

ON NUMERICAL SOLUTIONS OF THE TIME-DEPENDENT EULER EQUATIONS FOR INCOMPRESSIBLE FLOW

J.-H. SAIAC

C.N.A.M. 292 rue Saint Martin, 75003 Paris, France

SUMMARY

This paper presents finite element methods for the non-stationary Euler equations of a two dimensional inviscid and incompressible flow. For the time discretization, we compare numerical results obtained by the use of a leap-frog scheme and a semi-implicit scheme of order two.

INTRODUCTION

The two-dimensional non-stationary Euler equations for an inviscid and incompressible fluid have been used to describe large-scale atmospheric motions. Non-linear computational instability of finite difference schemes is the major difficulty to overcome for long-term integration of this equation.

Arakawa¹ pointed out that the properties of conservation of kinetic energy and mean-square vorticity or 'enstrophy' must be maintained when the continuous problem is approximated by a numerical scheme.

In a previous paper,² we showed that the use of finite-element spatial discretizations leads to conservative and convergent schemes as soon as the numerical integration of the non-linear term is exact. We gave an error analysis of finite element approximations and we derived stability and convergence theorems for time-differencing schemes.

This time we present numerical results, computational aspects and comparisons in practice of the schemes introduced in Reference 2. In particular, non-trivial periodic solutions of the stream-function have been found and numerically tested during long-term integration.

STATEMENT OF THE PROBLEM

Let Ω be a simply connected plane domain with boundary Γ . We consider the non-stationary Euler equations for an incompressible and inviscid fluid:

$$\frac{\partial \mathbf{u}}{\partial t} + \sum_{i=1}^2 u_i \frac{\partial \mathbf{u}}{\partial x_i} = - \text{grad } p, \quad \text{in } \Omega, \quad (1)$$

$$\text{div } \mathbf{u} = 0, \quad \text{in } \Omega, \quad (2)$$

$$\mathbf{u} \cdot \boldsymbol{\nu} = 0, \quad \text{on } \Gamma, \quad (3)$$

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \quad \text{in } \Omega, \quad (4)$$

where $\boldsymbol{\nu}$ is the unit outward normal to Γ .

Since the flow is incompressible, there is a stream function ψ satisfying

$$\mathbf{u} = \left(\frac{\partial \psi}{\partial x_2}, -\frac{\partial \psi}{\partial x_1} \right). \quad (5)$$

Then introducing the vorticity ω by

$$\omega = \text{curl } \mathbf{u} \quad (6)$$

and using the identity

$$\text{curl } \mathbf{u} = -\Delta \psi \quad (7)$$

we obtain the stream-function formulation of the Euler equations

$$-\Delta \psi = \omega, \quad \text{in } \Omega, \quad (8)$$

$$\frac{\partial \omega}{\partial t} = \frac{\partial \psi}{\partial x_1} \frac{\partial \omega}{\partial x_2} - \frac{\partial \psi}{\partial x_2} \frac{\partial \omega}{\partial x_1} = J(\psi, \omega), \quad \text{in } \Omega, \quad (9)$$

$$\psi(\mathbf{x}, t) = 0, \quad \text{on } \Gamma, \quad (10)$$

$$\psi(\mathbf{x}, 0) = \psi_0(\mathbf{x}), \quad \text{in } \Omega. \quad (11)$$

We now introduce function spaces in order to write these equations in variational form and to derive finite element methods. Let (\cdot, \cdot) denote the scalar product in $L^2(\Omega)$ and $\|\cdot\|_{0,\Omega}$ denote the norm in $L^2(\Omega)$. Let then

$$H^1(\Omega) = \left\{ v \mid v \in L^2(\Omega), \frac{\partial v}{\partial x_i} \in L^2(\Omega), \quad i = 1, 2 \right\}, \quad (12)$$

which is a Hilbert space for the norm

$$\|v\|_{1,\Omega} = \left(\|v\|_{0,\Omega}^2 + \sum_{i=1}^2 \left\| \frac{\partial v}{\partial x_i} \right\|_{0,\Omega}^2 \right)^{1/2}. \quad (13)$$

Let us also consider the following semi-norm in $H^1(\Omega)$

$$|v|_{1,\Omega} = \left(\sum_{i=1}^2 \left\| \frac{\partial v}{\partial x_i} \right\|_{0,\Omega}^2 \right)^{1/2}. \quad (14)$$

It is well known that this defines a norm in $H_0^1(\Omega)$, where

$$H_0^1(\Omega) = \{v \mid v \in H^1(\Omega), v|_{\Gamma} = 0\}. \quad (15)$$

Let $a(u, v) \rightarrow a(u, v)$ be the real bilinear form on $(H_0^1(\Omega))^2$:

$$a(u, v) = \int_{\Omega} \left(\frac{\partial u}{\partial x_1} \frac{\partial v}{\partial x_1} + \frac{\partial u}{\partial x_2} \frac{\partial v}{\partial x_2} \right) dx \quad (16)$$

associated with the variational formulation of the harmonic Dirichlet problem, and let us consider $b:(u, v, w) \rightarrow b(u, v, w)$ defined by

$$b(u, v, w) = \int_{\Omega} \left(\frac{\partial u}{\partial x_1} \frac{\partial v}{\partial x_2} - \frac{\partial u}{\partial x_2} \frac{\partial v}{\partial x_1} \right) w dx. \quad (17)$$

The Euler equation can be expressed in variational form as follows:

$$a(\psi(t), u) = (\omega(t), u), \quad \forall u \in H_0^1(\Omega), \quad (18)$$

$$\left(\frac{d}{dt} \omega(t), v \right) = b(\psi(t), \omega(t), v), \quad \forall v \in H^1(\Omega), \quad (19)$$

$$\psi(0) = \psi_0. \quad (20)$$

We saw in Reference 2 that the form b satisfies the identity

$$b(u, v, w) = b(v, w, u) = b(w, u, v), \quad (21)$$

which leads to the following conservation properties:

$$\frac{1}{2} \frac{d}{dt} \|\omega(t)\|_{0,\Omega}^2 = b(\psi(t), \omega(t), \omega(t)) = 0, \quad (22)$$

$$\frac{1}{2} \frac{d}{dt} |\psi(t)|_{1,\Omega}^2 = b(\psi(t), \omega(t), \psi(t)) = 0, \quad (23)$$

so that the kinetic energy

$$\frac{1}{2} |\psi(t)|_{1,\Omega}^2$$

and the enstrophy

$$\|\omega(t)\|_{0,\Omega}^2$$

are conserved with time.

The fact that numerical schemes keep the properties of the form b and then ensure the conservation of kinetic energy and enstrophy is the main condition of their convergence.

FINITE ELEMENT APPROXIMATION

Standard conforming elements enable us to construct two finite dimensional spaces;

$$U_h \subset H_0^1(\Omega), \quad (24)$$

$$V_h \subset H^1(\Omega). \quad (25)$$

Then we approximate the Euler equations by the following discrete formulation:

$$a(\psi_h(t), u_h) = (\omega_h(t), u_h), \quad \forall u_h \in U_h, \quad (26)$$

$$\left(\frac{d}{dt} \omega_h(t), v_h \right) = b(\psi_h(t), \omega_h(t), v_h), \quad \forall v_h \in V_h, \quad (27)$$

$$\psi_h(0) = \psi_{0,h}. \quad (28)$$

Examples

Triangular elements. For each integer k , we denote by P_k , the space of all polynomials defined on \mathbb{R}^2 of degree less or equal to k . Then, for the construction of the finite dimensional spaces U_h and V_h , we use a conforming finite element method associated with a regular triangulation made with triangles K and, on each K , polynomials of P_k .

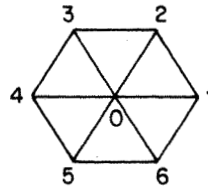


Figure 1

Quadrilateral elements. In this case, we denote by \hat{K} the unit reference square with vertices $\hat{a}_1 = (0, 0)$, $\hat{a}_2 = (1, 0)$, $\hat{a}_3 = (1, 1)$, $\hat{a}_4 = (0, 1)$, and we denote by F_K the mapping that maps the unit square \hat{K} onto the quadrilateral element K . Now for each integer k , we denote by Q_k the space of all polynomials spanned by $x_1^{\alpha_1} x_2^{\alpha_2}$ with $0 \leq \alpha_i \leq k$ for $i = 1, 2$.

Then for the construction of the finite dimensional spaces U_h and V_h we shall use a conforming finite element method associated with quadrilaterals K and, on each K , the functions obtained by composition with F_K^{-1} of polynomials of Q_k .

In both cases of triangular and quadrilateral elements we derived the following error bound as soon as the exact solution (ψ, ω) is sufficiently smooth:

$$|\psi(t) - \psi_h(t)|_{1,\Omega} + \|\omega(t) - \omega_h(t)\|_{0,\Omega} \leq Ch^k, \quad \forall t \in [0, T]. \tag{29}$$

In particular, let us choose a regular ‘triangulation’ of Ω made with equal rectangles, and let us choose as V_h the space of continuous functions on $\bar{\Omega}$ such as their restrictions on each rectangle K belong to Q_1 . We obtain the conservative scheme introduced by Arakawa for the Jacobian (see also Reference 3):

$$\begin{aligned} J(\psi, \omega)_{i,j} = & \frac{1}{12\Delta x \Delta y} \{ \omega_{i+1,j+1}(\psi_{i+1,j} - \psi_{i,j+1}) \\ & + \omega_{i-1,j+1}(\psi_{i,j+1} - \psi_{i-1,j}) + \omega_{i-1,j-1}(\psi_{i-1,j} - \psi_{i,j-1}) \\ & + \omega_{i+1,j-1}(\psi_{i,j-1} - \psi_{i+1,j}) \\ & + \omega_{i+1,j}(\psi_{i,j-1} - \psi_{i,j+1} + \psi_{i+1,j-1} - \psi_{i+1,j+1}) \\ & + \omega_{i,j+1}(\psi_{i+1,j} - \psi_{i-1,j} + \psi_{i+1,j+1} - \psi_{i-1,j+1}) \\ & + \omega_{i-1,j}(\psi_{i,j+1} - \psi_{i,j-1} + \psi_{i-1,j+1} - \psi_{i-1,j-1}) \\ & + \omega_{i,j-1}(\psi_{i-1,j} - \psi_{i+1,j} + \psi_{i-1,j-1} - \psi_{i+1,j-1}) \}. \end{aligned} \tag{30}$$

It is also possible to consider a regular triangulation of Ω made of equilateral triangles and to take as V_h the space of continuous functions such that their restrictions on each triangle K belong to P_1 . There is an advantage to do so from the point of view of numerical quadrature. We obtain the following analogue of the Jacobian (see Figure 1):

$$\begin{aligned} J(\psi, \omega)_0 = & \frac{1}{6h^2} [\omega_1(\psi_6 - \psi_2) + \omega_2(\psi_1 - \psi_3) + \omega_3(\psi_2 - \psi_4) \\ & + \omega_4(\psi_3 - \psi_5) + \omega_5(\omega_4 - \psi_6) + \omega_6(\psi_5 - \psi_1)]. \end{aligned} \tag{31}$$

TIME DISCRETIZATION

The most generally used time discretization scheme for the Euler equation is the leap-frog scheme. But this scheme may presents instability and explosive growth of energy and enstrophy in the case

of long-term integration. The importance of these properties leads us to consider numerical schemes satisfying the conservation of kinetic energy and enstrophy. This implies the choice of semi-implicit schemes.

The leap-frog scheme

Let us choose a positive integer N ; let Δt denote the corresponding time step $\Delta t = T/N$ and $t_n = n\Delta t$ for $n = 0, \dots, N$. Let ψ_h^n and ω_h^n denote the approximate values of $\psi(t_n)$ and $\omega(t_n)$. The leap-frog scheme can be written as follows:

$$a(\psi_h^n, u_h) = (\omega_h^n, u_h), \quad \forall u_h \in U_h, \tag{32}$$

$$(\omega_h^{n+1}, v_h) = (\omega_h^{n-1}, v_h) + 2\Delta t b(\psi_h^n, \omega_h^n, v_h), \quad \forall v_h \in V_h, \tag{33}$$

$$\omega_h^0 = \omega_0, \tag{34}$$

$$(\omega_h^1, v_h) = (\omega_h^0, v_h) + \Delta t b(\psi_h^0, \omega_h^0, v_h), \quad \forall v_h \in V_h. \tag{35}$$

We derived in Reference 2 the following stability and convergence results. Under the following stability hypothesis:

- (i) a Courant–Friedrichs–Lewy type condition

$$\frac{C\Delta t}{h} \sup_{\mathbf{x} \in \Omega} |\mathbf{u}_h^n(\mathbf{x})| < 1, \quad \forall n = 0, \dots, N, \tag{36}$$

where the constant C depends on the type of finite elements used and where $\sup_{\mathbf{x} \in \Omega} |\mathbf{u}_h^n(\mathbf{x})|$ represents the maximum value of the velocity in Ω

- (ii) there exists a constant $A > 0$ such that

$$\frac{C}{h} \sup_{\mathbf{x} \in \Omega} |\mathbf{u}_h^n(\mathbf{x}) - \mathbf{u}_h^{n-1}(\mathbf{x})| < A, \quad \forall n = 1, \dots, N \tag{37}$$

we have the following bound:

$$\|\omega_h^n\|_{0,\Omega}^2 + \|\omega_h^{n+1}\|_{0,\Omega}^2 \leq C (\|\omega_h^0\|_{0,\Omega}^2 + \|\omega_h^1\|_{0,\Omega}^2), \quad \forall n = 0, \dots, N-1. \tag{38}$$

If we assume furthermore sufficient regularity conditions on the exact solution (ψ, ω) , we can get the error estimate

$$|\psi(t_n) - \psi_h^n|_{1,\Omega} + \|\omega(t_n) - \omega_h^n\|_{0,\Omega} \leq C(h^k + \Delta t^2), \quad \forall n = 0, \dots, N. \tag{39}$$

Remark 1. Condition (37) seems rather theoretical. In fact, the constant C in (38) may become very large with the integration time T . Namely

$$C = 2K \exp(\text{AKT}), \tag{40}$$

with

$$K = \max_{0,N} \frac{1}{1 - \frac{C\Delta t}{h} \sup_{\mathbf{x} \in \Omega} |\mathbf{u}_h^n(\mathbf{x})|}. \tag{41}$$

Furthermore, it may be observed that the stability condition (36) does not imply the condition (37). Condition (37) shows itself to be more restrictive than (36). A practical way to relax this condition is to use a mixing between the successive steps; For example, a periodic mixing at every

M steps, M being adjusted to maintain conservative properties:

$$\begin{aligned}\tilde{\omega}_n &= \omega_{n-2} + 2\Delta t J(\psi_{n-1}, \omega_{n-1}), \\ \tilde{\omega}_{n+1} &= \omega_{n-1} + 2\Delta t J(\tilde{\psi}_n, \tilde{\omega}_n), \\ \omega_n &= \frac{\tilde{\omega}_{n+1} + 2\tilde{\omega}_n + \omega_{n-1}}{4},\end{aligned}\tag{42}$$

or a mixing at every time step of the form

$$\begin{aligned}\tilde{\omega}_n &= \omega_{n-2} + 2\Delta t J(\psi_{n-1}, \omega_{n-1}), \\ \tilde{\omega}_{n+1} &= \omega_{n-1} + 2\Delta t J(\tilde{\psi}_n, \tilde{\omega}_n), \\ \omega_n &= (1 - \varepsilon)\tilde{\omega}_n + \varepsilon \frac{\tilde{\omega}_{n+1} + 2\tilde{\omega}_n + \omega_{n-1}}{4},\end{aligned}\tag{43}$$

where ε is a parameter to fit in order to keep kinetic energy and enstrophy constant.

Remark 2. It is well known now,^{1,3} and our numerical experiments have confirmed it, that the non-linear term must be exactly integrated. For the L^2 scalar product

$$(\omega_h, v_h)\tag{44}$$

on the contrary, we can use approximate integration. Doing this, the leap-frog scheme is explicit and there is no linear system to solve at each time step. This is what is called ‘lumping’. We obtain it by the use of an approximate quadrature formula involving the interpolation points.

For example, with finite elements Q_k , the formula using interpolation points is exact for polynomials of degree less than or equal to $2k - 1$. This ensures an integration error of order k , that is the same order as finite element interpolation error.⁴

Thus, from a theoretical point of view, there is no loss of order of accuracy when the mass matrix is replaced by a lumped diagonal matrix. Numerical comparisons are presented in the next section. Numerical tests have shown that the ‘lumping’ actually stabilizes the leap-frog scheme in the case of Euler equations.

A semi-implicit scheme of order two

The following scheme ensures exact conservation of enstrophy. Numerical experiments have shown that kinetic energy is also constant with time when the exactly integrated formulation b of the non-linear term (17) is used.

Let ψ_h^n and ω_h^n be approximations of $\psi(t_n)$ and $\omega(t_n)$, we write

$$a(\psi_h^n, u_h) = (\omega_h^n, u_h), \quad \forall u_h \in U_h,\tag{45}$$

$$a(\psi_h^{n+1/2}, u_h) = (\omega_h^n, u_h) + \frac{\Delta t}{2} b(\psi_h^n, \omega_h^n, u_h), \quad \forall u_h \in U_h,\tag{46}$$

$$(\omega_h^{n+1}, v_h) = (\omega_h^n, v_h) + \frac{\Delta t}{2} b(\psi_h^{n+1/2}, \omega_h^n + \omega_h^{n+1}, v_h), \quad \forall v_h \in V_h,\tag{47}$$

$$\omega_h^0 = \omega_{0,h}.\tag{48}$$

We call this scheme a semi-implicit scheme because it is only implicit with respect to ω_h . It

ensures the exact conservation of approximate enstrophy:

$$\|\omega_h^n\|_{0,h}^2 = \|\omega_h^0\|_{0,h}^2, \quad \forall n = 0, \dots, N. \quad (49)$$

This scheme appears to be a mid-point rule for ψ associated with a Crank-Nicolson scheme for ω .

For the exact solution (ψ, ω) , we have

$$\omega(t_{n+1}) = \omega(t_n) + \frac{\Delta t}{2} J(\psi(t_{n+1/2}), \omega(t_n) + \omega(t_{n+1})) + O(\Delta t^3), \quad (50)$$

with

$$\psi(t_{n+1/2}) = \psi(t_n) + \frac{\Delta t}{2} \frac{\partial \psi}{\partial t}(t_n) + O(\Delta t^2) \quad (51)$$

such as

$$-\Delta \psi(t_{n+1/2}) = \omega(t_n) + \frac{\Delta t}{2} J(\psi(t_n), \omega(t_n)) + O(\Delta t^2), \quad (52)$$

so that this scheme is clearly of order two in time.

More precisely, suppose that the exact solution (ψ, ω) is sufficiently smooth, we derived in (Reference 2) the following error bound:

$$\|\omega(t_n) - \omega_h^n\|_{0,\Omega} + |\psi(t_n) - \psi_h^n|_{1,\Omega} < C(h^k + \Delta t^2) \quad (53)$$

Computational aspects of the semi-implicit scheme

This scheme implies the resolution of two discrete Dirichlet problems at each time step. This has been done by factorizing the discrete Laplacian matrix once and for all using Cholesky's algorithm.

The resolution of the vorticity equation (47) is more tricky. The problem can be written as follows: ω_h^n and ψ_h^n being already calculated, find ω_h^{n+1} such that

$$(\omega_h^{n+1}, v_h) - \frac{\Delta t}{2} b(\psi_h^{n+1/2}, \omega_h^{n+1}, v_h) = (\omega_h^n, v_h) + \frac{\Delta t}{2} b(\psi_h^{n+1/2}, \omega_h^n, v_h), \quad \forall v_h \in V_h. \quad (54)$$

If p denotes the number of nodes of the finite element method, this leads to the following inversion matrix problem in \mathbb{R}^p at each time step:

$$(\mathbf{I} - \mathbf{J}_n) \bar{\omega}^{n+1} = (\mathbf{I} + \mathbf{J}_n) \bar{\omega}^n, \quad (55)$$

\mathbf{I} being the $p \times p$ identity matrix; \mathbf{J}_n is a skew-symmetric matrix depending on $\psi_h^{n+1/2}$ and $\bar{\omega}^n$ is the vector solution for each time t_n .

This problem has a unique solution since

$$((\mathbf{I} - \mathbf{J}_n) \mathbf{V}, \mathbf{V})_{\mathbb{R}^p} = \|\mathbf{V}\|_{\mathbb{R}^p}^2, \quad \forall \mathbf{V} \in \mathbb{R}^p. \quad (56)$$

Let us now review different algorithms to solve it.

The relaxation algorithm. Let us consider the system of linear equations

$$\mathbf{A} \mathbf{x} = \mathbf{b},$$

where $\mathbf{A} = (a_{i,j})$ is a given $p \times p$ real matrix and \mathbf{b} is a given real p -vector.

Let \mathbf{x}^0 be a given vector; the computational scheme can be written as follows, with $a_{i,i} = 1, \forall i$:

$$\begin{aligned} x_i^{n+1} = & \omega(b_i - a_{i,1}x_1^{n+1} \cdots - a_{i,i-1}x_{i-1}^{n+1}) \\ & + (1 - \omega)x_i^n + \omega(-a_{i,i+1}x_{i+1}^n \cdots - a_{i,p}x_p^n), \quad \text{for } i = 1, \dots, p, \end{aligned}$$

where ω is the relaxation parameter. Numerical experiments have shown that the optimal values of ω were

$$\omega = 0.95, \quad \text{for } \frac{\Delta t}{h} \simeq 0.2,$$

$$\omega = 0.90, \quad \text{for } \frac{\Delta t}{h} \simeq 0.5.$$

This algorithm is very easy to implement, but the extra-diagonal coefficients of the matrix become more important when h decreases or when Δt increases. Then the relaxation algorithm becomes less efficient. It may even diverge when the matrix \mathbf{A} is not diagonally dominant. The necessity to reduce, in this case, the time step Δt leads to higher computational costs.

The generalized conjugate gradient method. This was introduced by Concus and Golub.⁵ Let us consider the system of linear equations

$$\mathbf{Ax} = \mathbf{b}$$

where \mathbf{A} is a $p \times p$ real matrix and \mathbf{b} a given real p -vector. We rewrite the system

$$\mathbf{Mx} = \mathbf{Nx} + \mathbf{b},$$

where $\mathbf{M} = \mathbf{M}^T = (\mathbf{A} + \mathbf{A}^T)/2$ is the symmetric part of \mathbf{A} , and

$$-\mathbf{N} = \mathbf{N}^T = (\mathbf{A} - \mathbf{A}^T)/2$$

is its skew-symmetric part.

Assuming that \mathbf{M} is positive definite and that it is a simpler computational task to solve

$$\mathbf{Mz} = \mathbf{d},$$

the generalized conjugate gradient method for the splitting $(\mathbf{A} + \mathbf{A}^T)/2$ is summarized as follows:

Algorithm

Let \mathbf{x}^0 be a given vector and arbitrarily define \mathbf{x}^k for $k = 0, 1$

1. Solve $\mathbf{MZ}^k = \mathbf{r}^k$ when $\mathbf{r}^k = \mathbf{b} - \mathbf{Ax}^k$.

2. Compute $\omega_{k+1} = \left(1 + \frac{\mathbf{Z}^{kT}\mathbf{MZ}^k}{\mathbf{Z}^{k-1T}\mathbf{MZ}^{k-1}} \frac{1}{\omega_k}\right)^{-1}$ for $k \geq 1$, with $\omega_1 = 1$.

3. Compute $\mathbf{x}^{k+1} = \mathbf{x}^{k-1} + \omega_{k+1}(\mathbf{Z}^k + \mathbf{x}^k - \mathbf{x}^{k-1})$

Properties of the method in our case.

Let us recall that $\mathbf{A} = \mathbf{I} - \mathbf{J}_n$, where \mathbf{J}_n is skew symmetric, so that the symmetric part of \mathbf{A} is \mathbf{I} , and the first step of the algorithm is removed.

This method is then very easy to implement in this case.

The Chebyshev iteration. We present a short overview of the Chebyshev iteration for a non-symmetric matrix according to Van der Vorst.⁶ Let us consider the linear system

$$\mathbf{Ax} = \mathbf{b}.$$

The computational scheme can be written as follows:

Algorithm

Given \mathbf{x}^0 , define

$$\mathbf{r}^0 = \mathbf{b} - \mathbf{Ax}^0,$$

$$\mathbf{p}^0 = \frac{1}{d}\mathbf{r}^0,$$

$$\alpha^0 = \frac{2}{d},$$

then

$$\mathbf{x}^k = \mathbf{x}^{k-1} + \mathbf{p}^{k-1},$$

$$\mathbf{r}^k = \mathbf{b} - \mathbf{Ax}^k,$$

$$\alpha^k = \left[d - \left(\frac{c}{2} \right)^2 \alpha^{k-1} \right]^{-1},$$

$$\beta^k = d\alpha^k - 1,$$

$$\mathbf{p}^k = \alpha^k \mathbf{r}^k + \beta^k \mathbf{p}^{k-1},$$

for $k = 1, 2, \dots$, where the constant c is the focal distance and d is the centre of the ellipse that encloses the spectrum of \mathbf{A} .

In our case $\mathbf{A} = \mathbf{I} - \mathbf{J}_n$ has all its eigenvalues of the form $\lambda = 1 + i\mu$, where μ is real. Then the ellipse enclosing the spectrum of \mathbf{A} degenerates into a segment $(1 - ic, 1 + ic)$ (see Figure 2).

The bi-conjugate gradient method. This was introduced by D. A. H. Jacobs.⁷ This method extends the principle of the conjugate gradient method to non-symmetric matrices. Let us consider a system

$$\mathbf{Ax} = \mathbf{b}$$

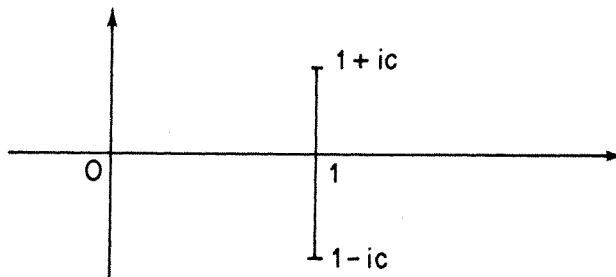


Figure 2

where \mathbf{A} is non-symmetric. We consider

$$\begin{aligned}\mathbf{r} &= \mathbf{b} - \mathbf{A}\mathbf{x}, \\ \bar{\mathbf{r}} &= \mathbf{b} - \mathbf{A}^T\mathbf{x},\end{aligned}$$

which leads to the following algorithm:

Algorithm

Given \mathbf{x}^0 , calculate the initial residual

$$\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0$$

and the initial bi-residual

$$\bar{\mathbf{r}}^0 = \mathbf{r}^0.$$

Set the first direction vector

$$\mathbf{p}^0 = \mathbf{r}^0$$

and the first bi-direction vector

$$\bar{\mathbf{p}}^0 = \bar{\mathbf{r}}^0.$$

Then for each $k = 0, 1, \dots$ calculate

$$\alpha^k = \frac{(\bar{\mathbf{r}}^k, \mathbf{p}^k)}{(\bar{\mathbf{p}}^k, \mathbf{A}\mathbf{p}^k)}.$$

Calculate the new estimate

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{p}^k.$$

Determine the new residual

$$\mathbf{r}^{k+1} = \mathbf{r}^k - \alpha^k \mathbf{A}\mathbf{p}^k$$

and the new bi-residual

$$\bar{\mathbf{r}}^{k+1} = \bar{\mathbf{r}}^k - \alpha^k \mathbf{A}^T \bar{\mathbf{p}}^k.$$

Determine the coefficient

$$\beta^k = \frac{-(\mathbf{r}^{k+1}, \mathbf{A}\mathbf{p}^k)}{(\bar{\mathbf{p}}^k, \mathbf{A}\mathbf{p}^k)}.$$

Set the new direction vector

$$\mathbf{p}^{k+1} = \mathbf{r}^{k+1} + \beta^k \mathbf{p}^k$$

and the new bi-direction

$$\bar{\mathbf{p}}^{k+1} = \bar{\mathbf{r}}^{k+1} + \beta^k \bar{\mathbf{p}}^k.$$

The multigrid method. This will be too long to present here the multigrid algorithm. It is described, for example, by Brandt.⁸

NUMERICAL EXPERIMENTS

We tested the above schemes on the problems described below. Let Ω be the square $(0, \pi) \times (0, \pi)$. We solve equations (8)–(11) in Ω with various initial conditions.

1. We first tried the following initial condition (see Figure 3):

$$\psi_0(x, y) = \sin(x)\sin(2y) \quad (\text{C1})$$

which is an eigenfunction of the Dirichlet problem in the square Ω . This initial condition gives a stationary solution. The finite element scheme satisfies the stationarity without errors for long integration times. The choice of either a leap-frog scheme or a semi-implicit scheme does not modify the numerical results in this case.

2. Then after Arakawa and Lamb (Reference 9), and for meteorological modelling purposes, we chose the two initial conditions obtained by addition of eigenfunctions of the Laplace operator:

$$\psi_0(x, y) = \sin(x)\sin(y) + \sin(2x)\sin(y), \quad \text{on } \Omega, \quad (\text{C2})$$

$$\psi_0(x, y) = \sin(2x)\sin(2y) + \sin(2x)\sin(3y), \quad \text{on } \Omega \quad (\text{C3})$$

(see Figures 4 and 5, respectively.)

For the initial condition C2 the initial value of enstrophy is

$$\|\omega\|_{0,\Omega}^2 = 71.6.$$

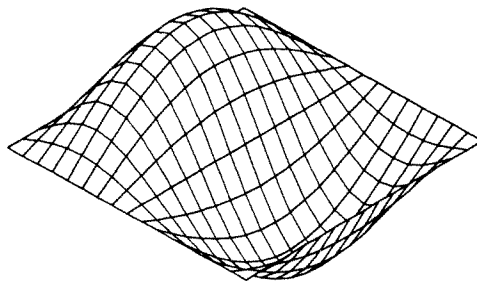


Figure 3. Initial condition C1: $\psi_0(x, y) = \sin(x)\sin(2y)$. Stationary solution for $T = 0$ to $T = 100$

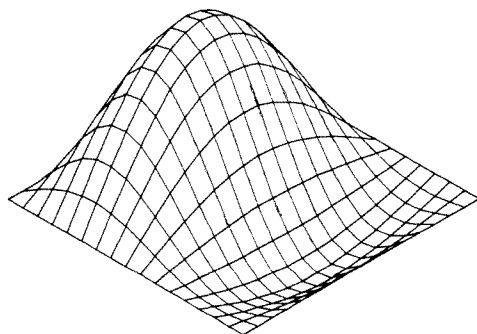


Figure 4. Initial condition C2: $\psi_0(x, y) = \sin(x)\sin(y) + \sin(2x)\sin(y)$. $T = 0$

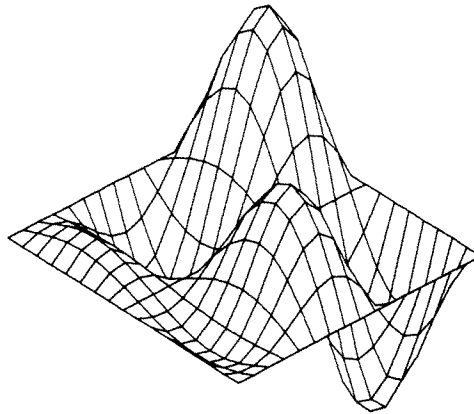


Figure 5. Initial condition C3: $\psi_0(x, y) = \sin(2x)\sin(2y) + \sin(2x)\sin(3y)$. $T = 0$

The value of energy corresponds to

$$|\psi|_{1,\Omega}^2 = 17.4.$$

Likewise for the initial condition C3, we have

$$\|\omega\|_{0,\Omega}^2 = 575,$$

$$|\psi|_{1,\Omega}^2 = 52.9.$$

For each initial condition, we perform several computations with various values of the mesh size h and the time step Δt . We first compare the behaviours of leap-frog and semi-implicit time discretization schemes.

The test is to maintain the conservation laws on kinetic energy and enstrophy as long as possible.

Time discretization

Leap-frog scheme. The leap-frog scheme appears often to be unstable in the case of long-term integration or with rough initial conditions. Its stability implies that a Courant-Friedrichs-Lewy

Table I. Leap-frog scheme (exactly integrated)

		$T = 0$	$T = 10$	$T = 20$
<i>Initial condition (C2)</i>				
$h = \frac{\pi}{10}$				
$\Delta t = 0.05$	Enstrophy	71.6	80.7	93.4
	Energy	17.7	17.6	17.8
$\Delta t = 0.1$			unstable	
$h = \frac{\pi}{16}$				
$\Delta t = 0.02$	Enstrophy	71.6	78.6	85.6
	Energy	17.4	17.4	17.2
$\Delta t = 0.05$			unstable	

type condition must be set on Δt . This constrains us to choose a smaller time step Δt when the mesh h is smaller, which then increases the computational cost of the leap-frog scheme.

We tested the leap-frog scheme in the case of initial conditions (C2) and (C3) with different values of h . And we tried to maintain the conservation constraints and stability until the integration time $T = 100$. We present the numerical results in Tables I and II.

Numerical tests presented in Tables I and II have clearly shown that lumping stabilizes the leap-frog scheme in the case of Euler equations.

After lumping, we remark that in the case of smooth solutions corresponding to initial condition (C2), the behaviour of the leap-frog scheme is satisfactory, whereas for initial condition (C3), the

Table II. Leap-frog scheme (with lumping of the mass matrix)

		$T = 0$	$T = 50$	$T = 100$
<i>Initial condition (C2)</i>				
$h = \frac{\pi}{10}$				
$\Delta t = 0.05$	Enstrophy	71.6	71.6	71.6
	Energy	17.7	17.7	17.7
$\Delta t = 0.1$	Enstrophy	71.6	71.6	71.8
	Energy	17.7	17.7	17.7
$\Delta t = 0.2$	Enstrophy	71.6	91.5	overflow
	Energy	17.7	18.2	at $T = 51$
<i>Initial condition (C2)</i>				
$h = \frac{\pi}{16}$				
$\Delta t = 0.02$	Enstrophy	71.6	71.6	71.6
	Energy	17.4	17.4	17.4
$\Delta t = 0.05$	Enstrophy	71.6	71.7	71.8
	Energy	17.4	17.4	17.4
$\Delta t = 0.1$	Enstrophy	71.5	overflow	
	Energy	17.4	at $T = 20$	
<i>Initial condition (C3)</i>				
$h = \frac{\pi}{10}$				
$\Delta t = 0.02$	Enstrophy	575.4	580	616.8
	Energy	54.5	55.8	52.8
$\Delta t = 0.05$	Enstrophy	575.4	597.1	overflow
	Energy	54.5	54.3	at $T = 80$
$\Delta t = 0.1$	Enstrophy	575.4	overflow	
	Energy	54.5	at $T = 40$	
$h = \frac{\pi}{16}$				
$\Delta t = 0.02$	Enstrophy	575.4	601	overflow
	Energy	52.9	55.7	at $T = 60$
$\Delta t = 0.05$	Enstrophy	575.4	687	overflow
	Energy	52.9	56.2	at $T = 51$

Table III. Leap-frog with ε mixing.

$h = \frac{\pi}{10}$		$T = 0$	$T = 50$	$T = 100$
<i>Initial condition (C2)</i>				
$\Delta t = 0.2$				
$\varepsilon = 0.06$	Enstrophy	71.6	70.8	77.7
	Energy	17.7	17.7	17.7
$\varepsilon = 0.07$	Enstrophy	71.6	70.1	69.4
	Energy	17.7	17.5	17.3
<i>Initial condition (C3)</i>				
$\Delta t = 0.1$				
$\varepsilon = 0.1$	Enstrophy	575.4	500.4	432.3
	Energy	54.5	52	50
$\varepsilon = 0.02$	Enstrophy	575.4	568.5	591.2
	Energy	54.5	54.9	48.2
$\Delta t = 0.05$				
$\varepsilon = 0.02$	Enstrophy	575.4	569.5	554.7
	Energy	54.4	54.4	53.9
$\varepsilon = 0.01$	Enstrophy	575.4	573.4	575.4
	Energy	54.5	54.1	55.1

stability condition requires very small Δt . In fact, in this case, we have not been able to carry out the computation until $T = 100$ even with time step $\Delta t = 0.02$. This is the reason why smoothing procedures have been tried in order to stabilize the leap-frog scheme. We have tested several kinds of smoothing among the most commonly used. Our conclusion is the following: the mixing procedures (42) and (43) can be efficient if the parameters M and ε have been carefully chosen. The numerical results depend very sensibly on the values of M and ε . For small variations of these parameters, the scheme may produce either too much damping or instability.

However, the most efficient procedure, for our point of view, is the ε mixing defined in (43). But this procedure makes necessary twice as many computations and, in some cases, it is a better choice to halve the time step.

We present in Table III some results of computations with a mixing procedure.

Semi-implicit scheme. Numerical experiments have confirmed the remarkable conservation properties of this scheme. These properties allow us to take larger time steps and then to decrease the computational cost of the semi-implicit scheme. Comparisons between computational times of leap-frog and semi-implicit schemes are presented in Table IV. They show that in the case of a rough initial condition and long-term integration, the semi-implicit scheme succeeded where the leap-frog scheme failed and that its computational cost became competitive.

So far, the algorithm used to solve the inversion matrix problem (55), which occurs with the semi-implicit scheme, has been the relaxation algorithm. But computational instabilities make it divergent when the extra-diagonal coefficients of the matrix become dominant, for example when the ratio $\Delta t/h$ increases. This sets the problem of mesh refinements. This is the reason why we have tested other iterative methods well-suited for solving non-symmetric systems. We summarize in Table V the first results of our computations.

Table IV.

		$T = 100$	Computer time, s
<i>Initial condition (C2)</i>			
$h = \frac{\pi}{10}$			
Leap-frog scheme			
$\Delta t = 0.05$	Enstrophy	71.6	720
	Energy	17.7	
$\Delta t = 0.1$	Enstrophy	71.8	350
	Energy	17.7	
Semi-implicit scheme			
$\Delta t = 0.1$	Enstrophy	71.6	1320
	Energy	17.7	
$\Delta t = 0.2$	Enstrophy	71.6	710
	Energy	17.5	
$h = \frac{\pi}{16}$			
Leap-frog scheme			
$\Delta t = 0.02$	Enstrophy	71.6	3810
	Energy	17.4	
$\Delta t = 0.05$	Enstrophy	71.8	1529
	Energy	17.4	
Semi-implicit scheme			
$\Delta t = 0.1$	Enstrophy	71.6	4850
	Energy	17.4	
<i>Initial condition (C3)</i>			
$h = \frac{\pi}{10}$			
Leap-frog scheme			
$\Delta t = 0.02$	Enstrophy	616.8	3845
	Energy	52.8	
Semi-implicit scheme			
$\Delta t = 0.05$	Enstrophy	575.4	2920
	Energy	54.2	
$\Delta t = 0.1$	Enstrophy	575.1	1510
	Energy	51	
$h = \frac{\pi}{16}$			
Leap-frog scheme			
$\Delta t = 0.02$		overflow after $T = 60$	
Semi-implicit scheme			
$\Delta t = 0.05$	Enstrophy	575	8719
	Energy	51.4	

Table V.

	Average computer time per time step, s
$h = \frac{\pi}{16}, \Delta t = 0.1$	
Relaxation, $\omega = 0.90$	5.0
Concus and Golub G.C.G	5.2
Chebysheff iteration	3.8
Bi-conjugate gradient	7.0
Multigrid method	5.5
$h = \frac{\pi}{32}, \Delta t = 0.05$	
Relaxation, $\omega = 0.90$	44
Concus and Golub G.C.G	57
Chebyshev iteration	39
Bi-conjugate gradient	59
Multigrid method	43

Computations have been performed on a VAX mini-computer of Digital Equipment Company

Graphical representations and spatial discretizations

In order to make precise the behaviour of approximate solutions, we have been led to use graphical representations of the stream-function ψ at different times. These representations have shown two noticeable results.

First, from a theoretical point of view, a consequence of the possibility, with the semi implicit scheme, to follow the stream function solution during long-term integration, has been to exhibit a periodic solution of Euler equations in the case of initial condition (C2)—see Figures 6–9.

Then, from a numerical point of view, we perform computations in the case of initial condition (C2) with several values of the mesh size h . The study of the graphical representations of the stream function solutions at $T = 20$ for different h , shows that if for every choice of the mesh size, the shape

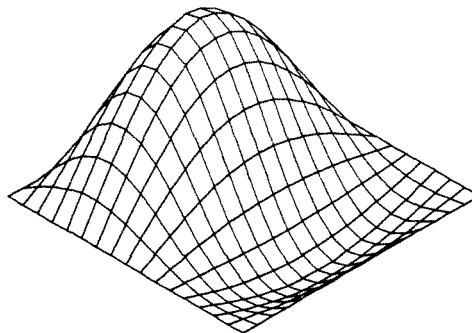


Figure 6. Initial condition C2. $T = 0$; 'Enstrophy' = 71.6; 'Energy' = 17.4

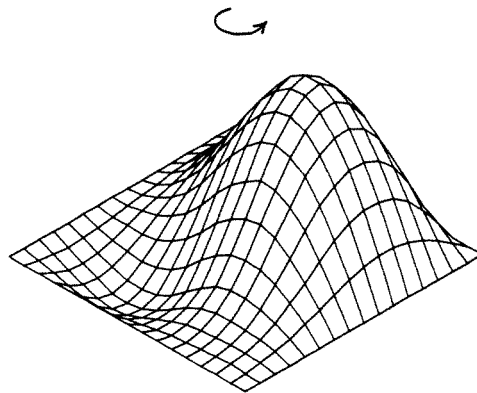


Figure 7. Initial condition C2. $T = 10$; 'Enstrophy' = 71.6; 'Energy' = 17.4

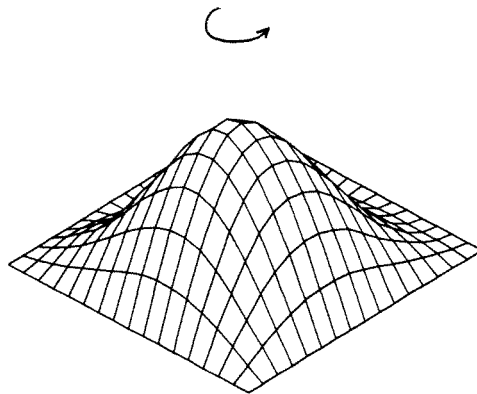


Figure 8. Initial condition C2. $T = 20$; 'Enstrophy' = 71.6; 'Energy' = 17.4

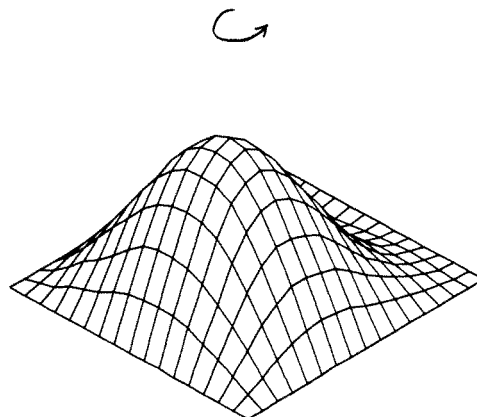


Figure 9. Initial condition C2. $T = 100$; 'Enstrophy' = 71.6; 'Energy' = 17.4

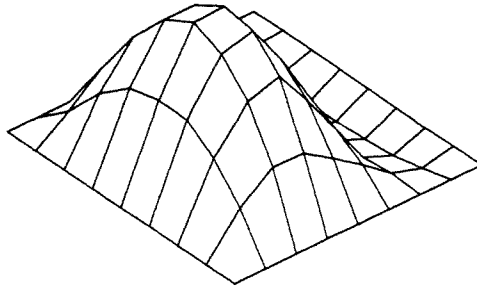


Figure 10. Initial condition C2. Stream function solution at $T = 20$. $h = \frac{\pi}{8}$; 'Enstrophy' = 71.6; 'Energy' = 17.9

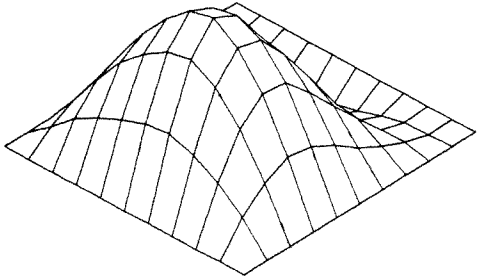


Figure 11. Initial condition C2. Stream function solution at $T = 20$. $h = \frac{\pi}{10}$; 'Enstrophy' = 71.6; 'Energy' = 17.7

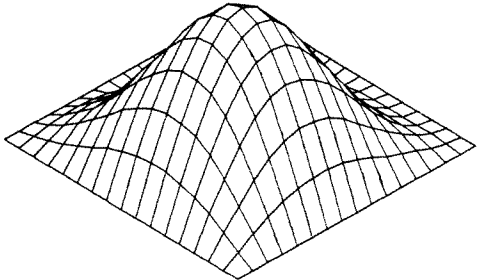


Figure 12. Initial condition C2. Stream function solution at $T = 20$. $h = \frac{\pi}{16}$; 'Enstrophy' = 71.6; 'Energy' = 17.4

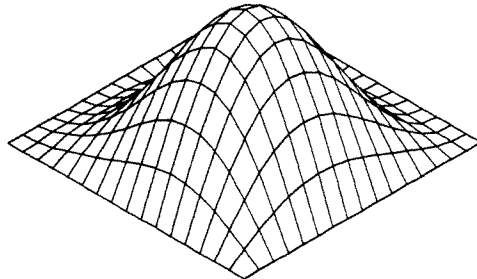


Figure 13. Initial condition C2. Stream function solution at $T = 20$. $h = \frac{\pi}{32}$; 'Enstrophy' = 71.6; 'Energy' = 17.4

of the stream function is qualitatively maintained, the enstrophy and kinetic energy remain exactly constant; on the other hand, the velocity of the rotation phenomenon depends on the mesh size h . It seems that we have the appearance of group velocity errors in two dimensions as mentioned in the papers of Trefethen¹⁰ and Bamberger *et al.*¹¹ So far, owing to the non-linearity of the equations, it has not been possible to specify the variations of the velocity of the phenomenon with the mesh size h . Complementary work is to be done in that direction (see Figures 10–13).

Further experiments

We also test the above schemes on open ocean modelling equations, after Haidvogel *et al.*¹² We consider the forced non-linear box mode problem (see Reference 12, p. 34)

$$\frac{\partial}{\partial t} \nabla^2 \psi + \varepsilon J(\psi, \nabla^2 \psi) + \psi_x = F(x, y, t),$$

where

$$F(x, y, t) = \left\{ \frac{\partial}{\partial t} \nabla^2 + \varepsilon J(\psi_a, \nabla^2) + \frac{\partial}{\partial x} \right\} \psi_a(x, y, t),$$

with

$$\psi_a(x, y, t) = \sin x \sin y \cos(ax + by + ct), \quad 0 \leq x, y \leq \pi,$$

We choose¹²

$$a = \frac{1}{\sqrt{2}}, \quad b = \frac{1}{\sqrt{2}}, \quad c = \frac{1}{2}, \quad \varepsilon = 0.2.$$

The results are presented in Table VI. These results are very close to the corresponding results of Haidvogel *et al.*¹²

Table VI.

	RMS(ω)	RMS(ψ)	NDIF
$h = \frac{\pi}{16} \quad dt = \frac{\pi}{64}$	6.35×10^{-2}	1.17×10^{-2}	7.29×10^{-2}
$h = \frac{\pi}{32} \quad dt = \frac{\pi}{64}$	5.93×10^{-2}	5.76×10^{-3}	6.23×10^{-2}
$h = \frac{\pi}{32} \quad dt = \frac{\pi}{128}$	4.69×10^{-2}	3.96×10^{-3}	7.81×10^{-3}

Duration 2 periods

$$\text{RMS}(\psi) = \left\{ \iint (\psi - \psi_a)^2 dA / \iint (\psi_a)^2 dA \right\}^{1/2}$$

$$\text{RMS}(\omega) = \left\{ \iint (\omega - \omega_a)^2 dA / \iint (\omega_a)^2 dA \right\}^{1/2}$$

$$\text{NDIF} = \left\{ \iint |\nabla \psi|^2 dA - \iint |\nabla \psi_a|^2 dA \right\} / \iint |\nabla \psi_a|^2 dA$$

CONCLUSION

Let us recall that our purpose was to compute the stream-function and the vorticity solutions of the two-dimensional non-stationary Euler equation in long-term integration while keeping constant kinetic energy and enstrophy, as necessary in atmospheric motions model.

The finite element associated with a semi-implicit time scheme of order two provides a reliable solution of that problem. This way, we have been able to follow the solution during long-term integration. Moreover we exhibit numerically a solution of the Euler equations which seems to be periodic.

From a practical point of view, although the semi-implicit scheme may appear already competitive in some cases, complementary works have to be done in order to reduce its computational cost.

REFERENCES

1. A. Arakawa, 'Computational design for long-term integrations of fluid-motion', *J. Comput. Phys.*, **1**, 119–149 (1966).
2. J.-H. Saiaç, 'Finite element method for time dependent Euler equation', *Math. Meth. in the Appl. Sci.*, **5**, 22–39 (1983).
3. D. C. Jaspersen, 'Arakawa's method is a finite element method', *J. Comput. Phys.*, **16**, 383–390 (1974).
4. P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North Holland, Amsterdam, 1978.
5. P. Concus and G. H. Golub, 'Generalized conjugate gradient method for non-symmetric systems of linear equations', *Stanford University Res. Rep.*, January 1976.
6. H. Van der Vorst, 'Iterative solution methods for certain sparse linear-systems with a non symmetric matrix arising from P.D.E. problems', *J. Comput. Phys.*, **44**, 119 (1981).
7. D. A. H. Jacobs, 'Generalizations of the conjugate gradient method for solving non-symmetric and complex systems of algebraic equations', Central Electricity Research Laboratory, 1980.
8. A. Brandt, 'Multi-level adaptative techniques for partial differential equations. Ideas and software', in J. Rice (ed.) *Mathematical Software III*, Academic Press, N.Y. pp. 273, 314.
9. A. Arakawa and V. R. Lamb, 'The U.C.L.A. general circulation model', *Methods in Computational Physics*, **17**, 000–000 (1977).
10. L. Trefethen, 'Group velocity in finite differences schemes', *SIAM, Review*, **24** (2) 113–137 (1982).
11. A. Bamberger, G. Chavent and P. Lailly, 'étude de schemas numériques pour les equations de l'elastodynamique linéaire', *Res. Rep. 41*, INRIA, France, 1980.
12. D. B. Haidvogel, A. R. Robinson and E. E. Schulman, 'The efficiency of open ocean models', *J. Comput. Phys.*, **34** 1, (1980).
13. V. Arnold, *Methodes Mathématiques de la Mécanique*, pp. 333–344.
14. C. Bardos, 'Existence et unicite de la solution de l'equation d'Euler', *J. Math. Anal. Appl.*, **40**, 769–790 (1972).
15. C. Bardos, M. Bercovier and O. Pironneau, 'The vortex method with finite elements', *Math. Comp.*, **36**, 153, (1981).
16. D. G. Ebin and J. Marsden, 'Groups of diffeomorphism and the motion of an incompressible fluid', *Ann. Math.*, 102–163 (1970).
17. G. J. Fix, 'Finite element models for ocean circulation problems', *SIAM J. Appl. Math.*, **29**, (3), (1975).
18. V. Girault and P. A. Raviart, 'Finite element approximations of the Navier–Stokes equations', *Lecture Notes in Math no. 749*, Springer Verlag, Berlin, 1979.
19. T. Kato, 'On classical solutions of the two-dimensional non-stationary Euler equation', *Arch. Rat. Mech. Anal.*, 302–324 (1967).
20. P. Lesaint, 'Sur la resolution des systemes hyperboliques du premier ordre par des methodes d'elements finis', *Thèse Université de Paris VI*, 1975.
21. J. L. Lions, 'Queques methodes de resolution des problemes aux limites non-linéaires', Dunod, Paris, 1969.
22. R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial Value Problems*, Interscience Publishers, Wiley, 1967.
23. R. Temam, 'Local existence of C^∞ solutions of the Euler equations of incompressible perfect fluids. Turbulence and Navier-Stokes Equations', *Orsay 1975, Lectures Notes in Mathematics*, Springer Verlag.