THE USE OF INTEGRAL CONSTRAINTS IN DESIGNING FINITE-DIFFERENCE SCHEMES FOR THE TWO-DIMENSIONAL ADVECTION EQUATION

A. Arakawa
Department of Atmospheric Sciences
University of California, Los Angeles, USA

1. INTRODUCTION

In this lecture I shall discuss how integral constraints can be used in designing finite-difference schemes for the two-dimensional advection equation with a space-dependent current. Taking the vorticity equation for two-dimensional nondivergent flow as an example, the possible cause for nonlinear computational instability is demonstrated. Then, following Arakawa (1966), a method that prevents the instability through the conservation of enstrophy and/or energy in a finite-difference vorticity equation is described and some numerical examples are presented. This method is subsequently applied to the problem of computational boundary conditions. It is then generalized to the two-dimensional advection equation with a divergent current. The method is also applied to the design of dissipative schemes and schemes that conserve non-quadratic quantities.

2. THE POSSIBLE CAUSE FOR NONLINEAR COMPUTATIONAL INSTABILITY

2.1 Aliasing error

When grid points with the interval Δx are used, wavelengths shorter than $2\Delta x$ (wavenumbers larger than $k_{max} = \pi/\Delta x$) cannot be resolved. To see what happens in such a situation, consider a wave with the wavenumber $k > k_{max}$. At the grid points $x = i\Delta x$, where i is an integer, $2k_{max}x = 2\pi i$. Thus

$$\sin kx = -\sin (2k_{max} - k)x, \qquad \cos kx = \cos (2k_{max} - k)x$$

at the grid points. Since $k > k_{max}$, $2k_{max} - k < k_{max}$. This means that, recognizing only the grid point values, the grid misrepresents the unresolvable wave with the wavenumber k as a resolvable wave with the wavenumber $k^* = 2k_{max} - k$. The error due this spectral misrepresentation is called aliasing error.

In numerical integration of a nonlinear equation (or a linear equation whose coefficients vary in space), the product term in the equation can generate waves that have wavenumbers larger than k_{max}. The spectrum of the solution can then be affected by the aliasing error, and even an instability known as nonlinear computational instability can occur, as shown below.

2.2 A finite-difference approximation to the vorticity equation for two-dimensional nondivergent flow

Let us consider two-dimensional nondivergent flow as an example. By assumption

$$\nabla \cdot \mathbf{v} = 0, \tag{1}$$

where $\mathbf{v}$ is the two-dimensional velocity and ∇ is the two-dimensional del operator. The time change of the flow is governed by the vorticity equation

$$\frac{\partial \zeta}{\partial t} = - \mathbf{v} \cdot \nabla \zeta, \tag{2}$$

where t is the time, ζ is the vorticity, $\mathbf{k} \cdot \nabla \times \mathbf{v}$, and $\mathbf{k}$ is the unit vector normal to the plane of motion. Using (1), we may rewrite (2) as

$$\frac{\partial \zeta}{\partial t} = - \nabla \cdot (\mathbf{v} \zeta). \tag{3}$$

The form (2) will be referred to as the advection form and (3) as the flux convergence form.

The continuity equation (1) is identically satisfied with

$$\mathbf{v} = \mathbf{k} \times \nabla \psi, \tag{4}$$

where ψ is the streamfunction. The vorticity is then given by

$$\zeta = \nabla^2 \psi . \tag{5}$$

Using the streamfunction and cartesian coordinates (x,y), the vorticity equation may be written as

$$\frac{\partial \zeta}{\partial t} = J(\zeta, \psi), \tag{6}$$

where $J(\zeta,\psi)$ is the Jacobian operator defined by

$$J(\zeta,\psi) = \frac{\partial \zeta}{\partial x} \frac{\partial \psi}{\partial y} - \frac{\partial \zeta}{\partial y} \frac{\partial \psi}{\partial x} . \tag{7}$$

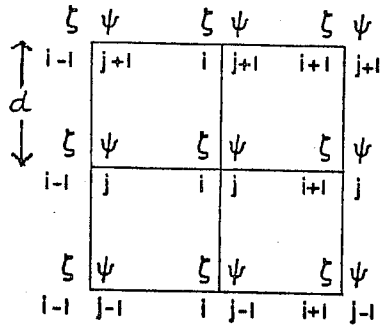

Fig. 1 A square grid showing the indices for grid points

For a square grid as shown by Fig. 1, use of the simplest centered finite-difference approximations to the Jacobian and Laplacian operators gives

$$\frac{\partial}{\partial t}\zeta_{i,j} = \frac{1}{4d^2}\left[\left(\zeta_{i+1,j} - \zeta_{i-1,j} \right)\left(\psi_{i,j+1} - \psi_{i,j-1} \right) - \left(\zeta_{i,j+1} - \zeta_{i,j-1} \right)\left(\psi_{i+1,j} - \psi_{i-1,j} \right) \right] \tag{8}$$

and

$$\zeta_{i,j} = \frac{1}{d^2}\left(\psi_{i+1,j} + \psi_{i-1,j} + \psi_{i,j+1} + \psi_{i,j-1} - 4\psi_{i,j} \right). \tag{9}$$

2.3 An example of nonlinear computational instability

Following Phillips (1959), we show that the aliasing error can cause computational instability when the above finite-difference Jacobian is used. Let us consider the initial condition such that

$$\psi = \psi_1 + \psi_2 , \tag{10}$$

where

$$(\psi_1)_{i,j} = \left(C\cos\frac{\pi i}{2} + S\sin\frac{\pi i}{2} \right)\sin\frac{2\pi j}{3}, \quad (\psi_2)_{i,j} = U\cos\pi i \sin\frac{2\pi j}{3} . \tag{11}$$

Substituting (10) and (11) into (9) and the right hand side of (8), we obtain, after some manipulations,

$$\frac{\partial}{\partial t}\zeta_{i,j} = \frac{\sqrt{3}U}{4d^4}\left[-C\left(\sin\frac{3\pi i}{2} - \sin\frac{\pi i}{2} \right) + S\left(\cos\frac{3\pi i}{2} + \cos\frac{\pi i}{2} \right) \right]\sin\frac{4\pi j}{3} . \tag{12}$$

But due to aliasing,

$$\sin\frac{3\pi i}{2} \to -\sin\frac{\pi i}{2}, \quad \cos\frac{3\pi i}{2} \to \cos\frac{\pi i}{2}, \quad \sin\frac{4\pi j}{2} \to -\sin\frac{2\pi j}{3}. \tag{13}$$

Then (12) becomes

$$\frac{\partial}{\partial t}\zeta_{i,j} = -\frac{\sqrt{3}U}{2d^4}\left(C\sin\frac{\pi i}{2} + S\cos\frac{\pi i}{2} \right)\sin\frac{2\pi j}{3} . \tag{14}$$

From (14) and (9), we find

$$\frac{\partial}{\partial t}\psi_{i,j} = \frac{\sqrt{3}U}{10d^2}\left(C\sin\frac{\pi i}{2} + S\cos\frac{\pi i}{2} \right)\sin\frac{2\pi j}{3} . \tag{15}$$

Substituting (10) into the left hand side of (15), we see that (10) is the solution for all t, due to aliasing, if

$$\frac{dC}{dt} = \frac{\sqrt{3}U}{10d^2}S , \quad \frac{dS}{dt} = \frac{\sqrt{3}U}{10d^2}C , \quad \frac{dU}{dt} = 0 . \tag{16}$$

Then U = constant and

$$C = S \propto e^{\sqrt{3}Ut/10d^2} \quad \text{or} \quad C = -S \propto e^{-\sqrt{3}Ut/10d^2} . \tag{17}$$

Note that (15) is in phase with ψ_1 for the growing mode and 180° out of phase for the decaying mode. Typically, $U/d^2 \sim 10^{-5}$ sec^{-1}; then the growth rate of the growing mode is given by $\sqrt{3}U/10d^2 \sim 1.7 \times 10^{-6}$ sec^{-1} $\sim$ (6.7 days)$^{-1}$.

As we can see from the above example, nonlinear computational instability can occur even when time is continuous. This is in sharp contrast to the more

familiar linear computational instability which occurs in commonly used schemes when the Courant-Friedrichs-Lewy stability criterion is violated.

3. PROPERTIES OF THE JACOBIAN OPERATOR AND CONSERVATION LAWS FOR TWO-DIMENSIONAL NONDIVERGENT FLOW

Before proceeding to the problem of overcoming nonlinear computational instability, let us consider the Jacobian $J(p,q)$ in its differential form, where p and q are arbitrary functions of space. We note that

$$J(p,q) = \mathbf{k}\cdot\nabla p\times\nabla q = \mathbf{k}\cdot\nabla\times(p\nabla q) = -\mathbf{k}\cdot\nabla\times(q\nabla p) \tag{18}$$

and

$$J(p,q) = -J(q,p). \tag{19}$$

Using (18) and Stokes' theorem, we find

$$\overline{J(p,q)} = 0, \tag{20}$$

where the overbar denotes the area mean over a domain along the boundary of which either p or q is constant, or over a periodic domain. Furthermore, we can easily show that

$$\overline{pJ(p,q)} = 0 \tag{21}$$

and

$$\overline{qJ(p,q)} = 0 \tag{22}$$

for the same domain.

In our problem $p = \zeta$ and $q = \psi$. Let us assume a bounded domain along the boundary of which ψ = constant so that there is no inflow or outflow across the boundary. From (6) and (20), we obtain

$$\frac{\partial}{\partial t}\overline{\zeta} = 0 \tag{23}$$

Thus the mean vorticity is conserved. From (6) and (21), we obtain

$$\frac{\partial}{\partial t}\frac{1}{2}\overline{\zeta^2} = 0. \tag{24}$$

Thus the enstrophy defined by $E \equiv \frac{1}{2}\overline{\zeta^2}$ is conserved. From (6) and (22), on the other hand, we obtain $\overline{\psi\partial\zeta/\partial t} = 0$. Using $\zeta = \nabla^2\psi$ and (23), we may rewrite this as

$$\frac{\partial}{\partial t}\frac{1}{2}\overline{(\nabla\psi)^2} = 0. \tag{25}$$

Since $\mathbf{v}^2 = (\mathbf{k}\times\nabla\psi)^2 = (\nabla\psi)^2$, (25) means that the kinetic energy $K \equiv \frac{1}{2}\overline{\mathbf{v}^2}$ is conserved.

We can express the streamfunction ψ as a series of orthogonal harmonic functions. Then

$$\psi = \sum_n \psi_n, \tag{26}$$

where the functions ψ_n are eigenfunctions of

$$\nabla^2\psi_n + \lambda_n^2 \psi_n = 0. \tag{27}$$

The parameter λ_n is the generalized wavenumber of the component ψ_n. Due to orthogonality, we find

$$K = \sum_n K_n, \quad K_n = \frac{1}{2}\overline{(\nabla\psi_n)^2}\ . \tag{28}$$

Also we can show from (27) that

$$E = \sum_n E_n, \quad E_n = \lambda_n^2 K_n. \tag{29}$$

Since both K and E are conserved, their ratio is also conserved. Then the average wavenumber λ defined by

$$\lambda^2 \equiv \frac{E}{K} = \frac{\sum_n \lambda_n^2 K_n}{\sum_n K_n} \tag{30}$$

is conserved. Since λ^2 is the mean of λ_n^2 with the weight K_n, conservation of λ^2 means that a systematic energy cascade toward higher wavenumbers is not possible in two-dimensional nondivergent flow, as pointed out by Fjørtoft (1953).

4. THE ENSTROPHY AND ENERGY CONSERVING FINITE-DIFFERENCE JACOBIAN

We now return to the problem of numerical solution of (6). If a finite-difference scheme could be constructed so as to conserve discrete analogs of either the enstrophy or the kinetic energy, nonlinear computational instability would be prevented. Furthermore, if both could be conserved, the average wavenumber would not change, and, therefore, a systematic cascade of energy toward higher wavenumbers would not be possible. Arakawa (1966) has pointed this out and showed that such finite-difference schemes can indeed be constructed.

For convenience, let us introduce a single index $\underline{i}$ to identify a grid point. The finite-difference approximation of the Jacobian at the point $\underline{i}$ may be written, in a general form, as

$$\sigma_{\underline{i}} J_{\underline{i}}(\zeta,\psi) = \sum_{\underline{i}'} \sum_{\underline{i}''} c_{\underline{i}',\underline{i}''} \zeta_{\underline{i}+\underline{i}'} \psi_{\underline{i}+\underline{i}''}\ , \tag{31}$$

where $\sigma_{\underline{i}}$ is the area represented by the grid point $\underline{i}$. In the case of a square grid with the grid interval d, $\sigma_{\underline{i}} = d^2$. The choice of coefficients $c_{\underline{i}',\underline{i}''}$ in (31) determines the order of accuracy and other properties of the finite-difference

approximation. Here we are particularly concerned with the properties

$$\sum_{\underline{i}} \sigma_{\underline{i}}\zeta_{\underline{i}}J_{\underline{i}}(\zeta,\psi) = 0 \tag{32}$$

and

$$\sum_{\underline{i}} \sigma_{\underline{i}}\psi_{\underline{i}}J_{\underline{i}}(\zeta,\psi) = 0. \tag{33}$$

When divided by the total area, these are discrete analogs of the integral constraints (21) and (22) on Jacobian, which lead to the enstrophy and energy conservation when $p = \zeta$ and $q = \psi$.

To find the condition to satisfy (32), it is convenient to rewrite (31) as

$$\sigma_{\underline{i}}J_{\underline{i}}(\zeta,\psi) = \sum_{\underline{i}'} a_{\underline{i},\underline{i+i}'}\zeta_{\underline{i+i}'} \quad , \tag{34}$$

where

$$a_{\underline{i},\underline{i+i}'} \equiv \sum_{\underline{i}''} c_{\underline{i}',\underline{i}''}\psi_{\underline{i+i}''} \quad . \tag{35}$$

Multiplying (34) by $\zeta_{\underline{i}}$, we obtain

$$\sigma_{\underline{i}}\zeta_{\underline{i}}J_{\underline{i}}(\zeta,\psi) = \sum_{\underline{i}'} a_{\underline{i},\underline{i+i}'}\zeta_{\underline{i}}\zeta_{\underline{i+i}'} \, . \tag{36}$$

The term $a_{\underline{i},\underline{i+i}'}\zeta_{\underline{i}}\zeta_{\underline{i+i}'}$ can be interpreted as the contribution of the interaction between the grid points $\underline{i}$ and $\underline{i+i}'$ to $\sigma_{\underline{i}}J_{\underline{i}}(\zeta,\psi)$. Similarly, $a_{\underline{i+i}',\underline{i}}\zeta_{\underline{i+i}'}\zeta_{\underline{i}}$ can be interpreted as the contribution of the same interaction to $\sigma_{\underline{i+i}'}J_{\underline{i+i}'}(\zeta,\psi)$. These two must have the same magnitude and the opposite sign to satisfy (32). Thus we must require

$$a_{\underline{i+i}',\underline{i}} = -a_{\underline{i},\underline{i+i}'} \quad , \quad \text{for all } \underline{i} \text{ and } \underline{i}', \tag{37}$$

to conserve enstrophy.

The finite-difference Jacobian (31) may also be written as

$$\sigma_{\underline{i}}J_{\underline{i}}(\zeta,\psi) = \sum_{\underline{i}''} b_{\underline{i},\underline{i+i}''}\psi_{\underline{i+i}''} \quad , \tag{38}$$

where

$$b_{\underline{i},\underline{i+i}''} = \sum_{\underline{i}'} c_{\underline{i}',\underline{i}''}\zeta_{\underline{i+i}'} \quad . \tag{39}$$

An argument similar to the above shows that we must require

$$b_{\underline{i+i}'',\underline{i}} = -b_{\underline{i},\underline{i+i}''} \quad \text{for all } \underline{i} \text{ and } \underline{i}'' \tag{40}$$

to conserve energy.

In these arguments, we did not specify the finite-difference form of the Laplacian, $\nabla^2\psi$. We can simply assume that the form of the Laplacian and the form of the kinetic energy, $\frac{1}{2}(\nabla\psi)^2$, maintain the relation $\overline{\psi\nabla^2\psi} = -\overline{(\nabla\psi)^2}$. The finite-difference Laplacian (9) satisfies this relation.

In the case of the finite-difference Jacobian (8) for a square grid with $\sigma_{\underline{i}} = d^2$, the coefficients $a_{\underline{i},\underline{i+i}'}$ can be identified as

$$a_{i,j;i+1,j} = (\psi_{i,j+1} - \psi_{i,j-1})/4 \tag{41}$$

$$a_{i,j;i-1,j} = -(\psi_{i,j+1} - \psi_{i,j-1})/4 \tag{42}$$

$$a_{i,j;i,j+1} = -(\psi_{i+1,j} - \psi_{i-1,j})/4 \tag{43}$$

$$a_{i,j;i,j-1} = (\psi_{i+1,j} - \psi_{i-1,j})/4 \,. \tag{44}$$

Here the original double indecies are used instead of the single index $\underline{i}$ to identify a grid point. Replacing i in (42) by i+1, we obtain

$$a_{i+1,j;i,j} = (\psi_{i+1,j+1} - \psi_{i+1,j-1})/4 \,. \tag{45}$$

By comparing (45) with (41) we see that (37) is not satisfied by these coefficients and, similarly, by the coefficients $a_{i,j+1;i,j}$ and $a_{i,j;i,j+1}$. We also find that (40) is not satisfied. Thus, the finite-difference Jacobian (8) maintains conservation of neither enstrophy nor energy.

To find a finite-difference Jacobian that satisfies (37) and (40), let us first consider the following three basic finite-difference Jacobians:

$$(J_1)_{i,j} \equiv \frac{1}{4d^2}\,[\,(\zeta_{i+1,j} - \zeta_{i-1,j})(\psi_{i,j+1} - \psi_{i,j-1}) - (\zeta_{i,j+1} - \zeta_{i,j-1})(\psi_{i+1,j} - \psi_{i-1,j})\,], \tag{46}$$

$$(J_2)_{i,j} \equiv \frac{1}{4d^2}\,[\,-(\zeta_{i+1,j+1} - \zeta_{i+1,j-1})\psi_{i+1,j} + (\zeta_{i-1,j+1} - \zeta_{i-1,j-1})\psi_{i-1,j} + (\zeta_{i+1,j+1} - \zeta_{i-1,j+1})\psi_{i,j+1} - (\zeta_{i+1,j-1} - \zeta_{i-1,j-1})\psi_{i,j-1}\,], \tag{47}$$

$$(J_3)_{i,j} \equiv \frac{1}{4d^2}\,[\,\zeta_{i+1,j}(\psi_{i+1,j+1} - \psi_{i+1,j-1}) - \zeta_{i-1,j}(\psi_{i-1,j+1} - \psi_{i-1,j-1}) - \zeta_{i,j+1}(\psi_{i+1,j+1} - \psi_{i-1,j+1}) + \zeta_{i,j-1}(\psi_{i+1,j-1} - \psi_{i-1,j-1})\,] \,. \tag{48}$$

These finite-difference Jacobians use the grid points shown in Fig. 2 and correspond to the following expressions of the differential Jacobian, respectively:

$$J(\zeta,\psi) = \frac{\partial\zeta}{\partial x}\frac{\partial\psi}{\partial y} - \frac{\partial\zeta}{\partial y}\frac{\partial\psi}{\partial x} = -\frac{\partial}{\partial x}(\frac{\partial\zeta}{\partial y}\psi) + \frac{\partial}{\partial y}(\frac{\partial\zeta}{\partial x}\psi) = \frac{\partial}{\partial x}(\zeta\frac{\partial\psi}{\partial y}) - \frac{\partial}{\partial y}(\zeta\frac{\partial\psi}{\partial x}).$$

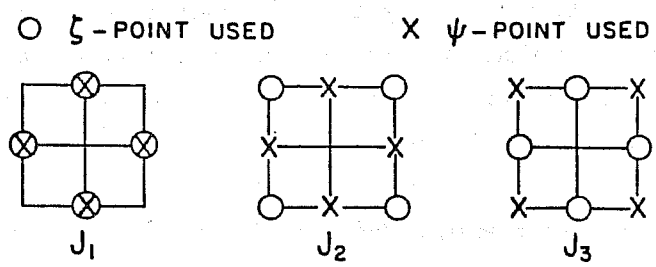

Fig. 2 Grid points used for J_1, J_2 and J_3

The finite difference Jacobian J_1 given above is that used in (8). It can be shown that all of J_1, J_2 and J_3 have second-order accuracy and satisfy a discrete analog of the integral constraint (20).

More general second-order finite-difference Jacobians can be obtained by linear combinations of the above three; thus we let

$$J_{i,j}(\zeta,\psi) = \alpha(J_1)_{i,j} + \gamma(J_2)_{i,j} + \beta(J_3)_{i,j} \quad , \tag{49}$$

where $\alpha + \gamma + \beta = 1$. Identifying the coefficients $a_{\underline{i,i+i}}$, for this finite-difference Jacobian, we find $\alpha = \beta$ is required to satisfy (37). Thus $\alpha(J_1+J_3) + \gamma J_2$, where $2\alpha + \gamma = 1$, is enstrophy conserving. An example is J_2 itself. Similarly, we find $\alpha = \gamma$ is required to satisfy (40). Thus $\alpha(J_1+J_2) + \beta J_3$, where $2\alpha + \beta = 1$, is energy-conserving. An example is J_3 itself.

The finite-difference Jacobian that satisfies both (37) and (40) is then given by $\alpha = \beta = \gamma = \frac{1}{3}$. Thus the finite-difference Jacobian given by

$$J_A = \frac{1}{3}(J_1 + J_2 + J_3) \tag{50}$$

conserves both enstrophy and energy. When (50) is used, the vorticity equation is given by

$$\begin{aligned} d^2\frac{\partial}{\partial t}\zeta_{i,j} = &-\frac{2}{3}\Big[F_{i+\frac{1}{2},j}\frac{\zeta_{i+1,j}+\zeta_{i,j}}{2} - F_{i-\frac{1}{2},j}\frac{\zeta_{i,j}+\zeta_{i-1,j}}{2} \\ &+ G_{i,j+\frac{1}{2}}\frac{\zeta_{i,j+1}+\zeta_{i,j}}{2} - G_{i,j-\frac{1}{2}}\frac{\zeta_{i,j}+\zeta_{i,j-1}}{2}\Big] \\ &-\frac{1}{3}\Big[F'_{i+\frac{1}{2},j+\frac{1}{2}}\frac{\zeta_{i+1,j+1}+\zeta_{i,j}}{2} - F'_{i-\frac{1}{2},j-\frac{1}{2}}\frac{\zeta_{i,j}+\zeta_{i-1,j-1}}{2} \\ &+ G'_{i-\frac{1}{2},j+\frac{1}{2}}\frac{\zeta_{i-1,j+1}+\zeta_{i,j}}{2} - G'_{i+\frac{1}{2},j-\frac{1}{2}}\frac{\zeta_{i,j}+\zeta_{i+1,j-1}}{2}\Big], \end{aligned} \tag{51}$$

where

$$F_{i+\frac{1}{2},j} \equiv \frac{1}{4}(\psi_{i,j-1} + \psi_{i+1,j-1} - \psi_{i,j+1} - \psi_{i+1,j+1}), \tag{52}$$

$$G_{i,j+\frac{1}{2}} \equiv \frac{1}{4}(\psi_{i+1,j} + \psi_{i+1,j+1} - \psi_{i-1,j} - \psi_{i-1,j+1}), \tag{53}$$

$$F'_{i+\frac{1}{2},j+\frac{1}{2}} \equiv \frac{1}{2}(\psi_{i+1,j} - \psi_{i,j+1}), \tag{54}$$

$$G'_{i-\frac{1}{2},j+\frac{1}{2}} \equiv \frac{1}{2}(\psi_{i,j+1} - \psi_{i-1,j}). \tag{55}$$

When divided by d^2, (51) corresponds to the flux convergence form of the vorticity equation, (3). The terms involving F, G, F' and G' correspond to the vorticity fluxes in the x, y, x' and y' directions shown in Fig. 3, respectively.

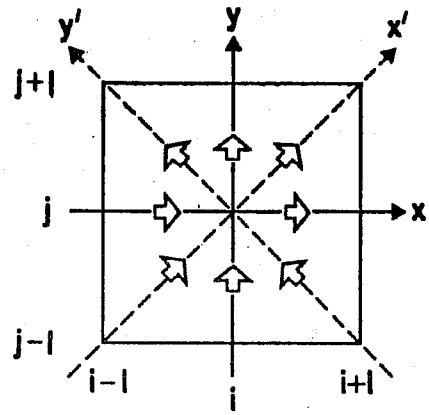

Fig. 3 Fluxes appearing in the finite-difference vorticity equation (51).

From the definitions of F, G. F' and G' given above, we see that

$$F_{i+\frac{1}{2},j} - F_{i-\frac{1}{2},j} + G_{i,j+\frac{1}{2}} - G_{i,j-\frac{1}{2}} = 0, \tag{56}$$

$$F'_{i+\frac{1}{2},j+\frac{1}{2}} - F'_{i-\frac{1}{2},j-\frac{1}{2}} + G'_{i-\frac{1}{2},j+\frac{1}{2}} - G'_{i+\frac{1}{2},j-\frac{1}{2}} = 0 . \tag{57}$$

When divided by d^2, (56) and (57) correspond to the continuity equation (1). Using these in the right hand side of (50), we find that the sum of all the coefficients on $\zeta_{i,j}$ vanishes; and thus $\zeta_{i,j}$ can be replaced by $-\zeta_{i,j}$. When this is done, we obtain a form that corresponds to the advection form of the vorticity equation, (2).

Multiplying (51) by $\zeta_{i,j}$ and using (56) and (57), we obtain

$$\begin{aligned} d^2\frac{\partial}{\partial t}\frac{1}{2}\zeta^2_{i,j} = &-\frac{1}{3}\left[F_{i+\frac{1}{2},j}\zeta_{i,j}\zeta_{i+1,j} - F_{i-\frac{1}{2},j}\zeta_{i-1,j}\zeta_{i,j} \right. \\ &\left. + G_{i,j+\frac{1}{2}}\zeta_{i,j}\zeta_{i,j+1} - G_{i,j-\frac{1}{2}}\zeta_{i,j-1}\zeta_{i,j} \right] \\ &-\frac{1}{6}\left[F'_{i+\frac{1}{2},j+\frac{1}{2}}\zeta_{i,j}\zeta_{i+1,j+1} - F'_{i-\frac{1}{2},j-\frac{1}{2}}\zeta_{i-1,j-1}\zeta_{i,j} \right. \\ &\left. + G'_{i-\frac{1}{2},j+\frac{1}{2}}\zeta_{i,j}\zeta_{i-1,j+1} - G'_{i+\frac{1}{2},j-\frac{1}{2}}\zeta_{i+1,j-1}\zeta_{i,j} \right], \end{aligned} \tag{58}$$

When divided by d^2, this corresponds to the flux convergence form for $\partial\frac{1}{2}\zeta^2/\partial t$ of the continuous case,

$$\frac{\partial}{\partial t}\frac{1}{2}\zeta^2 = -\nabla\cdot(\mathbf{v}\frac{1}{2}\zeta^2), \tag{59}$$

which can be obtained from (1) and (2). As this example for a square grid shows, the product $\zeta_{\underline{i}}J_{\underline{i}}(\zeta,\psi)$ can be written in a flux convergence form when the requirement (37) is satisfied. In such a form, however, as in (58), each of the fluxes of $\frac{1}{2}\zeta^2$ is based on a product of ζ which is not necessarily positive. This is anticipated since (51) reduces to

$$\frac{\partial\zeta_i}{\partial t} = -U\frac{\zeta_{i+1} - \zeta_{i-1}}{2d} \tag{60}$$

when it is linearized with respect to a small y-independent perturbation on a uniform flow in the x-direction, U. Multiplying (60) by ζ_i, we obtain

$$\frac{\partial}{\partial t}\frac{1}{2}\zeta_i^2 = -\frac{1}{d}(U\frac{1}{2}\zeta_i\zeta_{i+1} - U\frac{1}{2}\zeta_{i-1}\zeta_i), \tag{61}$$

which is the flux convergence form corresponding to (59) for this linearized case. The products $\zeta_i\zeta_{i+1}$ and $\zeta_{i-1}\zeta_i$ are negative when averaged in time for waves whose wavelengths are shorter than 4d. This is consistent with the fact that the group velocity obtained from (60) is against the direction of U for such short waves.

As the above example shows, the finite-difference Jacobian J_A is a generalization of the right hand side of (60) to the two-dimensional nonlinear case; and therefore it does not eliminate (or decrease) any deficiencies that scheme (60) might have, such as the computational dispersion of short waves. It should be noted that, however, there are a variety of ways in generalizing (60) to the nonlinear case. Use of J_A, or its fourth-order version (see Arakawa, 1966), eliminates computational instability that may occur in a nonlinear system (or in a system in which the basic current varies in space). Moreover, it prevents a false systematic computational cascade of energy into small-scale motions; and because there is then relatively little energy in the small-scale motions, where errors are large, the overall error is small. In this way, other statistical properties of the solution, such as conservation of the higher moments of the statistical distribution of vorticity, are approximately maintained.

If the energy in the shortest scale is the result of a spurious computational energy cascade, a decrease of the grid size does not help insofar as the long-term simulation of nonviscous flow is concerned (Arakawa and Lamb, 1977). Such a result is completely different from that which might be expected from the usual analysis of truncation error, which is a measure of the formal difference of the finite-difference equation from the original differential equation. The paradox occurs because a decrease of the grid size allows a further computational cascade of energy into the

added part of the spectral domain. After a sufficient period of integration, the cascading energy will again reach and accumulate in the shortest resolvable scale. The overall error will become large again and the prediction of some of the statistical properties will become even worse than with the coarser grid.

The existence of lateral viscosity can make a false computational cascade of energy less harmful. Since such viscosity is more effective for smaller scales, however, a spurious computational energy cascade into these scales falsely enhances the total amount of energy dissipation.

The spectral forms of the finite-difference Jacobians J_1, J_2, J_3 and J_A were derived and investigated by Lilly (1965). He pointed out that when J_A is used for the two-component system given by (10) and (11), we simply obtain $dS/dt = dC/dt = dU/dt = 0$, so that the aliasing error has no influence for this system.

It is interesting that when a regular triangular-hexagonal grid is used, the simplest second-order finite-difference Jacobian satisfies the requirements (37) and (40) (Sadourny et al., 1968; Williamson, 1968; Masuda, 1969). Jespersen (1974) showed that J_A and its counterpart on a regular triangular-hexagonal grid can be derived with a finite-element method.

6. EXAMPLES OF NUMERICAL TIME INTEGRATION WITH DIFFERENT FINITE-DIFFERENCE JACOBIANS

In the last section we let time t be continuous. In actual numerical integrations, time must also be discretized.

When the trapezoidal implicit scheme is used as the time differencing, with J_A for a square grid, the vorticity equation (6) becomes

$$\frac{\zeta^{n+1} - \zeta^n}{\Delta t} = J_A\left(\frac{\zeta^{n+1} + \zeta^n}{2}, \frac{\psi^{n+1} + \psi^n}{2}\right), \tag{62}$$

where the superscript denotes time level and Δt is the time interval. The index for space grid has been omitted. Multiplying (62) by $\frac{1}{2}(\zeta^{n+1} + \zeta^n)$ and using the property of J_A, we obtain

$$\overline{\frac{1}{2}(\zeta^{n+1})^2} = \overline{\frac{1}{2}(\zeta^n)^2}\ , \tag{63}$$

where the overbar denotes the mean over all grid points. This means that enstrophy is conserved in the time-discrete case also.

When the more common leapfrog scheme is used, we have

$$\frac{\zeta^{n+1} - \zeta^{n-1}}{2\Delta t} = J_A(\zeta^n, \psi^n)\ . \tag{64}$$

Multiplying (64) by ζ^n and again using the property of J_A, we obtain

$$\overline{\frac{1}{2}\zeta^n\zeta^{n+1}} = \overline{\frac{1}{2}\zeta^{n-1}\zeta^n} \tag{65}$$

instead of (63). If the solution is "smooth" in the sense of

$$\zeta^n \fallingdotseq \frac{1}{2}(\zeta^{n-1} + \zeta^{n+1}), \tag{66}$$

(65) gives

$$\overline{\frac{1}{2}(\zeta^{n+1})^2} \fallingdotseq \overline{\frac{1}{2}(\zeta^{n-1})^2} \tag{67}$$

so that the enstrophy at every other time level remains approximately constant.

It is well-known, however, that there are two sets of solutions with the leapfrog scheme: one "physical" mode and the other computational mode (see, for example, Mesinger and Arakawa, 1976). With a sufficiently small Δt, the computational mode is characterized by an oscillation from level to level in time with the period of $2\Delta t$. It is also well-known that the computational mode is unconditionally unstable for a decay equation (see also Mesinger and Arakawa, 1976).

In a nonlinear system there are wave-wave interactions and, even when the entire system does not include any dissipative effect, a wave can decay by transfering its energy to another wave. Presumably due to this decaying effect acting on each wave component, numerical integrations of a nonlinear system with the leapfrog scheme usually show gradual amplifications of the computational mode even when the system is stable in the time-continuous case. (See, for example, Lilly, 1965). A common way to overcome this problem is to periodically insert a time-differencing scheme that is free from a computational mode, or to introduce a time-smoothing operator.

Numerical tests have been made with the finite-difference Jacobians presented in the last section. In these tests, the initial condition was given by

$$\psi = \Psi \sin(\pi i/8)[\cos(\pi j/8) + 0.1 \cos(\pi j/4)] \tag{68}$$

and Δt was chosen such that $\Delta t/d^2 = 0.7$. The leapfrog scheme was used for most time steps. In order to eliminate the gradual separation of the solutions at even and odd time steps due to a growth of the computational mode, a two-level scheme was inserted every 240 time steps. The simplest five-point Laplacian given by (9) was used. Fig. 4 shows the time change of enstrophy and energy obtained with these Jacobians. The expected conservation properties are observed even though the trapezoidal implicit scheme was not used. The energy-conserving J_3 shows considerable increase of enstrophy. On the other hand, the enstrophy-conserving J_2 approximately conserves energy in spite of the lack of a formal guarantee. This is reasonable because ens-

trophy is more sensitive to shorter waves for which errors are large. J_A conserves both quantities, with only negligible errors arising from the leapfrog scheme.

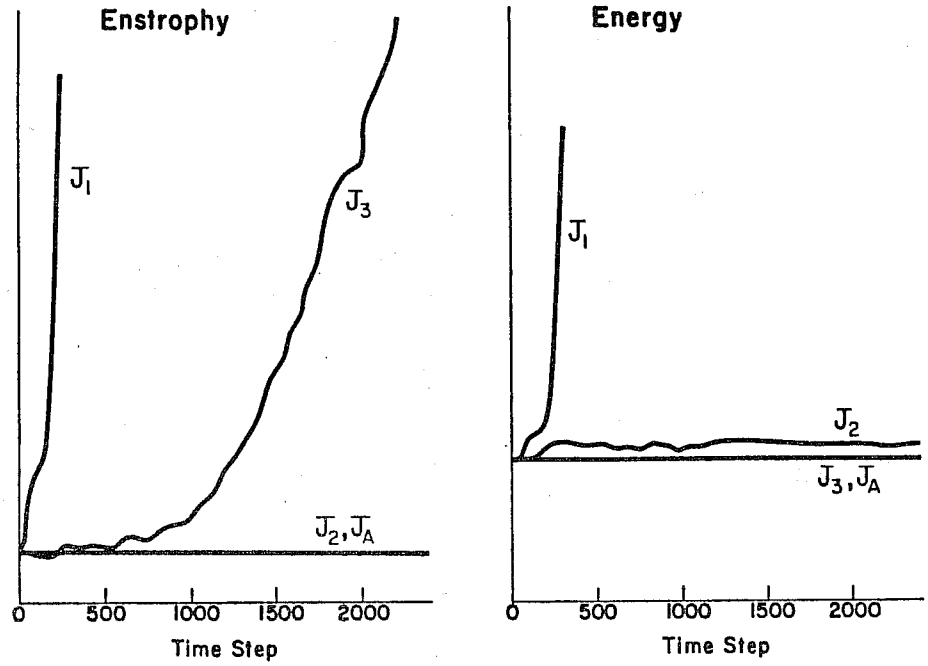

Fig. 4 Comparison of the time variation of enstrophy and energy during a numerical integration with the finite-difference Jacobians J_1, J_2, J_3 and J_A.

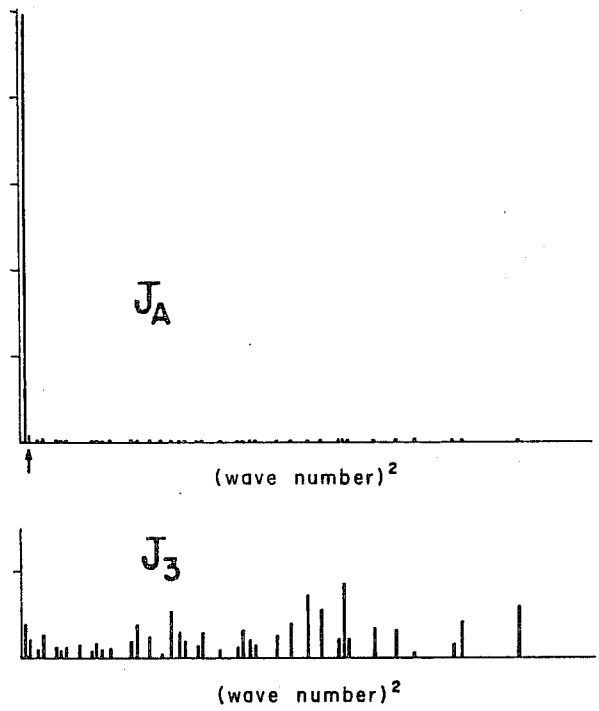

Fig. 5 Comparison of the spectral distribution of energy obtained with J_3 and J_A after a numerical integration of 2400 time steps.[3] The arrow shows the wavenumber that contained most of the energy at the initial time.

Fig. 5 shows the spectral distribution of kinetic energy obtained by J_A and by J_3 at the end of the calculations. The small arrow shows the wavenumber for $\sin(\pi i/8)\cos(\pi j/8)$, which contained almost all of the energy at the initial time. Although energy was approximately conserved with J_3 there was a considerable spurious energy cascade into the high wavenumbers, whereas with J_A more energy went into a lower wavenumber than into the higher wavenumbers, in agreement with the conservation of the average wavenumber as given by (30).

7. COMPUTATIONAL BOUNDARY CONDITIONS

In this section we consider the problem of integrating the vorticity equation (6) for a limited domain. The boundary of the domain can be either a rigid wall or an artificial open boundary within the fluid. To solve for $\partial\psi/\partial t$ from known $\partial\zeta/\partial t(=\nabla^2\partial\psi/\partial t)$, we need a boundary condition for $\partial\psi/\partial t$. When discrete grid points are used, we need an additional boundary condition, which may be called computational boundary condition, as shown below.

Let the heavy line in Fig. 6, $i = I$, be the boundary of the computational domain. The point (I,j) shown by the open circle in the figure is an outflow point, a rigid wall point, and an inflow point, depending on $\psi_{I,j-1} - \psi_{I,j+1} \gtrless 0$. We assume that the finite-difference vorticity equation (51) is used for all inner grid points, $i \leq I-1$. By substituting $i = I-1$ into (51), we see that ζ at $i = I$ must be known, or must be somehow determined, in order to compute $\partial\zeta_{I-1,j}/\partial t$.

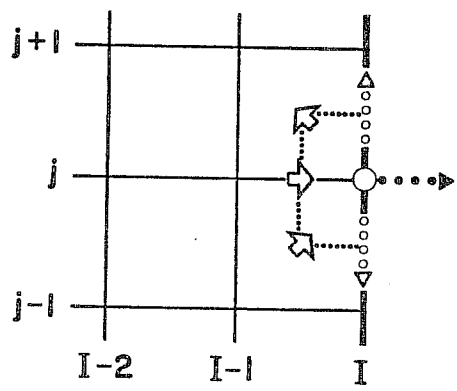

Fig. 6 Fluxes from or toward a grid point at the boundary of the computational domain. See text for explanation.

Here we follow an approach in which $\zeta_{I,j}$ is prognostically determined. We first let time t be continuous and consider the vorticity budget for the area represented by the point (I,j), the area surrounded by the boundary and the dotted lines in the figure. Taking (51) applied to $i = I-1$ into consideration, the vorticity budget equation may be written as

$$\frac{d^2}{2}\frac{\partial}{\partial t}\zeta_{I,j} = \frac{2}{3}F_{I-\frac{1}{2},j}\frac{\zeta_{I,j}+\zeta_{I-1,j}}{2}$$

$$+\frac{1}{3}\left[F'_{I-\frac{1}{2},j-\frac{1}{2}}\frac{\zeta_{I,j}+\zeta_{I-1,j-1}}{2} - G'_{I-\frac{1}{2},j+\frac{1}{2}}\frac{\zeta_{I-1,j+1}+\zeta_{I,j}}{2}\right]$$

$$-\hat{G}_{j+\frac{1}{2}}\frac{\zeta_{I,j+1}+\zeta_{I,j}}{2} + \hat{G}_{j-\frac{1}{2}}\frac{\zeta_{I,j}+\zeta_{I,j-1}}{2} - \hat{F}_j\frac{\hat{\zeta}_j+\zeta_{I,j}}{2}\,. \qquad (69)$$

The factor $d^2/2$ appears on the left hand side since it is reasonable to assume that the boundary points represent one half of the area represented by the inner points. The terms involving F, F' and G' are the vorticity fluxes from or toward points (I-1,j), (I-1,j-1) and (I-1,j+1), respectively, as shown by the double-stroke arrows in the figure. The terms involving $\hat{G}$, the form of which is yet to be specified, are the vorticity fluxes along the boundary as shown by the open dot arrows. The term involving $\hat{F}$, the form of which is also yet to be specified, is the vorticity flux across the boundary as shown by the solid dot arrow. This flux also depends on an unspecified hypothetical vorticity $\hat{\zeta}$.

The forms of F, F' and G' are given by (52), (54) and (55). For $\hat{G}$ we choose the following uncentered form:

$$\hat{G}_{j+\frac{1}{2}} = \alpha\,(\psi_{I,j} + \psi_{I,j+1} - \psi_{I-1,j} - \psi_{I-1,j+1})\,, \qquad (70)$$

where α is an unspecified positive constant. For $\hat{F}$ we choose

$$\hat{F}_j = \beta\,(\psi_{I,j-1} - \psi_{I,j+1}), \qquad (71)$$

where β is another unspecified positive constant, so that $\hat{F}_j \gtrless 0$ as $\psi_{I,j-1} - \psi_{I,j+1} \gtrless 0$.

Since the right hand side of (69) should vanish when ζ is constant, we require

$$\frac{2}{3}F_{I-\frac{1}{2},j} + \frac{1}{3}\left[F'_{I-\frac{1}{2},j-\frac{1}{2}} - G'_{I-\frac{1}{2},j+\frac{1}{2}}\right] - \hat{G}_{j+\frac{1}{2}} + \hat{G}_{j-\frac{1}{2}} - \hat{F}_j = 0\,. \qquad (72)$$

Substituting (52), (54), (55), (70) and (71), we find that (72) is identically satisfied if and only if

$$\alpha = \frac{1}{6}\,, \quad \beta = \frac{1}{2}\,. \qquad (73)$$

Let us now consider the enstrophy budget equation. Multiplying (69) by $\zeta_{I,j}$ and using (72), we obtain

$$\frac{d^2}{2}\frac{\partial}{\partial t}\frac{1}{2}\zeta^2_{I,j} = \frac{1}{3}F_{I-\frac{1}{2},j}\zeta_{I,j}\zeta_{I-1,j}$$

$$+\frac{1}{6}\left[F'_{I-\frac{1}{2},j-\frac{1}{2}}\zeta_{I,j}\zeta_{I-1,j-1} - G'_{I-\frac{1}{2},j+\frac{1}{2}}\zeta_{I,j}\zeta_{I-1,j+1}\right]$$

$$-\frac{1}{2}\hat{G}_{j+\frac{1}{2}}\zeta_{I,j}\zeta_{I,j+1} + \frac{1}{2}\hat{G}_{j-\frac{1}{2}}\zeta_{I,j}\zeta_{I,j-1} - \frac{1}{2}\hat{F}_j\hat{\zeta}_{I,j}\hat{\zeta}_j\,. \qquad (74)$$

The first three terms on the right hand side of (74) cancel when the sum of (74) and (58) applied to the inner points is taken. The two terms involving $\hat{G}$ cancel when the sum of (74) over j along the boundary is taken. The net effect on the entire domain is, therefore, through the term involving $\hat{F}$, which represents the enstrophy flow across the boundary.

We must now choose $\hat{\zeta}$ properly.

(a) Rigid boundary ($\hat{F}_j = 0$)
In this case $\hat{\zeta}$ in (64) has no effect since it is multiplied by $\hat{F}_j$.

(b) Outflow boundary ($\hat{F}_j > 0$)
In this case $\hat{\zeta}_j = \zeta_{I,j}$ seems to be the best choice since $\hat{F}_j\hat{\zeta}_j\zeta_{I,j} = \hat{F}_j\zeta^2_{I,j} > 0$ and, therefore, the enstrophy flux across the boundary is in fact outward.

(c) Inflow boundary ($\hat{F}_j < 0$)

In this case we have more freedom. We may wish to prescribe $\zeta_{I,j}$ and ignore the prognostic equation (69). Or we may wish no inflow of vorticity, with $\hat{\zeta}_j = -\zeta_{I,j}$ in (69). With this choice the last term of (74) has a damping effect on $\frac{1}{2}\zeta^2_{I,j}$ until $\zeta_{I,j}$ itself becomes zero. Or we may wish no inflow of enstrophy with $\hat{\zeta}_j = 0$.

To compare with the results of other authors (e.g. Nitta, 1962; Matsuno, 1966; Gustafsson et al., 1972; Elvis and Sundstrom, 1973) on outflow boundary conditions and obtain a guide in time-differencing, let us apply (51) and (69) to the case of the one-dimensional advection equation with constant current. This can be done by formally making ψ constant in i, ζ constant in j, and $(\psi_{j-1} - \psi_j)/d = (\psi_j - \psi_{j+1})/d = U$ for all j. Then (51) with (52), (53), (54) and (55) for inner points becomes (69), i.e.,

$$\frac{\partial \zeta_i}{\partial t} = -\frac{U}{2d}(\zeta_{i+1} - \zeta_{i-1}), \quad i = \cdots, I-2, I-1; \tag{75}$$

while (69) with (52), (54), (55), (70), (71) and (73) becomes

$$\frac{\partial \zeta_I}{\partial t} = -\frac{U}{d}(\hat{\zeta} - \zeta_{I-1}). \tag{76}$$

Now let U be positive so that i = I is an outflow point as shown in Fig. 7. With the choice $\hat{\zeta} = \zeta_I$ (see the argument in (b) above), (76) becomes

$$\frac{\partial \zeta_I}{\partial t} = -\frac{U}{d}(\zeta_I - \zeta_{I-1}). \tag{77}$$

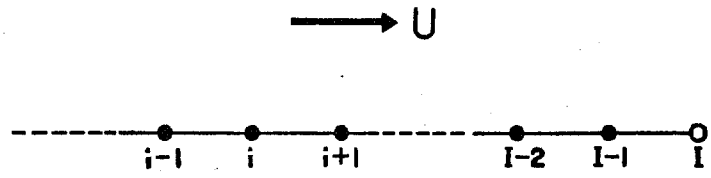

Fig. 7 A one-dimensional grid, showing an outflow point i = I by the open circle.

Eq. (77) is based on uncentered upstream space-differencing. For the case of constant U being considered now, it is equivalent to

$$\frac{\partial \zeta_I}{\partial t} = -\frac{U}{2d}(\zeta_{I+1} - \zeta_{I-1}) \tag{78}$$

with ζ_{I+1} given by the extrapolation

$$\zeta_{I+1} = 2\zeta_I - \zeta_{I-1}. \tag{79}$$

Matsuno (1966) discussed false reflection of waves at an outflow boundary with boundary conditions based on extrapolation. Nitta (1962) examined various boundary conditions through actual numerical integrations. Their results show that (78) with (79), or equivalently (77), is a reasonable choice as far as space differencing is concerned. We must then find a suitable time differencing for (77).

Suppose that we are using the leapfrog scheme for (75). Then

$$\frac{\zeta_i^{n+1} - \zeta_i^{n-1}}{2\Delta t} = -\frac{U}{2d}(\zeta_{i+1}^n - \zeta_{i-1}^n), \quad i = \cdots, I-2, I-1. \tag{80}$$

We may formally write a time-discrete version of (76) as

$$\frac{\zeta_I^{n+1} - \zeta_I^{n-1}}{2\Delta t} = -\frac{U}{d}(\hat{\zeta}^n - \zeta_{I-1}^n). \tag{81}$$

If we replace $\hat{\zeta}^n$ in (81) simply by ζ_I^n, we obtain the leapfrog scheme for (77),

$$\frac{\zeta_I^{n+1} - \zeta_I^{n-1}}{2\Delta t} = -\frac{U}{d}(\zeta_I^n - \zeta_{I-1}^n). \tag{82}$$

But from the argument given in (b) above, it is clear that (75) and (77) form a damping system. The damping effect is due to the leading term in (77). This means that when the leapfrog scheme (82) is used, the solution is unstable.

We must therefore seek a suitable alternative. To this end, we temporarily write a scheme for (78) as

$$\frac{\zeta_I^{n+1} - \zeta_I^{n-1}}{2\Delta t} = -\frac{U}{2d}\left(\zeta_{I+1}^{n} - \zeta_{I-1}^{n}\right) . \tag{83}$$

Eqs. (81) and (83) are equivalent when

$$\hat{\zeta}^n = \frac{1}{2}\left(\zeta_{I-1}^{n} + \zeta_{I+1}^{n}\right) . \tag{84}$$

If we use the "rhomboidal" extrapolation and "inward-backward" extrapolation for ζ_{I+1}^{n} (Gustafsson et al., 1972; Elvis and Sundström, 1973; see also Sundström and Elvis, 1977) given by

$$\zeta_{I+1}^{n} = \zeta_I^{n+1} + \zeta_I^{n-1} - \zeta_{I-1}^{n} \tag{85}$$

and

$$\zeta_{I+1}^{n} = 2\zeta_I^{n-1} - \zeta_{I-1}^{n-2} , \tag{86}$$

we obtain from (84)

$$\hat{\zeta}^n = \frac{1}{2}\left(\zeta_I^{n+1} + \zeta_I^{n-1}\right) \tag{87}$$

and

$$\hat{\zeta}^n = \zeta_I^{n-1} + \frac{1}{2}\left(\zeta_{I-1}^{n} - \zeta_{I-1}^{n-2}\right) , \tag{88}$$

respectively. One of these expressions, with the subscript j added, should be a good choice of $\hat{\zeta}$ for use in (69).

We must also be careful in time-differencing of (69) at inflow point if $\hat{\zeta}$ is chosen in such a way that it adds to the entire system a damping effect, as in the case of $\hat{\zeta}_j = -\zeta_{I,j}$ (see the argument in (c) above).

8. FINITE-DIFFERENCE SCHEMES FOR THE GENERAL TWO-DIMENSIONAL ADVECTION EQUATION

In this section we consider the general two-dimensional advection equation

$$\frac{\partial q}{\partial t} = -\mathbf{v}\cdot\nabla q, \tag{89}$$

where q is an arbitrary variable and v is the advective velocity, which is not necessarily nondivergent. The two-dimensional continuity equation may be written as

$$\frac{\partial m}{\partial t} = -\nabla\cdot(m\mathbf{v}), \tag{90}$$

where m is the mass of the layer bounded by two material surfaces, per unit horizontal area, as shown in Fig. 8. Combining (89) with (90), we obtain the

flux convergence form for q given by

$$\frac{\partial}{\partial t}(mq) = - \nabla\cdot(mvq). \tag{91}$$

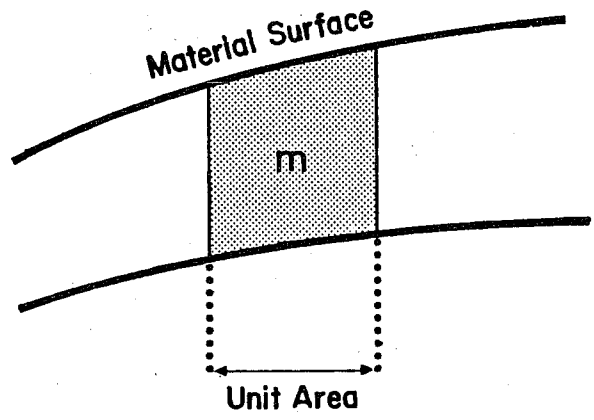

Fig. 8 A material layer whose mass per unit horizontal area is m.

Let F(q) be an arbitrary function of q and $F'(q) \equiv dF(q)/dq$. Multiplying (89) by F'(q), we obtain

$$\frac{\partial F}{\partial t} = - \mathbf{v}\cdot\nabla F. \tag{92}$$

Combining (92) with (90), we obtain the flux convergence form for F(q),

$$\frac{\partial}{\partial t}(mF) = - \nabla\cdot(m\mathbf{v}F). \tag{93}$$

Let an overbar denote the area mean over a domain along the boundary of which the normal component of v is zero, or over a periodic domain. Then, from (90), (91) and (93), we have

$$\frac{\partial \overline{m}}{\partial t} = 0 , \tag{94}$$

$$\frac{\partial \overline{mq}}{\partial t} = 0 , \tag{95}$$

$$\frac{\partial \overline{mF}}{\partial t} = 0 . \tag{96}$$

When multiplied by the total area, (94) represents conservation of the total mass, while (95) and (96) represent conservation of the mass-weighted integrals of q and F(q), respectively.

We are now interested in constructing finite-difference schemes in which discrete analogs of $\overline{m}$ and $\overline{mq}$ are conserved and in which a discrete analog of $\overline{mF}$ is either conserved or bounded by a selected F(q) other than q itself.

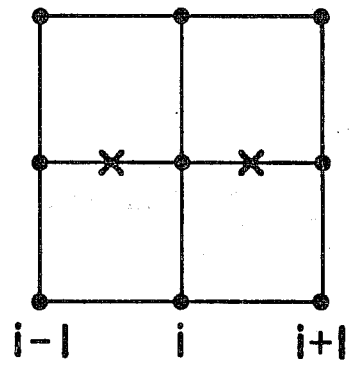

Fig. 9 See text for explanation.

Consider a square grid as shown by Fig. 9. We assume that the dot points carry both m and q. The velocity components may be carried at other points, as in the case of a staggered or a semi-staggered grid. In any case, we may write a finite-difference approximation to (90) as

$$\frac{\partial m_i}{\partial t} = -\frac{1}{d}\left[(mu)_{i+\frac{1}{2}} - (mu)_{i-\frac{1}{2}} \right], \tag{97}$$

Here only the mass flux convergence in the x-direction has been written and the index j has been omitted for simplicity. The mass flux (mu) is defined at the cross points in the figure; and its actual expression generally involves space-averaging of m and, depending on the staggering, space-averaging of u as well. Regardless of its form, however, successive cancellations of (mu) will take place when the sum of (97) over all grid points is taken, leading to conservation of the total mass.

Similarly, conservation of a discrete analog of $\overline{mq}$ can be achieved by writing a finite-difference approximation to (91) as

$$\frac{\partial}{\partial t}(m_i q_i) = -\frac{1}{d}\left[(muq)_{i+\frac{1}{2}} - (muq)_{i-\frac{1}{2}} \right], \tag{98}$$

regardless of the actual expression for (muq).

From (97) and (98), we obtain

$$m_i \frac{\partial q_i}{\partial t} = -\frac{1}{d}\left[\left\{ (muq)_{i+\frac{1}{2}} - (mu)_{i+\frac{1}{2}} q_i \right\} + \left\{ (mu)_{i-\frac{1}{2}} q_i - (muq)_{i-\frac{1}{2}} \right\} \right]. \tag{99}$$

When divided by m_i, (99) is a finite difference approximation to the advection equation (89).

To obtain the corresponding finite-difference approximation to (93), we first multiply (99) by $F'(q_i)$. Then

$$m_i \frac{\partial F_i}{\partial t} = -\frac{1}{d}\Big[\{ (muq)_{i+\frac{1}{2}} - (mu)_{i+\frac{1}{2}} q_i \} F'_i + \{ (mu)_{i-\frac{1}{2}} q_i - (muq)_{i-\frac{1}{2}} \} F'_i \Big] , \tag{100}$$

where $F_i \equiv F(q_i)$ and $F'_i \equiv F'(q_i)$. Adding $F_i \times$(97) to (100) we obtain

$$\frac{\partial}{\partial t}(m_i F_i) = -\frac{1}{d}\Big[\{ (muq)_{i+\frac{1}{2}} F'_i - (mu)_{i+\frac{1}{2}} (qF' - F)_i - \{ (muq)_{i-\frac{1}{2}} F'_i - (mu)_{i-\frac{1}{2}} (qF' - F)_i \} \Big] . \tag{101}$$

We are now concerned with the sum of (101) over all grid points. When the contribution from the boundary condition can be omitted, as in the case of the periodic domain, we obtain

$$\frac{\partial}{\partial t} \sum_i (m_i F_i) = -\frac{1}{d} \sum_i \Big[- (muq)_{i+\frac{1}{2}} (F'_{i+1} - F'_i) + (mu)_{i+\frac{1}{2}} \{ (qF' - F)_{i+1} - (qF' - F)_i \} \Big] \tag{102}$$

In deriving (102) the index i under the sum has been adjusted in such a way that only (muq) and (mu) at the point $i+\frac{1}{2}$ appear. Let us define $\hat{q}_{i+\frac{1}{2}}$ by

$$(muq)_{i+\frac{1}{2}} = (mu)_{i+\frac{1}{2}} \hat{q}_{i+\frac{1}{2}} . \tag{103}$$

Then (102) may be rewritten as

$$\frac{\partial}{\partial t} \sum_i (m_i F_i) = \frac{1}{d} \sum_i \Big[(mu)_{i+\frac{1}{2}} \{ \hat{q}_{i+\frac{1}{2}} (F'_{i+1} - F'_i) - (qF' - F)_{i+1} + (qF' - F)_i \} \Big] . \tag{104}$$

If we require that $\sum_i (m_i F_i)$ be conserved for any $(mu)_{i+\frac{1}{2}}$, the following form for $\hat{q}$ must be used:

$$\hat{q}_{i+\frac{1}{2}} = \frac{(qF' - F)_{i+1} - (qF' - F)_i}{F'_{i+1} - F'_i} \quad . \tag{105}$$

This is a finite-difference approximation to the identity

$$q \equiv \frac{d(qF' - F)}{dF'} \quad . \tag{106}$$

Some examples for the choice of F(q) and the corresponding $\hat{q}$ are given below.

$$\text{(a)} \quad F = q^2: \qquad \hat{q}_{i+\frac{1}{2}} = \frac{q_i + q_{i+1}}{2} \tag{107}$$

$$\text{(b)} \quad F = \sqrt{q}: \qquad \hat{q}_{i+\frac{1}{2}} = \sqrt{q_i q_{i+1}} \tag{108}$$

$$\text{(c)} \quad F = \frac{1}{q}: \qquad \hat{q}_{i+\frac{1}{2}} = \frac{2q_i q_{i+1}}{q_i + q_{i+1}} \tag{109}$$

$$\text{(d)} \quad F = \ell nq \qquad \hat{q}_{i+\frac{1}{2}} = \frac{\ell nq_{i+1} - \ell nq_i}{1/q_i - 1/q_{i+1}} \tag{110}$$

Note that $q_{i+\frac{1}{2}}$ given by (107), (108) and (109) are the arithmetic, geometric and harmonic means of q_i and q_{i+1}. Obviously, the choices (b), (c) and (d) can be made only when q is a non-zero variable. The actual choice for F(q) should be guided by the physical meaning of q and by judging whether conservation of a discrete analog of $\overline{mF}$ is an effective computational constraint or not.

In the case of the vorticity equation for nondivergent flow discussed in earlier sections we have m = constant and $q = \zeta$. For this case (a) is the most reasonable choice because conservation of a discrete analog of the enstrophy then follows.

The choice (a), however, does not necessarily give an effective computational constraint for other variables. For example, when q is the mixing ratio of an atmospheric constituent, such as that of water vapor, conservation of a discrete analog of $\overline{mq^2}$ generally does not prevent generation of negative values for q. The choice (c) has some benefit in such a case since conservation of a discrete analog of $\overline{m(1/q)}$ prevents generation of $q = 0$ at any grid point if $q \neq 0$ initially at all grid points. However, this choice has difficulty at those points where $q = 0$ initially or $q = 0$ is generated by an added sink term.

A promising approach in such a case as above is to abandon exact conservation, except for discrete analogs of $\overline{m}$ and $\overline{mq}$, but to require that discrete analogs of $\overline{mF}$ be bounded for more than one selected functional forms of F(q). To illustrate the approach, let us consider two functions of q, F(q) and G(q),

that are positive definite for $q > 0$, and require that

$$\frac{\partial}{\partial t}\sum_i (m_i F_i) \leq 0 \tag{111}$$

and

$$\frac{\partial}{\partial t}\sum_i (m_i G_i) \leq 0. \tag{112}$$

From (104) we see that(111) and (112) are satisfied if we choose $\hat{q}$ for which all of the following conditions are met:

$$\hat{q}_{i+\frac{1}{2}} \leq \frac{(qF' - F)_{i+1} - (qF' - F)_i}{F'_{i+1} - F'_i} \quad \text{when } (mu)_{i+\frac{1}{2}}(F'_{i+1} - F'_i) \geq 0, \tag{113}$$

$$\hat{q}_{i+\frac{1}{2}} \geq \frac{(qF' - F)_{i+1} - (qF' - F)_i}{F'_{i+1} - F'_i} \quad \text{when } (mu)_{i+\frac{1}{2}}(F'_{i+1} - F'_i) \leq 0, \tag{114}$$

and the same conditions as (113) and (114) but based on G.

An interesting possibility is the choice $F(q) = q^2$ and $G(q) = 1/q$. Then the above conditions become

$$\left.\begin{aligned} \hat{q}_{i+\frac{1}{2}} &\leq \frac{q_i + q_{i+1}}{2} \\ \hat{q}_{i+\frac{1}{2}} &\leq \frac{2q_i q_{i+1}}{q_i + q_{i+1}} \end{aligned}\right\} \quad \text{when } (mu)_{i+\frac{1}{2}}(q_{i+1} - q_i) \geq 0 \tag{115}$$

and

$$\left.\begin{aligned} \hat{q}_{i+\frac{1}{2}} &\geq \frac{q_i + q_{i+1}}{2} \\ \hat{q}_{i+\frac{1}{2}} &\geq \frac{2q_i q_{i+1}}{q_i + q_{i+1}} \end{aligned}\right\} \quad \text{when } (mu)_{i+\frac{1}{2}}(q_{i+1} - q_i) \leq 0 . \tag{116}$$

Since we have been assuming $q > 0$, the arithmetic mean of q_i and q_{i+1} is larger than their harmonic mean. Then (115) and (116) are satisfied with $\hat{q}_{i+\frac{1}{2}}$ such that

$$\hat{q}_{i+\frac{1}{2}} = \frac{2q_i q_{i+1}}{q_i + q_{i+1}} \quad \text{when } (mu)_{i+\frac{1}{2}}(q_{i+1} - q_i) \geq 0 \tag{117}$$

and

$$\hat{q}_{i+\frac{1}{2}} = \frac{q_i + q_{i+1}}{2} \quad \text{when } (mu)_{i+\frac{1}{2}}(q_{i+1} - q_i) \leq 0 . \tag{118}$$

This choice of $\hat{q}$ uses the harmonic mean when the mass flux is from a grid point with smaller q to a grid point with larger q, and the arithmetic mean otherwise. Consequently, when q at a certain grid point approaches zero, the outflow of q from that grid point also approaches zero while the inflow to that grid point

can remain finite. In this way, there is no possibility of generating negative values of q if time truncation errors are negligible. A version of this scheme is being used in the UCLA general circulation model for predicting water vapor and ozone mixing ratios.

References

Arakawa, A., 1966: Computational design for long-term numerical integration of equations of fluid motion: Two dimensional incompressible flow. Part I. J. Comput. Phys., 1, 119-143.

Arakawa, A. and V.R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. Methods in Computational Physics, Vol. 17, General circulation models of the atmosphere, ed. by J. Chang, Academic Press, New York, 173-265.

Elvius, T. and A. Sundstrom, 1973: Computationally efficient schemes and boundary conditions for a fine-mesh model based on the shallow-water equations. Tellus, 25, 132-156.

Fjørtoft, R., 1953: On the changes in the spectral distribution of kinetic energy for two-dimensional, nondivergent flow. Tellus, 5, 225-230.

Gustafsson, B., H.-O. Kreiss, and A. Sundström, 1972: Difference approximations to mixed initial boundary value problems, II. Math. Comp. 26, 649-686.

Jespersen, D.C., 1974: Arakawa's method is a finite element method. J. Comput. Phys., 16, 383-390.

Masuda, Y., 1969: A finite difference scheme by making use of hexagonal mesh-points. Proc. WMO/IUGG Symposium on Numerical Weather Prediction, Tokyo, VII-35-44.

Matsuno, T., 1966b: False reflection of waves at the boundary due to the use of finite differences. J. Meteor. Soc. Japan, II, 44, 145-157.

Mesinger, F. and A. Arakawa, 1976: Numerical methods used in atmospheric models, Vol. I. GARP Publication Series No. 17, 64 pp.

Nitta, T., 1962: The outflow boundary condition in numerical time integration of adjective equations. J. Meteor. Soc. Japan, II, 40, 13-24.

Phillips, N.A., 1959: An example of non-linear computational instability. The Atmosphere and the Sea in Motion. Rossby Memorial Volume, Rockefeller Institute Press, 501-504.

Sadourny, R., A. Arakawa and Y. Mintz, 1968: Integration of the nondivergent barotropic vorticity equation with an icosahedral-hexagonal grid for the sphere. Mon. Wea. Rev., 96, 351-356.

Sundström, A., and T. Elvis, 1979: Computational problems related to limited-area modeling. Numerical methods used in atmospheric models, Vol. II. GARP Publication Series No. 17, 379-416.

Williamson, D.L., 1968: Integration of the barotropic vorticity equation on a spherical geodesic grid. Tellus, 20, 642-653.