with $h_2 \sim \epsilon\cdot k$, if $\epsilon \ll 1$.

Comparing (43a) and (43b), and remembering (34) shows that the speed of uniform convergence, i.e., the speed of decrease of $\bar{d}_k$, compares as follows for the choices of $P_k(Z)$ under consideration—namely, the "optimum" choice of $P_k(Z)$ [i.e., (41)], and the "ordinary" (no means!) choice of $P_k(Z)$ (i.e., $\equiv Z^k$): In the first case the increase of k that e^{-1}-folds $\bar{d}_k$ (asymptotically!) is $\Delta k = 1/\sqrt{2\epsilon}$, in the second case that increase is $\Delta k = 1/\epsilon$. Thus the first choice accelerates the convergence over the second choice in the ratio $\sqrt{2\epsilon}:\epsilon = \sqrt{2/\epsilon}$.

9. Let us now return to the definition (38) of $R_k(Z)$ [on which the "optimum" definition (41) of $P_k(Z)$ is based]. (38) is transcendental, the equivalent (39) is irrational. It is desirable to replace these by a rational definition. Such a definition obtains, in the form of a two-step recursion, from the identity

$$\cos[(k+1)u] + \cos[(k-1)u] \equiv 2\cos u \cos(ku).$$

In view of (38) this gives

$$R_{k+1}(Z) + R_{k-1}(Z) \equiv 2ZR_k(Z),$$

i.e.,

$$\text{(44)} \qquad R_{k+1}(Z) \equiv 2ZR_k(Z) - R_{k-1}(Z) \qquad (k = 1, 2, \cdots).$$

This relation, together with the "starting conditions"

$$\text{(45)} \qquad R_0(Z) \equiv 1, \qquad R_1(Z) \equiv Z,$$

defines the $R_k(Z)$ completely.

Now (41) permits us to pass to $P_k(Z)$. Then (44) becomes

$$P_{k+1}(Z) \equiv 2\frac{Z}{1-\epsilon}\frac{R_k\left(\frac{1}{1-\epsilon}\right)}{R_{k+1}\left(\frac{1}{1-\epsilon}\right)}P_k(Z) - \frac{R_{k-1}\left(\frac{1}{1-\epsilon}\right)}{R_{k+1}\left(\frac{1}{1-\epsilon}\right)}P_{k-1}(Z),$$

i.e.,

$$P_{k+1}(Z) \equiv 2Z \frac{a_{k+1}}{1-\epsilon} P_k(Z) - a_{k+1} a_k P_{k-1}(Z), \tag{46}$$

where

$$a_l = \frac{R_{l-1}\left(\frac{1}{1-\epsilon}\right)}{R_l\left(\frac{1}{1-\epsilon}\right)}. \tag{47}$$

Putting $Z = 1/(1-\epsilon)$ in (44) and dividing by $R_k[1/(1-\epsilon)]$ gives

$$\frac{1}{a_{k+1}} = \frac{2}{1-\epsilon} - a_k. \tag{48}$$

Through (41), (45) becomes

$$P_0(Z) \equiv 1, \qquad P_1(Z) \equiv Z. \tag{49}$$

Also, (45) gives

$$a_1 = 1 - \epsilon. \tag{50}$$

It is convenient to introduce

$$b_l = \frac{a_l}{1-\epsilon}. \tag{51}$$

Then (50), (48) give

$$b_1 = 1, \tag{52a}$$

$$b_{k+1} = \frac{1}{2 - (1-\epsilon)^2 b_k} \qquad (k = 1, 2, \cdots). \tag{52b}$$

Next, (46) gives

$$P_{k+1}(Z) \equiv 2b_{k+1} Z P_k(Z) - (1-\epsilon)^2 b_{k+1} b_k P_{k-1}(Z).$$

Owing to (52b)

$$(1-\epsilon)^2 b_{k+1} b_k = 2b_{k+1} - 1,$$

hence the above equation can also be written like this:

$$P_{k+1}(Z) = 2b_{k+1}[ZP_k(Z) - P_{k-1}(Z)] + P_{k-1}(Z) \qquad (k = 1, 2, \cdots). \tag{53}$$

Finally by (34), (42)

$$\bar{d}_k = \frac{1}{R_k\left(\frac{1}{1-\epsilon}\right)},$$

hence by (45), (47)

$$\bar{d}_k = a_1 \cdots a_k,$$

and so by (51)

$$\bar{d}_k = (1 - \epsilon)^k b_1 \cdots b_k . \tag{54}$$

10. We can now pass from the $P_k(Z)$ to the $\mathfrak{n}^{(k)}$, of course with the help of (16). We replace Z by the $2n$th order matrix E in both equations of (49) as well as in (53). Thus in all three equations both sides become $2n$th order matrices. We apply these to the $2n$th order vector $\{\xi, \alpha\}$. In this way three equations obtain, each one having $2n$th order vectors on both sides. These are as follows:

From the first equation of (49), using (16):

$$\{\mathfrak{n}^0, \alpha\} = \{\xi, \alpha\},$$

i.e.,

$$\mathfrak{n}^0 = \xi. \tag{55}$$

From the second equation of (49), using (16):

$$\{\mathfrak{n}^1, \alpha\} = E\{\xi, \alpha\},$$

i.e., recalling (2), (3):

$$\mathfrak{n}^1 = F\{\xi, \alpha\}. \tag{56}$$

From (53), using (16):

$$\{\mathfrak{n}^{k+1}, \alpha\} = 2b_{k+1}(E\{\mathfrak{n}^k, \alpha\} - \{\mathfrak{n}^{k-1}, \alpha\}) + \{\mathfrak{n}^{k-1}, \alpha\},$$

i.e., again recalling (2), (3):

$$\{\mathfrak{n}^{k+1}, \alpha\} = 2b_{k+1}[(F\{\mathfrak{n}^k, \alpha\}, \alpha) - \{\mathfrak{n}^{k-1}, \alpha\}] + \{\mathfrak{n}^{k-1}, \alpha\}$$
$$= 2b_{k+1}(F\{\mathfrak{n}^k, \alpha\} - \mathfrak{n}^{k-1}) + \{\mathfrak{n}^{k-1}, \alpha\},$$

i.e.,

$$\mathfrak{n}^{k+1} = 2b_{k+1}(F\{\mathfrak{n}^k, \alpha\} - \mathfrak{n}^{k-1}) + \mathfrak{n}^{k-1}, \qquad (k = 1, 2, \cdots). \tag{57}$$

11. We have obtained an inductive definition of the sequence $\mathfrak{n}^0, \mathfrak{n}^1, \mathfrak{n}^2, \cdots$. This is based on another, inductively defined, (numerical) sequence $b_1, b_2, \cdots$. Actually the two inductions can proceed concurrently. We will now restate these.

The b_k induction is given by (52a), (52b):

$$b_1 = 1, \tag{Ia}$$

$$b_{k+1} = \frac{1}{2 - (1 - \epsilon)^2 b_k} \qquad (k = 1, 2, \cdots). \tag{Ib}$$

The $\mathfrak{n}^k$ induction is given by (55), (56), (57):

$$\mathfrak{n}^0 = \xi, \tag{IIa}$$

$$\mathfrak{n}^1 = F\{\xi, \alpha\}, \tag{IIb}$$

$$\mathfrak{n}^{k+1} = 2b_{k+1}(F\{\mathfrak{n}^k, \alpha\} - \mathfrak{n}^{k-1}) + \mathfrak{n}^{k-1} \qquad (k = 1, 2, \cdots). \tag{IIc}$$

We also restate the formula (54) for $\bar{d}_k$:

$$\bar{d}_k = (1 - \epsilon)^k b_1 \cdots b_k . \tag{III}$$

This can be expressed inductively:

(IIIa) $$\bar{d}_0 = 1,$$

(IIIa) $$\bar{d}_{k+1} = (1 - \epsilon)b_{k+1}\bar{d}_k \qquad (k = 0, 1, 2, \cdots).$$

It is worthwhile to compare this process, and in particular its central piece (II) (which produces the sequence $\mathfrak{n}^0, \mathfrak{n}^1, \mathfrak{n}^2, \cdots$), with the ordinary iterative process, i.e., with (9) (which produces the sequence $\xi^0, \xi^1, \xi^2, \cdots$). (II) and (9) give the same $\mathfrak{n}^k$ and ξ^k for $k = 0, 1$, but they differ for $k = 2, 3, \cdots$, i.e., for $\mathfrak{n}^{k+1}$ and ξ^{k+1} for $k = 1, 2, \cdots$. Even here the first step in forming $\mathfrak{n}^{k+1}$ is the same as the (only) step in forming ξ^{k+1}, the application of the original correction step $F\{\cdots, \mathfrak{a}\}$ (cf. 2). In forming $\mathfrak{n}^{k+1}$, however, this is followed by the further step $2b_{k+1}(\cdots - \mathfrak{n}^{k-1}) + \mathfrak{n}^{k-1}$. This is clearly an extrapolation from $\mathfrak{n}^{k-1}$ with the (excess) factor $2b_{k+1} - 1$. Note, that (I) implies $\frac{1}{2} < b_l < 1$ (for all $l = 1, 2, \cdots$), hence $0 < 2b_l - 1 < 1$. Thus the extrapolation (excess) factor lies between 0 and 1.

Now it is by no means unusual that an iterative correction method is improved by combination with extrapolation steps. The noteworthy circumstance is rather, that, in going from $\mathfrak{n}^k$ to $\mathfrak{n}^{k+1}$, the extrapolation should issue from $\mathfrak{n}^{k-1}$. It is also of interest, that a "universal" and "optimum" sequence of extrapolation factors (i.e., the $2b_{k+1} - 1$) could be determined [by (I)].

12. The procedure summarized in **11** is complete, but it is based on the knowledge of a Hermitian G fulfilling (32a) [or equivalently (32b)]. Thus there remains the problem of constructing such a G.

More precisely, we need the F of (2), i.e., the G, H of (4). These are linked by the relation (7), which we restate:

(58) $$G = I - HA.$$

A is, of course, given. Thus H is arbitrary, it determines G by (58), and this G must then be Hermitian and fulfilling (32a). These conditions can also be stated in terms of $I - G = HA$: The Hermitian character of G is equivalent to that of HA. (32a) is equivalent to

(59) $$\epsilon \leqq \lambda \leqq 2 - \epsilon$$

for all characteristic values of λ of HA.

We repeat: We are looking for an H that makes HA Hermitian and fulfills (59).

We will now describe two procedures that achieve this:

First, put

(60) $$H = \alpha A^* \qquad (\alpha > 0).\dagger$$

Then (58) gives

(61) $$G = I - \alpha A^* A.$$

$HA = \alpha A^* A$ is obviously Hermitian, it is also positive-definite. Hence the smallest characteristic value of HA is $|\alpha A^* A|_l = \alpha |A^* A|_l = \alpha(|A|_l)^2$, and the largest characteristic value of HA is $|\alpha A^* A|_u = \alpha |A^* A|_u = \alpha(|A|_u)^2$. Consequently

† A^* is the "adjoint" of A, i.e., its complex-conjugate transposed.

(59) means that

$$\alpha(|A|_l)^2 \geqq \epsilon, \tag{62a}$$

$$\alpha(|A|_u)^2 \leqq 2 - \epsilon. \tag{62b}$$

Assume that we know that

$$0 < a \leqq |A|_l \leqq |A|_u \leqq b, \tag{63}$$

i.e., so that a, b are known. Then (62a), (62b) can be guaranteed by prescribing

$$\alpha a^2 = \epsilon, \qquad \alpha b^2 = 2 - \epsilon, \tag{64}$$

i.e.,

$$\alpha = \frac{2}{a^2 + b^2}, \tag{65}$$

$$\epsilon = \frac{2a^2}{a^2 + b^2}. \tag{66}$$

Now put

$$f = \frac{b}{a}. \tag{67}$$

Then (66) becomes

$$\epsilon = \frac{2}{f^2 + 1}. \tag{68}$$

Second, assume that A is Hermitian and positive-definite. In this case put

$$H = \alpha I \qquad (\alpha > 0). \tag{69}$$

Then (58) gives

$$G = I - \alpha A. \tag{70}$$

$HA = \alpha A$ is clearly Hermitian and positive-definite. Hence the smallest characteristic value of HA is $|\alpha A|_l = \alpha |A|_l$, and the largest characteristic value of HA is $|\alpha A|_u = \alpha |A|_u$. Consequently (59) means that

$$\alpha |A|_l \geqq \epsilon, \tag{71a}$$

$$\alpha |A|_u \leqq 2 - \epsilon. \tag{71b}$$

Assuming again the validity of (63), we can guarantee (71a), (71b) by prescribing

$$\alpha a = \epsilon, \qquad \alpha b = 2 - \epsilon, \tag{72}$$

i.e.,

$$\alpha = \frac{2}{a + b}, \tag{73}$$

$$\epsilon = \frac{2a}{a + b}. \tag{74}$$

Using (67) again, (74) becomes

$$\epsilon = \frac{2}{f+1}, \tag{75}$$

13. The results of **12** deserve restatement and some comments. The common assumptions of both parts of **12** are (63), (67):

$$0 < a \leqq |A|_l \leqq |A|_u \leqq b, \tag{IVa}$$

$$f = \frac{b}{a}. \tag{IVb}$$

The result of the first part is contained in (60), (61), (65), (66), (68):

$$H = \alpha A^* \tag{Va}$$

$$G = I - \alpha A^* A, \tag{Vb}$$

$$\alpha = \frac{2}{a^2 + b^2}, \tag{Vc}$$

$$\epsilon = \frac{2a^2}{a^2 + b^2} = \frac{2}{f^2 + 1}, \tag{Vd}$$

A being otherwise unrestricted.

The result of the second part is contained in (69), (70), (73), (74), (75):

$$H = \alpha I, \tag{VIa}$$

$$G = I - \alpha A, \tag{VIb}$$

$$\alpha = \frac{2}{a+b}, \tag{VIc}$$

$$\epsilon = \frac{2a}{a+b} = \frac{2}{f+1}, \tag{VId}$$

A being assumed Hermitian and positive-definite.

(Va), (Vb) show that the first case is related to the iterative "steepest descent" methods; (VIa), (VIb) show that the second case is related to the iterative "relaxation" methods.

Our derivation makes it plausible why the former are of universal applicability, while the latter are limited to Hermitian and positive-definite matrices—i.e., if the problem arises from the difference equation treatment of partial differential equations of the elliptic type [2nd order, s ($= 2, 3, \cdots$) variables, cf. **1** and again **14**], to the self-adjoint, elliptic case.

In general $f \gg 1$. Then in the first case $\epsilon \sim 2f^{-2}$ [by (Vd)], and in the second case $\epsilon \sim 2f^{-1}$ [by (VId)]. Thus the first case gives a much smaller ϵ than the second case, i.e., in view of the remarks at the end of **8**, a much slower convergence of the iterative process.

This observation illustrates the general experience that whenever relaxation-type procedures are applicable, the convergence is significantly faster than otherwise.

14. We now pass to the consideration of the difference equation system for an elliptic partial differential equation [2nd order, s $(= 2, 3, \cdots)$ variables].

Let the partial differential equation be

$$-\sum_{i=1}^{s} \frac{\partial}{\partial x_i}\left(a^i \frac{\partial \xi}{\partial x_i}\right) = \alpha, \tag{76}$$

where $x_1, \cdots, x_s$ are the independent variables, $\xi \equiv \xi(x_1, \cdots, x_s)$ is the dependent variable, and $a^i \equiv a^i(x_1, \cdots, x_s)$, $(i = 1, \cdots, s)$ and $\alpha \equiv \alpha(x_1, \cdots, x_s)$ are known functions of $x_1, \cdots, x_s$. Also

$$0 < \bar{a}^i \leqq a^i(x_1, \cdots, x_s) \leqq \bar{b}^i \qquad (i = 1, \cdots, s), \tag{77}$$

the $\bar{a}^i$, $\bar{b}^i$ $(i = 1, \cdots, s)$ being known constants. Finally the domain of the $x_1, \cdots, x_s$ is

$$0 \leqq x_i \leqq L_i \qquad (i = 1, \cdots, s), \tag{78}$$

and the boundary condition is

$$\xi = 0 \quad \text{for} \quad x_i = 0 \quad \text{or} \quad L_i \qquad (i = 1, \cdots, s). \tag{79}$$

In order to pass to difference equations, we introduce a lattice

$$x_i = \eta_i \Delta x_i \left(\Delta x_i = \frac{L_i}{N_i}\right), \tag{80}$$

where in some cases

$$\eta_i = 0, 1 \cdots, N_i - 1, N_i, \tag{80a}$$

in others

$$\eta_i = 1, \cdots, N_i - 1, \tag{80b}$$

and in others again

$$\eta_i = \tfrac{1}{2}, \tfrac{3}{2}, \cdots, N_i - \tfrac{1}{2} \qquad (i = 1, \cdots, s). \tag{80c}$$

Of course, $N_i = 2, 3, \cdots$, and it expresses the fineness of this lattice in the x_i-direction.

We write

$$\xi(x_1, \cdots x_s) = \xi_{\eta_1 \cdots \eta_s}, \tag{81}$$

using (80a). These $\xi_{\eta_1 \cdots \eta_s}$ are the unknowns, but since (79) gives

$$\xi_{\eta_1 \cdots \eta_s} = 0 \quad \text{for} \quad \eta_i = 0, N_i \qquad (i = 1, \cdots, s), \tag{82}$$

the unknown character is actually restricted to (80b).

It is convenient to use with a^i (80b) for the η_j with $j \neq i$, and (80c) for η_i :

$$a^i(x_1, \cdots, x_s) = a^i_{\eta_1 \cdots \eta_s}, \tag{83}$$

and for α (80b) throughout:

$$\alpha(x_1, \cdots, x_s) = \alpha_{\eta_1 \cdots \eta_s}, \tag{84}$$

these being known quantities.

Now (76) is best stated for (80b). It becomes

$$(85)\quad -\sum_{i=1}^{s}\left(\frac{N_i}{L_i}\right)^2 [a^i_{\eta_1\cdots\eta_i+\frac{1}{2}\cdots\eta_s}(\xi_{\eta_1\cdots\eta_i+1\cdots\eta_s}-\xi_{\eta_1\cdots\eta_i\cdots\eta_s}) - a^i_{\eta_1\cdots\eta_i-\frac{1}{2}\cdots\eta_s}(\xi_{\eta_i\cdots\eta_i\cdots\eta_s}-\xi_{\eta_1\cdots\eta_i-1\cdots\eta_s})] = \alpha_{\eta_1\cdots\eta_s}.$$

We can view (85) as the equivalent of (1), with the following provisos: The complex $\eta_1, \cdots, \eta_s$, according to (80b), stands for the vector-index in (1). Hence the order of the matrix A is

$$(86)\quad \eta = \prod_{i}^{s}(N_i - 1).$$

The $\xi_{\eta_1\cdots\eta_s}$ are therefore the components of the (unknown) vector ξ, the $\alpha_{\eta_1\cdots\eta_s}$ are the components of the (known) vector α. The left-hand side of (85) then defines the (known) matrix A. Hence

$$(87)\quad A = \sum_{i=1}^{s}\left(\frac{N_i}{L_i}\right)^2 A_i,$$

where the matrix A_i is defined by

$$(88)\quad \begin{aligned} A_i\xi &= \xi^+,\\ \xi^+_{\eta_1\cdots\eta_s} &= -a^i_{\eta_1\cdots\eta_i+\frac{1}{2}\cdots\eta_s}(\xi_{\eta_1\cdots\eta_s+1\cdots\eta_s}-\xi_{\eta_1\cdots\eta_i\cdots\eta_s})\\ &\quad +a^i_{\eta_1\cdots\eta_i-\frac{1}{2}\cdots\eta_s}(\xi_{\eta_1\cdots\eta_i\cdots\eta_s}-\xi_{\eta_1\cdots\eta_i-1\cdots\eta_s}). \end{aligned}$$

Furthermore, clearly

$$(89)\quad A_i = B_i^* B_i$$

[cf. footnote on page 177, where

$$(90)\quad \begin{aligned} B_i\xi &= \xi^+,\\ \xi^+_{\eta_1\cdots\eta_s} &= \sqrt{a^i_{\eta_1\cdots\eta_i+\frac{1}{2}\cdots\eta_s}}\,(\xi_{\eta_1\cdots\eta_i+1\cdots\eta_s}-\xi_{\eta_1\cdots\eta_i\cdots\eta_s}). \end{aligned}$$

In order to apply the results of **13**, we now need estimates of $|A|_l$, $|A|_u$, in accordance with (IV) in **13.** From (87)

$$(91)\quad \left.\begin{aligned} |A|_l &\geqq \sum_{i=1}^{s}\left(\frac{N_i}{L_i}\right)^2 |A_i|_l,\\ |A|_u &\leqq \sum_{i=1}^{s}\left(\frac{N_i}{L_i}\right)^2 |A_i|_u. \end{aligned}\right\}$$

From (89)

$$(92)\quad \left.\begin{aligned} |A_i|_l &= (|B_i|_l)^2,\\ |A_i|_u &= (|B_i|_u)^2. \end{aligned}\right\}$$

Finally, designate A_i, B_i with $a^i_{\eta_1\cdots\eta_i\cdots\eta_s} = 1$ [cf. the remark preceding (83)!] by A_i^0, B_i^0. Then clearly

$$(93)\quad \left.\begin{aligned} |B_i|_l &\geqq \sqrt{\bar{a}^i}\,|B_i^0|_l,\\ |B_i|_u &\leqq \sqrt{\bar{b}^i}\,|B_i^0|_u. \end{aligned}\right\}$$

Combining both sides of (93) with (92) gives

$$\left.\begin{aligned} |A_i|_l &\geqq \bar{a}^i |A_i^0|_l, \\ |A_i|_u &\leqq \bar{b}^i |A_i^0|_u, \end{aligned}\right\} \tag{94}$$

and combining (94) with (91) gives

$$\left.\begin{aligned} |A|_l &\geqq \sum_{i=1}^{s} \bar{a}^i \left(\frac{N_i}{L_i}\right)^2 |A_i^0|_l, \\ |A|_u &\leqq \sum_{i=1}^{s} \bar{b}^i \left(\frac{N_i}{L_i}\right)^2 |A_i^0|_u. \end{aligned}\right\} \tag{95}$$

Now consider A_i^0. Applying (88) with $a^i_{\eta_1 \cdots \eta_i \cdots \eta_s} = 1$ shows, that the role of the η_j, $j \neq i$, is now irrelevant in determining $|A_i^0|_l$, $|A_i^0|_u$. Hence we may write in place of (88)

$$\left.\begin{aligned} A_i^0 \xi &= \xi^+, \\ \xi^+_{\eta_i} &= -\xi_{\eta_i+1} + 2\xi_{\eta_i} - \xi_{\eta_i-1}. \end{aligned}\right\} \tag{96}$$

This operator is Hermitian, its characteristic vectors are the $\xi^{m_i}(m_i = 1, \cdots, N_i - 1)$ with

$$\xi^{m_i}_{\eta_i} = \sin \frac{\pi m_i \eta_i}{N_i}, \tag{97}$$

the characteristic value of ξ^{m_i} being

$$\lambda^{\eta_i} = 2 - 2\cos\frac{\pi m_i}{N_i} = 4\sin^2\frac{\pi m_i}{2N_i}. \tag{98}$$

Hence

$$\left.\begin{aligned} |A_i^0|_l &= \operatorname*{Min}_{m_i} \lambda^{m_i} = \lambda^1 = 4\sin^2\frac{\pi}{2N_i}, \\ |A_i^0|_u &= \operatorname*{Max}_{m_i} \lambda^{m_i} = \lambda^{N_i-1} = 4\cos^2\frac{\pi}{2N_i}. \end{aligned}\right\} \tag{99}$$

Combining (95) and (99) gives

$$\left.\begin{aligned} |A|_l &\geqq 4\sum_{i=1}^{s} \bar{a}_i \left(\frac{N_i}{L_i}\right)^2 \sin^2\frac{\pi}{2N_i}, \\ |A|_u &\leqq 4\sum_{i=1}^{s} \bar{b}_i \left(\frac{N_i}{L_i}\right)^2 \cos^2\frac{\pi}{2N_i}. \end{aligned}\right\} \tag{100}$$

Hence we can put in (IVa)

$$\left.\begin{aligned} a &= 4\sum_{i=1}^{s} \bar{a}_i \left(\frac{N_i}{L_i}\right)^2 \sin^2\frac{\pi}{2N_i}, \\ b &= 4\sum_{i=1}^{s} \bar{b}_i \left(\frac{N_i}{L_i}\right)^2 \cos^2\frac{\pi}{2N_i}. \end{aligned}\right\} \tag{101}$$

Therefore (IVb) gives

$$f = \frac{\sum_{i=1}^{s} \bar{b}_i \left(\frac{N_i}{L_i}\right)^2 \cos^2 \frac{\pi}{2N_i}}{\sum_{i=1}^{s} \bar{a}_i \left(\frac{N_i}{L_i}\right)^2 \sin^2 \frac{\pi}{2N_i}}. \tag{102}$$

If

$$\operatorname*{Max}_{i=1,\cdots,s} \frac{\bar{b}_i}{\bar{a}_i} = \bar{M}, \qquad \operatorname*{Max}_{i=1,\cdots,s} N_i = \bar{N}, \tag{103}$$

Then (102) gives

$$f \leqq \bar{M} \cot^2 \frac{\pi}{2\bar{N}}. \tag{104}$$

and, since

$$\tan \frac{\pi}{2\bar{N}} \geqq \frac{\pi}{2\bar{N}}, \qquad \cot \frac{\pi}{2\bar{N}} \leqq \frac{2\bar{N}}{\pi},$$

a fortiori

$$f \leqq \frac{4}{\pi^2} \bar{M} \bar{N}^2. \tag{105}$$

Now (VId) in **13** gives $\epsilon = 2/(f+1)$ and the remarks at the end of **8** give for the error-e^{-1}-folding increase of k (asymptotically!)

$$\Delta k \sim \frac{1}{\sqrt{2\epsilon}} = \frac{1}{2}\sqrt{f+1} \sim \frac{1}{2}\sqrt{f},$$

i.e.,

$$\Delta k \sim \frac{1}{2}\sqrt{f}. \tag{106}$$

Hence, in view of (105),

$$\Delta k \lesssim \frac{1}{\pi}\sqrt{\bar{M}}\bar{N}. \tag{107}$$

15. We restate the results of **14.** The elliptic partial differential equation is given in (76), the subsidiary conditions in (77)–(79), the lattice is defined in (80) and (81)–(84), the difference equation system in (85). We do not restate these.

The a, b, f of (IV) in **13** are given in (101), (102):

$$a = 4 \sum_{i=1}^{s} \bar{a}_i \left(\frac{N_i}{L_i}\right)^2 \sin^2 \frac{\pi}{2N_i}, \tag{VIIa}$$

$$b = 4 \sum_{i=1}^{s} \bar{b}_i \left(\frac{N_i}{L_i}\right)^2 \cos^2 \frac{\pi}{2N_i}, \tag{VIIb}$$

$$f = \frac{\sum_{i=1}^{s} \bar{b}_i \left(\frac{N_i}{L_i}\right) \cos^2 \frac{\pi}{2N_i}}{\sum_{i=1}^{s} \bar{a}_i \left(\frac{N_i}{L_i}\right) \sin^2 \frac{\pi}{2N_i}}. \tag{VIIc}$$

If

$$\text{(VIIIa)} \qquad \operatorname*{Max}_{i=1,\cdots,s} \frac{\bar{b}_i}{\bar{a}_i} = \bar{M}, \qquad \operatorname*{Max}_{i=1,\cdots,s} N_i = \bar{N},$$

then

$$\text{(VIIIb)} \qquad f \leqq \bar{M} \cot^2 \frac{\pi}{2\bar{N}} \leqq \frac{4}{\pi^2} \bar{M}\bar{N}^2 .$$

Los Alamos, New Mexico

1. A. BLAIR, N. METROPOLIS, J. VON NEUMANN, A. H. TAUB & M. TSINGOU, *A Study of a Numerical Solution to a Two-dimensional Hydrodynamical Problem*, Los Alamos Report LA-2165, 1958.

2. S. BERNSTEIN, *Leçons sur les Propriétés Extrémales et la Meilleure Approximation des Fonctions Analytiques d'une Variable Réelle*, Gauthier-Villars, Paris, 1926, p. 7–8. Also P. Chebyshev, *Collected Works*, v. 2.