Product Formulas and Numerical Algorithms*

ALEXANDRE J. CHORIN
University of California, Berkeley

THOMAS J. R. HUGHES
California Institute of Technology

MARJORIE F. McCRACKEN
Courant Institute and Indiana University

AND

JERROLD E. MARSDEN
University of California, Berkeley

1. Introduction

Product formulas constitute one of several bridges between numerical and functional analysis. In numerical analysis, they represent algorithms intended to approximate some evolution equation and, in functional analysis, they are used to prove estimates, existence and representation theorems.

Our aim is to *survey* the setting for product formulas and to discuss some recent results. Needless to say, we do not attempt to accommodate all the complex variations which occur in practical algorithms, nor the sharpest possible theoretical results. Nevertheless, we hope that our middle ground approach and some of the examples will be of interest to both groups. Because of its survey nature, we have not hesitated to include some well-known examples which are important for understanding the ideas.

The general idea of product formulas is the following. Suppose one is interested in an initial value problem

$$\frac{du}{dt} = A(u),$$

* The research for this paper was done while the third author was a Visiting Member at the Courant Institute and supported by the National Science Foundation under Grant No. NSF MCS-76-07039. The work of the fourth author was partially supported by the National Science Foundation and based in part on a lecture (by J. E. Marsden) at the second Los Alamos Workshop, August 1976. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Communications on Pure and Applied Mathematics, Vol. XXXI, 205–256 (1978)
 0010-3640/78/0031-0205$01.00

where A is some linear or nonlinear operator on u in some space. Let $K_{\Delta t}$ be a step-forward operator corresponding to an algorithm intended to approximate this equation. If $\Delta t = t/n$ and we iterate n times,

$$K_{t/n}^n = K_{t/n} \circ \cdots \circ K_{t/n}$$

is supposed to approximate the evolution operator for the equation; $K_{\Delta t}$ may be defined on the same space of u's or on an approximating space.

If we let F_t denote the evolution operator for the equation, i.e., $F_t(u_0)$ is the solution with initial value u_0, then convergence of the algorithm may be written as

$$F_t = \lim_{n \to \infty} K_{t/n}^n$$

or as

$$F_t = \lim_{\Delta t \to 0} K_{\Delta t}^{[t/\Delta t]},$$

where $[t/\Delta t]$ is the greatest integer in $t/\Delta t$.

For time-dependent equations $du/dt = A(t, u)$, the evolution operator depends on t and s; $F_{t,s}(u_0)$ is the solution $u(t)$ with $u(s) = u_0$. Here the step-forward operator depends explicitly on time: $K_{\Delta t, t}$, and convergence may be written as

$$F_{t,s} = \lim_{\Delta t \to 0} \prod_{k=0}^{[(t-s)/\Delta t]} K_{\Delta t, s + k\,\Delta t},$$

where the product is ordered with smaller k's to the right. The time dependent case is technically more difficult and will not be discussed further (see Kato [31], [32] for the linear case and Crandall and Pazy [70] and Evans [72] for the nonlinear contractive case).

The history of product formulas is complicated by gaps between theory and practice, but we shall attempt a brief sketch here. The convergence of special algorithms for ordinary differential equations of course goes back to Euler and Picard. The same ideas prove the convergence of general algorithms for ordinary differential equations, as is given in numerical analysis texts (e.g., Gear [21]). We shall discuss this case in Section 2.

For $m \times m$ matrices A and B, the algorithm $K_{\Delta t} = \exp\{(\Delta t)A\} \exp\{(\Delta t)B\}$ for the equation $dx/dt = Ax + Bx$ leads to the 1875 formula of S. Lie [38]:

$$\exp\{A + B\} = \lim_{n \to \infty} (\exp\{A/n\} \exp\{B/n\})^n.$$

This and the related formula

$$\exp\{[A, B]\} = \lim_{n\to\infty} (\exp\{-B/\sqrt{n}\} \exp\{-A/\sqrt{n}\} \exp\{B/\sqrt{n}\} \exp\{A/\sqrt{n}\})$$

occur in the theory of Lie groups.

In the period from 1928–1950, a large number of specific iteration schemes were developed and their convergence established, primarily for linear partial differential equations, with some theory and a lot of practice in the nonlinear case. The bibliography of Richtmyer and Morton [54] contains some of the important references.

In the period 1950–1956 the von Neumann condition and the Lax equivalence theorem for linear systems were developed (see Lax and Richtmyer [37]). The Lax theorem gives conditions assuring convergence of the algorithm. One of these conditions, stability, was examined in detail through spectral properties, as in the Kreiss matrix theorem (again see [54]).

In 1958, Trotter [63] made some important improvements in the Lax theorem and applied this result in [64] to the formula $\exp\{A+B\} = \lim_{n\to\infty} (\exp\{A/n\} \exp\{B/n\})^n$ in case A and B are unbounded operators. A further important extension and simplification was given by Chernoff [5], [7] which will be discussed, along with some examples, in Section 3.

Trotter worked in the generality of approximating Banach spaces, a situation which is important in numerical work (see, e.g., Ciarlet and Raviart [13]); if the equations $\dot{u} = A(u)$ have solutions lying in a Banach space Y, the step forward operator $K_{\Delta t,n}$ may be defined on an approximating Banach space Y_n in this sense (see Kato [30]): there are uniformly bounded operators $P_n : Y \to Y_n$ such that, for $u \in Y$, $\|P_n u\| \to \|u\|$ as $n \to \infty$ and there is a constant C such that, for all $v \in Y_n$, there is a $u \in Y$ with $v = P_n u$ and $\|u\| \leqq C\|v\|$. It is fairly routine to extend the concepts and results from the case $K_{\Delta t} : Y \to Y$ to the approximating case $K_{\Delta t,n} : Y_n \to Y_n$. For example, convergence becomes phrased as follows: for $u \in Y$,

$$\lim_{n\to\infty} \|K^n_{t/n,n} P_n u - P_n F_t u\| = 0 .$$

Applications of product formulas for linear operators to quantum theory and the Feyman–Kac formula were initiated by Nelson [46]. For more recent references and applications, see Reed and Simon [53].

On the nonlinear side, a number of specific algorithms were discussed and methods developed in the context of numerical analysis. The papers of Strang [59] are representative. For contractive nonlinear semigroups, Brezis and Pazy [4] proved convergence of contractive algorithms. In practice, the contractive hypothesis may be difficult to arrange. Of course, many existence

theorems for nonlinear equations implicitly establish convergence of algorithms, as in Lions [40]. Product formulas were used by Ebin–Marsden [16] to establish convergence of the solutions of the Navier–Stokes equations to those of the Euler equations on regions with no boundary as the viscosity goes to zero. The method was abstracted by Marsden [42] and will be briefly discussed in Section 5. Finally, Chernoff [6] established under very general hypotheses that if $F_{t/n}^n$ converges, the limit must be a semiflow.

Although convergence is important, there are many situations, such as structural dynamics, in which one wants to take large time steps in order to mask high frequency modes of lesser interest. In doing so, stability problems are of central importance. Some of the more naive generalizations of linear techniques including an over-reliance on spectral methods seem beset with many difficulties. In Section 7 we shall discuss these ideas and present an energy preserving (implicit) algorithm for nonlinear Hamiltonian systems which generalizes the Crank–Nicolson method for linear systems. The method is also of theoretical interest, for it can be used to establish global weak solutions for a number of systems (e.g., those treated in Segal [56]).

The most complicated example treated here will be an algorithm for the Navier–Stokes equations first implemented by Chorin [9], and based on a heuristic model of boundary layer mechanics; see Lighthill [39] and Batchelor [2]. This algorithm was written as a product formula by Marsden [43]. Convergence of the algorithm and some discussion of its relevance for turbulent solutions of the Navier–Stokes equations is described in Section 6.

Acknowledgement. The authors are grateful to T. K. Caughey, P. R. Chernoff, T. Kato, S. A. Orszag, A. Pazy, and C. Peskin for their invaluable assistance. Some of the results here were done with their collaboration or, where noted, are due to them; they also pointed out a number of useful examples.

2. Ordinary Differential Equations

1. Introduction. To get the idea, consider a linear differential equation in $\mathbb{R}^m$.

$$\frac{dx}{dt} = Ax$$

with solution $x(t) = \exp\{tA\}x(0)$. Let $K_{\Delta t}$ be a linear map of $\mathbb{R}^m$ depending smoothly on Δt with $k_0 =$ identity and satisfying

$$\frac{d}{d\epsilon} K_\epsilon \Big|_{\epsilon=0} = A\,.$$

Then to show that $\lim_{n\to\infty} K_{t/n}^n = \exp\{tA\}$, note that

$$\log K_{t/n}^n = n \log K_{t/n} = t\frac{\log K_{t/n}}{t/n} \to tA$$

so that

$$K_{t/n}^n \to \exp\{tA\}.$$

(Taking logs is justified since K_ϵ is close to the identity for ϵ small.)

Unfortunately, this simple argument does not work for linear partial differential equations, nor in the nonlinear case.

2. A convergence theorem. Let us begin with the observation that the convergence theorem for ordinary differential equations works on Banach manifolds. It is worth recording this well-known proof for comparison with later results and also so that we can globalize the theorem.

First the notation: let X be a Banach manifold and let A be a locally Lipschitz vector field on X. Let F_t be the flow of A, maximally extended, so $F_t(x_0)$ is defined for (t, x_0) in an open set containing $\{0\}\times X$. Thus $F_t(x_0)$ is just the solution of $\dot{x} = A(x)$ with initial condition x_0 which is maximally extended in time (see any text in ordinary differential equations).

Let $K_\epsilon(x)$ be a map defined in some open set of $\mathbb{R}\times X$ containing $\{0\}\times X$ and taking on values in X and assume

(i) $K_0(x) = x$,

(ii) $K_\epsilon(x)$ is C^1 in ϵ with derivative continuous in (ϵ, x).

2.1. THEOREM. *Assume that the algorithm $K_\epsilon(x)$ is consistent with A in the sense that*

$$A(x) = \frac{d}{d\epsilon} K_\epsilon(x)\Big|_{\epsilon=0}.$$

Then, if (t, x) is in the domain of $F_t(x)$, $K_{t/n}^n(x)$ is defined for n sufficiently large and converges to $F_t(x)$. Conversely, if $K_{t/n}^n(x)$ is defined and converges for $0 \leqq t \leqq T$, then (T, x) is in the domain of F and the limit is $F_t(x)$.

Proof: First of all, we prove that convergence holds locally. We begin by showing that, for any x_0, the iterates $K_{t/n}^n(x_0)$ are defined if t is sufficiently small. Indeed, on a neighborhood of x_0, $K_\epsilon(x) = x + O(\epsilon)$; thus if $K_{t/j}^j(x)$ is defined for x in a neighborhood of x_0, $j = 1, \ldots, n-1$,

$$\begin{aligned} K_{t/n}^n(x) - x &= (K_{t/n}^n x - K_{t/n}^{n-1} x) + (K_{t/n}^{n-1} - K_{t/n}^{n-2} x) \\ &\quad + \cdots + (K_{t/n}(x) - x) \\ &= O(t/n) + \cdots + O(t/n) = O(t). \end{aligned}$$

This is small independent of n for t sufficiently small; so that, inductively, $K^n_{t/n}(x)$ is defined and remains in a neighborhood of x_0, for x near x_0.

Let β be a local Lipschitz constant for A such that $\|F_t(x)-F_t(y)\| \leqq \exp\{\beta|t|\}\|x-y\|$. Now write

$$\begin{aligned} F_t(x)-K^n_{t/n}(x) &= F^n_{t/n}(x)-K^n_{t/n}(x) \\ &= F^{n-1}_{t/n}F_{t/n}(x)-F^{n-1}_{t/n}K_{t/n}(x) \\ &\quad + F^{n-2}_{t/n}F_{t/n}(y_1)-F^{n-2}_{t/n}K_{t/n}(y_1) \\ &\quad + \cdots + F^{n-k}_{t/n}F_{t/n}(y_{k-1})-F^{n-k}_{t/n}K_{t/n}(y_{k-1}) \\ &\quad + \cdots + F_{t/n}(y_{n-1})-K_{t/n}(y_{n-1}), \end{aligned}$$

where $y_k = K^k_{t/n}(x)$. Thus

$$\begin{aligned} \|F_t(x)-K^n_{t/n}(x)\| &\leqq \sum_{k=1}^{n} \exp\left\{\frac{\beta|t|(n-k)}{n}\right\}\|F_{t/n}(y_{n-k-1})-K_{t/n}(y_{n-k-1})\| \\ &\leqq n\exp\{\beta|t|\}o(t/n)\to 0 \quad \text{as} \quad n\to\infty, \end{aligned}$$

since $F_\epsilon(y)-K_\epsilon(y)=o(\epsilon)$ by the consistency hypothesis.

Now suppose $F_t(x)$ is defined for $0\leqq t\leqq T$. We shall show that $K^n_{t/n}(x)$ converges to $F_t(x)$. By the above proof and compactness, if N is large enough, $F_{t/N}=\lim_{n\to\infty}K^n_{t/nN}$ uniformly on a neighborhood of the curve $t\to F_t(x)$. Hence, for $0\leqq t\leqq T$,

$$F_t(x)=F^N_{t/N}(x)=\lim_{n\to\infty}(K^n_{t/nN})^N(x).$$

By uniformity in t,

$$F_T(x)=\lim_{j\to\infty}K^j_{T/j}(x).$$

Conversely, let $K^n_{t/n}(x)$ converge to a curve $c(t)$, $0\leqq t\leqq T$. Let $S=\{t \mid F_t(x)$ is defined and $c(t)=F_t(x)\}$. From the local result, S is a nonempty open set. Let $t_k\in S$, $t_k\to t$. Thus $F_{t_k}(x)$ converge to $c(t)$; thus, by local existence theory, $F_t(x)$ is defined and, by continuity, $F_t(x)=x(t)$. Hence $S=[0,T]$ and the proof is complete.

3. Accuracy and some examples. The explicit Euler method merely chooses $K_\epsilon(x)=x+\epsilon A(x)$. For higher order accuracy, one uses the following consequence of the preceding proof.

2.2. COROLLARY. *Let $k \geqq 2$ and suppose $K_t(x)$ is C^{k+1} in t and A is of class C^{k+1}. If*

$$\frac{d^i}{(dt)^i} K_t(x)|_{t=0} = \frac{d^i}{(dt)^i} F_t(x)|_{t=0}, \qquad i = 1, 2, 3, \ldots, k,$$

in addition to the hypotheses of Theorem 2.1, *then the algorithm is k-th order accurate; i.e.,*

$$\|F_t(x) - K_{t/n}^n(x)\| \leqq \exp\{\beta |t|\} O((\Delta t)^k),$$

$\Delta t = t/n$, *as* $n \to \infty$.

Specifically, if

$$\frac{d^2}{dt^2} K_t(x)|_{t=0} = DA(x) \cdot A(x),$$

then the algorithm is second order accurate.

2.3. EXAMPLE. It is often convenient to use the notation

$$x_{n+1} = K_{\Delta t}(x_n), \qquad t = n\,\Delta t,$$

for the step-forward operator. Thus if we set

$$x_1 = K_{\Delta t}(x), \qquad x_2 = K_{\Delta t}(x_1), \ \cdots, \qquad x_n = K_{\Delta t}(x_{n-1}),$$

we have

$$x_n = K_{\Delta t}^n(x).$$

Consider an algorithm given implicitly by the form

$$x_{n+1} = x_n + \Delta t J(x_{n+1}, x_n, \Delta t),$$

with

$$A(x) = J(x, x, 0).$$

(So the step forward operator K_ϵ satisfies $K_\epsilon(x) = x + \epsilon J(K_\epsilon(x), x, \epsilon)$.) The condition that this be second order accurate is easy to work out and is

$$D_1 J(x, x, 0) \cdot J(x, x, 0) + 2D_3 J(x, x, 0) = D_2 J(x, x, 0)\, J(x, x, 0),$$

where D_i is the derivative with respect to the i-th factor. (Writing $J(x, y, \epsilon)$, this says that $(\partial J/\partial x) \cdot J + 2(\partial J/\partial \epsilon) = (\partial J/\partial y) \cdot J$ at $x = y, \quad \epsilon = 0$.)

The following are examples of second order accurate schemes for given $A(x)$:

Crank–Nicolson (Trapezoidal): $J(x, y, \epsilon) = \frac{1}{2}[A(x) + A(y)]$.

Mid-point: $J(x, y, \epsilon) = A(\frac{1}{2}(x+y))$.

Predictor–Corrector: $J(x, x, \epsilon) = \frac{1}{2}[A(x) + A(y + \epsilon A(y))]$.

Central-Difference (for $\ddot{x} = A(x)$): $J((x, \dot{x}), (y, \dot{y}), \epsilon) = (\dot{y} + \frac{1}{2}\epsilon A(y), \frac{1}{2}(A(x) + A(y))$, etc.

4. Application to the Euler equations. The results of this subsection can actually be applied to give some interesting product formulas for the Euler equations of a perfect fluid. On a compact manifold M, with C^∞ boundary, ∂M, and dimension n, these equations are

$$\frac{\partial u}{\partial t} + \nabla_u u = -\operatorname{grad} p\,, \tag{4.1}$$

$$\operatorname{div} u = 0\,,$$

$$u \text{ a vector field on } M \text{ parallel to } \partial M.$$

2.4. THEOREM. *In Sobolev function spaces H^s, $s > \frac{1}{2}n + 1$ (or $W^{s,p}$, $s > n/p + 1$) or $C^{k+\alpha}$, $k \geqq 1$, $0 < \alpha < 1$, let $G_{\Delta t}$ be the step-forward operator defined by*[1] $\partial u/\partial t + \nabla_u u = 0$.

Let $\mathbb{P}$ denote the L_2 orthogonal projection of vector fields to their divergence-free parts parallel to ∂M.

Then the solution $E_t(u)$ of the Euler equations, with initial data u is given by

$$E_t = \lim_{n \to \infty} (\mathbb{P} G_{t/n})^n\,. \tag{4.2}$$

This product formula converges for $0 \leqq t \leqq T$ and u fixed if and only if $E_t(u)$ exists (in the above function spaces), $0 \leqq t \leqq T$.

A proof of this product formula is due to Chorin [8], although the proof is somewhat complicated. The following proof (Marsden–Ebin–Fischer [45]) is more elementary in the sense that it relies only on Theorem 2.1, but more complicated in the sense that it relies on the results of Ebin–Marsden [16].

Proof of Theorem 2.4: (Sketch) Some key results of Ebin–Marsden are as follows: When transformed to Lagrangian coordinates (i.e., to the

[1] $G_{\Delta t}(u)$ is given as follows: let $\eta_u(x)$ be the geodesic starting at x in direction $u(x)$ (in Euclidean space $\eta_u(x) = x + \epsilon u(x)$), then $G_{\Delta t}(u) = u \circ \eta_{\Delta t}^{-1}$.

group of volume preserving diffeomorphisms of M), the vector fields generating E_t and G_t are C^∞ maps (with no loss of derivatives) and $\mathbb{P}$ transforms to a C^∞ bundle mapping. If we choose $K_t = \mathbb{P}G_t$ and transform to Lagrangian coordinates, the hypotheses of Theorem 2.1 are met, so convergence in Lagrangian coordinates follows. Since transformation back to Eulerian coordinates is a C^0 operator, convergence holds there as well.

The product formula in Theorem 2.4 may be useful in investigating the development of a singularity in the Taylor–Green problem (cf. [62]). This problem concerns the evolution of the following initial data on the three-torus under the Euler equations (or the Navier–Stokes equations):

$$\begin{aligned} v_1 &= \cos x_1 \sin x_2 \cos x_3 , \\ v_2 &= -\sin x_1 \cos x_2 \cos x_3 , \\ v_3 &= 0 . \end{aligned} \tag{4.3}$$

The flow determined by this initial data undergoes vorticity enhancement by a self-induced tornado type mechanism. There is considerable unpublished numerical work on this problem (Keller, Chorin, $\cdots$). This and work of Orszag ([51] and personal communication) indicates that

$$\Omega(t) = \int_{T^3} |\omega|^2 \, d\mu ,$$

where $\omega = \nabla \times v$ is the vorticity, blows up in a finite $t \geq 3.1$.

One would like to conclude, via Theorem 2.4 or a modification of it, that the strict Euler equations have a finite blow-up time.[2] In drawing such conclusions one must be careful of numerical blow-ups. For instance, in Herring, *et al.* [23], the accuracy of finite mode simulation of high Reynolds number flows is questioned (see also Chorin [8]). In this particular problem these objections do not seem relevant, for $t \leq 4$. (Orszag, private communication.)

Further computations are under way to confirm or deny the existence of a singularity and to detail its structure. Hopefully they will lead to a mathematical proof of finite blow-up time for the three-dimensional Euler equations. (For more details on the blow-up problem for the Euler and Navier–Stokes equations, see [45].)

[2] The algorithm of Theorem 2.4 is not strictly used in practice, but the modifications used (e.g., a leapfrog time differencing scheme on the nonlinear terms) for numerical stability are hopefully minor.

3. The Generalized Lax Equivalence Theorem

1. Some terminology. This section deals with product formulas in the *linear* case.

Let X be a Banach space and let A be the generator of a linear C^0 semigroup on X, $\exp\{tA\}$. Thus, if u lies in the domain of A, $\exp\{tA\}u = u(t)$ solves the equation

$$\dot{u} = Au.$$

As is well known, $\|\exp\{tA\}\| \leqq M \exp\{\beta t\}$ for constants M, β. We write $A \in G(M, \beta)$.

Let, for each $\epsilon \geqq 0$, $K_\epsilon : X \to X$ be a bounded linear operator and assume $K_0 = id$.

The algorithm K_ϵ is called *stable*[3] if, for any $T > 0$, there is a constant M_T such that

$$\|K_{t/n}^n\| \leqq M_T \tag{1.1}$$

for all $0 \leqq t \leqq T$, $n = 0, 1, 2, \cdots$.

The algorithm is called *consistent* if, for each x in the domain of A.

$$\frac{d}{d\epsilon} K_\epsilon(x)|_{\epsilon=0} = A(x). \tag{1.2}$$

The classical Lax theorem, subsequently improved by Trotter [64], states basically that under the assumption of consistency (along curves $\exp\{tA\}u = u(t)$) stability is equivalent to convergence. (See Richtmyer and Morton [54], or Lusternik and Sobolev [41] for a discussion and proof.)

The algorithm is *resolvent consistent* if, for $\lambda > 0$ sufficiently large,

$$(\lambda - A)^{-1} = \text{s-}\lim_{\epsilon \to 0} \left[\lambda - \frac{1}{\epsilon}(K_\epsilon - I)\right]^{-1} \tag{1.3}$$

(strong limit).

2. The generalized Lax theorem. For contractive semigroups and contractive algorithms, Chernoff [6] proved that resolvent consistency is equivalent to convergence. The general situation is as follows:

3.1. THEOREM. *Let A generate a C^0 semigroup on X (not necessarily contractive, bounded or quasi-contractive) and let $K_\epsilon \in B(X)$ be a given family*

[3] The terminology here is not quite consistent with that in Section 6, where stability is taken to mean uniform stability, i.e., that M_T is independent of T.

of bounded operators. $K_0 = Id$. *Then*

$$\left(\begin{array}{c} K_t \text{ is stable and} \\ \text{resolvent consistent} \end{array}\right) \Leftrightarrow \left(\begin{array}{c} \exp\{tA\} = \text{s-}\lim_{n\to\infty} K_{t/n}^n, \\ \text{uniformly on bounded } t\text{-intervals} \end{array}\right).$$

Although this can be derived from results in Kato [30], we shall give a self-contained proof kindly supplied by P. Chernoff.

Proof of Theorem 3.1:

Step 1. We first establish for technical convenience the equivalence of the following two versions of convergence:

$$\exp\{tA\} = \text{s-}\lim_{\epsilon\to 0} K(\epsilon)^{[t/\epsilon]} \quad \text{uniformly}, \qquad 0 \leqq t \leqq T, \tag{2.1}$$

$$\exp\{tA\} = \text{s-}\lim_{n\to\infty} K(t/n)^n \quad \text{uniformly}, \qquad 0 \leqq t \leqq T. \tag{2.2}$$

First suppose that (2.2) is false. Then there exists a vector x, a $\delta > 0$, a sequence t_k in $[0, T]$, and integers $n_k \to \infty$ such that, for all k,

$$\|K(t_k/n_k)^{n_k} x - \exp\{t_k A\}x\| \geqq \delta. \tag{2.3}$$

Let $\epsilon_k = t_k/n_k$; note that $\epsilon_k \to 0$. Also $K(t_k/n_k)^{n_k} = K(\epsilon_k)^{[t_k/\epsilon_k]}$, so that by (a) it follows that, for k sufficiently large,

$$\|K(\epsilon_k)^{[t_k/\epsilon_k]} x - \exp\{t_k A\}x\| \leqq \tfrac{1}{2}\delta.$$

This contradicts (2.3). Thus (2.1) implies (2.2).

Secondly, suppose that (2.1) is false. Then there is a vector x, a $\delta > 0$, and a sequence $\epsilon_k \to 0$, such that, for all k,

$$\|K(\epsilon_k)^{[t_k/\epsilon_k]} x - \exp\{t_k A\}x\| \geqq \delta. \tag{2.4}$$

This leads to an obvious contradiction if the numbers $[t_k/\epsilon_k]$ are bounded; thus, passing to a subsequence, we may assume that $[t_k/\epsilon_k] = n_k \to \infty$. Note that $\epsilon_k = \tau_k/n_k$, where $\tau_k - t_k \to 0$ as $k \to \infty$.

Then $K(\epsilon_k)^{[t_k/\epsilon_k]} x = K(\tau_k/n_k)^{n_k} x$, and, for large k, (2.2) implies the inequality

$$\|K(\tau_k/n_k)^{n_k} x - \exp\{\tau_k A\}x\| \leqq \tfrac{1}{2}\delta.$$

Note also that $\|\exp\{\tau_k A\}x - \exp\{t_k A\}x\| \leqq \frac{1}{4}\delta$ for large k by uniform continuity of $\exp\{tA\}x$ on $[0, T]$. Thus we have a contradiction of (2.4).

Step 2. *convergence implies stability:* Since $\exp\{tA\}$ is a C^0 semigroup we have an estimate of the form $\|\exp\{tA\}\| \leqq M_0 \exp\{\beta_0 t\}$. Suppose that $K(\epsilon)^{[t/\epsilon]}x \to \exp\{tA\}x$ uniformly, $0 \leqq t \leqq 1$, as $\epsilon \to 0$ for all x. Then there is a constant M such that, for all $\epsilon \in [0, 1]$ and all $t \in [0, 1]$, $\|K(\epsilon)^{[t/\epsilon]}\| \leqq M$.

Proof: We use the Banach–Steinhaus theorem. Suppose that, for some x, $\|K(\epsilon_n)^{[t_n/\epsilon_n]}x\|$ is not bounded. Clearly ϵ_n must tend to 0. We may assume that $t_n \to t \in [0, 1]$. But then, by assumption, $K(\epsilon_n)^{[t_n/\epsilon_n]}x \to \exp\{tA\}x$, so the norms must be bounded.

Next we prove that if $\text{s-lim}_{\epsilon \to 0} K(\epsilon)^{[t/\epsilon]} = \exp\{tA\}$ locally uniformly, we have an estimate

$$\|K(\epsilon)^{[t/\epsilon]}\| \leqq M \exp\{\beta t\}$$

for some constants M, β.

Proof: We have $\|K(\epsilon)^{[t/\epsilon]}\| \leqq M$ for $0 \leqq \epsilon \leqq 1$, $0 \leqq t \leqq 1$. Now suppose that $0 \leqq t \leqq n$. Write $t = t_1 + t_2 + \cdots + t_n$, $0 \leqq t_i \leqq 1$, and note that $[t/\epsilon] = [t_1/\epsilon] + \cdots + [t_n/\epsilon] + r$, where $0 \leqq r \leqq n$. Hence

$$\|K(\epsilon)^{[t/\epsilon]}\| \leqq \|K(\epsilon)^{[t_1/\epsilon]}\| \cdots \|K(\epsilon)^{[t_n/\epsilon]}\| \|K(\epsilon)^r\|$$

$$\leqq M^n \cdot M^r \leqq M^{2n}$$

if $n - 1 \leqq t \leqq n$, $2n \leqq 2t + 2$ so that

$$\|K(\epsilon)^{[t/\epsilon]}\| \leqq M^2 M^{2t} = M^2 \exp\{2t \log M\},$$

an estimate of the desired form.

Step 3. *Convergence implies stability and resolvent consistency:* If $\exp\{tA\} = \text{s-lim}_{\epsilon \to 0} K(\epsilon)^{[t/\epsilon]}$ uniformly, $0 \leqq t \leqq T$, then $K(\epsilon)$ is stable and resolvent consistent.

Proof: We already have proved in Step 2 that $K(\epsilon)$ is stable: $\|K(\epsilon)^n\| \leqq M \exp\{n\beta\epsilon\}$ for some constants M, β. Consider the resolvents (which we show make sense for $\lambda > (\exp\{\beta\epsilon\} - 1)/\epsilon$):

$$[\lambda - \epsilon^{-1}[K(\epsilon) - I]]^{-1} = \epsilon[1 + \lambda\epsilon - K(\epsilon)]^{-1}$$

$$= \frac{\epsilon}{1 + \lambda\epsilon}\left(1 - \frac{K(\epsilon)}{1 + \lambda\epsilon}\right)^{-1}$$

$$= \frac{\epsilon}{1 + \lambda\epsilon} \sum_{n=0}^{\infty} (1 + \lambda\epsilon)^{-n} K(\epsilon)^n.$$

Since $\|K(\epsilon)^n\| \leqq M \exp\{n\beta\epsilon\}$, the series converges for $1+\lambda\epsilon > \exp\{\beta\epsilon\}$, i.e., $\lambda > (\exp\{\beta\epsilon\}-1)/\epsilon$. The second series is equal to the integral

$$\frac{1}{1+\lambda\epsilon}\int_0^\infty (1+\lambda\epsilon)^{-[t/\epsilon]} K(\epsilon)^{[t/\epsilon]}\, dt$$

and this converges strongly, by the dominated convergence theorem, to

$$\int_0^\infty \exp\{-\lambda t\}\exp\{tA\}\, dt = [\lambda - A]^{-1}.$$

Thus we have proved resolvent consistency.

Step 4. *A key estimate:* Assume stability, i.e., $\|K(\epsilon)^{[t/\epsilon]}\| \leqq M \exp\{\beta t\}$. If $t/\epsilon = n$, we have $\|K(\epsilon)^n\| \leqq M \exp\{\beta\epsilon^n\}$. We now wish to estimate

$$[\exp\{n(K(\epsilon)-I)\} - K(\epsilon)^n]x = \sum_{k=0}^{\infty} \frac{\exp\{-n\}n^k}{k!}(K(\epsilon)^k x - K(\epsilon)^n x).$$

Consider an individual term. If $n > k$,

$$K(\epsilon)^n - K(\epsilon)^k = \{K(\epsilon)^{n-1} + K(\epsilon)^{n-2} + \cdots + K(\epsilon)^k\}(K(\epsilon)-I),$$

we estimate as follows:

$$\left\| \sum_{r=k}^{n-1} K(\epsilon)^r \right\| \leqq \sum_{r=k}^{n-1} M \exp\{r\beta\epsilon\} = M\left(\frac{\exp\{n\beta\epsilon\} - \exp\{k\beta\epsilon\}}{\exp\{\beta\epsilon\}-1}\right).$$

Hence

$$\|[\exp\{n(K(\epsilon)-I)\} - K(\epsilon)^n]x\| \leqq \frac{M}{\exp\{\beta\epsilon\}-1} \sum_{k=0}^{\infty} \frac{\exp\{-n\}n^k}{k!}(\exp\{n\beta\epsilon\} - \exp\{k\beta\epsilon\})\|K(\epsilon)x - x\|.$$

We estimate the series by the Schwarz inequality:

$$\sum_{k=0}^{\infty} \frac{\exp\{-n\}n^k}{k!}(\exp\{n\beta\epsilon\} - \exp\{k\beta\epsilon\}) \leqq \left\{ \sum_{k=0}^{\infty} \frac{\exp\{-n\}n^k}{k!}(\exp\{2n\beta\epsilon\} - 2\exp\{n\beta\epsilon\}\exp\{k\beta\epsilon\} + \exp\{2k\beta\epsilon\}) \right\}^{1/2}.$$

This series can be summed explicitly: it equals

$$\{\exp\{2n\beta\epsilon\}-2\exp\{n\beta\epsilon\}\exp\{-n\}\exp\{n\exp\{\beta\epsilon\}\} + \exp\{-n\}\exp\{n\exp\{2\beta\epsilon\}\}\}^{1/2}$$

$$=\exp\{n\beta\epsilon\}\{1-2\exp\{n(\exp\{\beta\epsilon\}-1-\beta\epsilon)\}+\exp\{n(\exp\{2\beta\epsilon\}-1-2\beta\epsilon)\}\}^{1/2}$$

$$=\exp\{n\beta\epsilon\}\{1-2\exp\{\tfrac{1}{2}n\beta^2\epsilon^2+nO(\beta^3\epsilon^3)\}+\exp\{n\cdot\tfrac{1}{2}4\beta^2\epsilon^2+nO(\beta^3\epsilon^3)\}\}^{1/2}$$

$$=\exp\{\beta\epsilon\}\{n\beta^2\epsilon^2+O(n\beta^3\epsilon^3)+O(n^2\beta^4\epsilon^4)\}^{1/2}.$$

Now take $n=[t/\epsilon]$. Then $n\beta^3\epsilon^3=O(\epsilon^2\beta^3 t)$ and $n^2\beta^4\epsilon^4=O(\epsilon^2\beta^4t^2)$, so the expression in brackets is $\beta^2\epsilon^2[t/\epsilon]+O(\epsilon^2)$ and thus

$$\sum_{k=0}^{\infty}\frac{\exp\{-n\}n^k}{k!}(\exp\{n\beta\epsilon\}-\exp\{k\beta\epsilon\})\leqq\exp\{\beta\epsilon[t/\epsilon]\}\cdot\beta\epsilon[t/\epsilon]^{1/2}+O(\epsilon^2)$$

uniformly $0\leqq t\leqq T$, $n=[t/\epsilon]$. Consequently we get our estimate:

$$\|[\exp\{[t/\epsilon](K(\epsilon)-I)\}-K(\epsilon)^{[t/\epsilon]}]x\| \leqq\left\{\frac{M\beta\epsilon\exp\{\beta\epsilon[t/\epsilon]}{\exp\{\beta\epsilon\}-1}[t/\epsilon]^{1/2}+O(\epsilon^2)\right\}\cdot\{\|K(\epsilon)x-x\|\}.$$

Step 5. Suppose that $K(\epsilon)$ is stable and resolvent consistent. Then $K(\epsilon)^{[t/\epsilon]}\to\exp\{tA\}$ (strongly, uniformly on $0\leqq t\leqq T$).

Proof: We consider the semigroups $\exp\{t\epsilon^{-1}\}[k(\epsilon)-I]$. We must show that they obey uniform estimates. We start with the estimate $\|K(\epsilon)^n\|\leqq M\exp\{\beta\epsilon n\}$. We then have

$$\|\exp\{t\epsilon^{-1}(K(\epsilon)-I)\}\|\leqq\exp\{-t/\epsilon\}\sum_{n=0}^{\infty}\left\|\frac{1}{n!}\left(\frac{t}{\epsilon}\right)^n K(\epsilon)^n\right\|$$

$$\leqq\exp\{-t/\epsilon\}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{t}{\epsilon}\right)^n\cdot M\exp\{n\beta\epsilon\}$$

$$=M\exp\left\{\frac{t}{\epsilon}(\exp\{\beta\epsilon\}-1)\right\}.$$

Now $\epsilon^{-1}(\exp\{\beta\epsilon\}-1)$ is bounded by some constant as long as $0<\epsilon\leqq 1$. Thus the semigroups $\exp\{t\epsilon^{-1}(K(\epsilon)-I)\}$ belong to a fixed class $G(M,\beta')$. By assumption we have resolvent convergence:

$$[\lambda-\epsilon^{-1}(K(\epsilon)-I)]^{-1}\to[\lambda-A]^{-1};$$

hence it follows that $\exp\{t\epsilon^{-1}(K(\epsilon)-I)\} \to \exp\{tA\}$ in the strong topology, uniformly on bounded t-intervals by the Trotter-Kato theorem ([30, page 502).

To complete the proof we just have to show that $K(\epsilon)^{[t/\epsilon]} - \exp\{t\epsilon^{-1}(K(\epsilon)-I)\} \to 0$ as $\epsilon \to 0$. Since these operators obey uniform estimates, it is enough to get convergence to 0 on a dense set. Fix a vector y. Define

$$x_\epsilon = [\lambda - \epsilon^{-1}(K(\epsilon)-I]^{-1} y .$$

Then

$$\epsilon^{-1}(K(\epsilon)-I)x_\epsilon = \lambda x_\epsilon - y ,$$

which converges to $\lambda(\lambda-A)^{-1}y - y = z$ as $\epsilon \to 0$. Now we apply our "key estimate" (Step 4):

$$\|[\exp\{[t/\epsilon](K(\epsilon)-I)\} - K(\epsilon)^{[t/\epsilon]}]x_\epsilon\| \leqq [\text{Const. } \epsilon^{-1/2} + O(\epsilon)^2)]\|K(\epsilon)x_\epsilon - x_\epsilon\|$$

uniformly on bounded t-intervals. Since $K(\epsilon)x_\epsilon - x_\epsilon = O(\epsilon)$ it follows that the right side goes to 0 uniformly as $\epsilon \to 0$, $0 \leqq t \leqq T$. Finally, note that $x_\epsilon \to x = (\lambda - A)^{-1}y = w$ and so our uniform bounds show that

$$\|[\exp\{[t/\epsilon](K(\epsilon)-I)\} - K(\epsilon)^{[t/\epsilon]}]w\| \to 0 \quad \text{uniformly}, \quad 0 \leqq t \leqq T.$$

But the set of such vectors w is dense.

We recover the Lax theorem as follows.

3.2. COROLLARY. *If K_ϵ is consistent, then it is stable if and only if it is convergent. Here for consistency it is enough that*

$$\frac{d}{d\epsilon} K(\epsilon)|_{\epsilon=0} x = Ax$$

for all x in a core D of A (i.e., a space D such that A is the closure of its restriction to D).

Proof: Necessity is immediate from Theorem 3.1. For sufficiency, write Δ_ϵ for $\epsilon^{-1}(K(\epsilon)-I)$. We must show that $(\lambda - \Delta_\epsilon)^{-1} \to (\lambda - A)^{-1}$ strongly, and by uniform bounds it is enough to do this on a dense set of vectors. We choose as our dense set the vectors of the form $y = (\lambda - A)x$ for x in the

core D. We then write

$$\begin{aligned}(\lambda-\Delta_\epsilon)^{-1}y-(\lambda-A)^{-1}y&=(\lambda-\Delta_\epsilon)^{-1}(\lambda-A)x-x\\&=(\lambda-\Delta_\epsilon)^{-1}(\lambda-\Delta_\epsilon+\Delta_\epsilon-A)x-x\\&=x+(\lambda-\Delta_\epsilon)^{-1}(\Delta_\epsilon-A)x-x\\&=(\lambda-\Delta_\epsilon)^{-1}(\Delta_\epsilon-A)x,\end{aligned}$$

which goes to 0 as $\epsilon\to 0$ since $\Delta_\epsilon x\to Ax$ while the norms of $(\lambda-\Delta_\epsilon)^{-1}$ remain bounded.

3.3. *Remark.* If consistency (on the domain of A) and stability hold, one can prove Corollary 3.2 directly by the same method used to prove Theorem 2.1. This alternative procedure (which is close to the original Lax proof) has a crucial advantage, namely

(a) for consistency, convergence is at the rate Δt, i.e., $1/n$,

(b) for resolvent consistency, convergence is at the rate $\sqrt{\Delta t}$, i.e., $1/\sqrt{n}$.

3. Some examples and applications. The difference pointed out in Remark 3.3 is often of computational significance. We give an example in Section 4, where (b) but not (a) holds and in which convergence is hard to detect on the computer (see Section 4). For now we give a simpler example which is standard in singular perturbations and boundary layer theory.

3.4. EXAMPLE. Let

$$A_\epsilon=\epsilon\frac{d^2}{dx^2}+\frac{d}{dx}\quad\text{in}\quad L_2([0,\infty])$$

with boundary condition $u(0)=0$. Let $K_\epsilon=\exp\{\epsilon A_\epsilon\}$ and let $A=d/dx$ with no boundary condition. In this example, resolvent consistency, but not consistency holds. Convergence of $K_{t/n}^n$ to $\exp\{tA\}$ in L_2 can in fact be seen directly from the Trotter–Kato theorem ([30], page 502). One has $(\lambda-A_\epsilon)^{-1}\to(\lambda-A)^{-1}$ in $L_2([0,\infty])$. This can be seen either using general techniques (see Section 4) or by direct computation. In fact, if $\mu_\pm=(-1\pm\sqrt{1+4\lambda\epsilon})/2\epsilon$, then

$$[(\lambda-A_\epsilon)^{-1}u](x)=\alpha\exp\{x\mu_-\}+\int_0^\infty\frac{\mu_+}{\lambda}\exp\{-s\mu_+\}u(s+x)\,ds,$$

where

$$\alpha = -\int_0^\infty \frac{\mu_+}{\lambda} \exp\{-s\mu_+\} u(s)\, ds\,;$$

this is easily seen to converge in L_2 to

$$[(\lambda - A)^{-1} u](x) = \int_0^\infty \exp\{-\lambda s\} u(s+x)\, ds\,.$$

In all but the simplest examples such explicit computation of the resolvents is difficult or impossible. Some further examples are given in [30].

As we shall see in Section 4, sometimes one can turn a resolvent consistent algorithm into a consistent one and thereby improve the convergence rate. Other pathologies can concievably happen, such as consistency and instability, but convergence on a dense set. If one expects an algorithm to converge only for very smooth functions, Theorem 3.1 or Corollary 3.2 might be applied to a power of A.

There are a number of applications of Theorem 3.1 and Corollary 3.2. The best known of these is the following result of Trotter [64].

3.5. Application. Suppose A and B are generators of quasi-contractive semi-groups (i.e., $\|\exp\{tA\}\| \leqq \exp\{t\alpha\}$, $\|\exp\{tB\}\| \leqq \exp\{t\beta\}$) and the closure of $A+B$, $C = \overline{A+B}$, is a C^0 generator. Then

$$\exp\{tC\} = \operatorname*{s-lim}_{n\to\infty} (\exp\{tA/n\} \exp\{tB/n\})^n\,.$$

Proof: Let $K_\epsilon = \exp\{\epsilon A\} \exp\{\epsilon B\}$. Then $\|K_{t/n}^n\| \leqq \exp\{\epsilon(\alpha+\beta)\}$, so that the algorithm is stable. Also, if $u \in D(A) \cap D(B)$,

$$\frac{1}{\epsilon}(K_\epsilon u - u) = \frac{1}{\epsilon} \exp\{\epsilon A\}(\exp\{\epsilon B\} u - u) + \frac{1}{\epsilon}(\exp\{\epsilon A\} u - u)$$

$$\to Bu + Au \quad \text{as} \quad \epsilon \to 0\,.$$

Since we have consistency on a core for C, the result follows by Corollary 3.2.

See Nelson [47] and Goldstein [22] for the corresponding formula for the semigroup of a bracket.

As Trotter points out, it is not always possible to renorm a space so that two semigroups simultaneously become quasi-contractive. This is possible for

one semigroup $\exp\{tA\}$ by this well-known trick of Feller: if $\|\exp\{tA\}\| \leqq M \exp\{t\beta\}$, let

$$|||x||| = \sup_{t \geqq 0} \|\exp\{-t\beta\} \exp\{tA\}x\| .$$

One might guess that a simultaneous norm in which $\exp\{tA\}$ and $\exp\{tB\}$ are quasi-contractive is necessary. The following example of Chernoff shows this is not so (see [64] and [7] for other examples of pathologies that can occur).

3.6. EXAMPLE. There are two (C_0) semigroups U_t, V_t on Hilbert space $\mathbb{H}$ such that:

(a) U_t, V_t are bounded (—in fact U_t is an isometry);

(b) if A, B are the generators of U_t and V_t, then $\overline{A+B} = C$ is a generator, and moreover the Trotter product formula

$$(U_{t/n} V_{t/n})^n \xrightarrow{s} \exp\{tC\} \quad \text{is valid;}$$

(c) but nevertheless $\mathbb{H}$ *cannot* be renormed in such a way that U_t, V_t are both quasi-contractive.

For the example: Let T_t be the translation group on $\mathbb{H} = L^2(-\infty, \infty)$: $T_t f(x) = f(x+t)$. Let $U_t = T_{-t}$. Then U_t is an isometry. Let $\sigma(x) = \sin(x^2)$. Let

$$V_t = \exp\{\sigma(x+2t) - \sigma(x)\} T_{2t} = R^{-1} T_{2t} R ,$$

where R is multiplication by $\exp\{\sigma(x)\}$.

Then V_t is a uniformly bounded C_0 semigroup. (In fact if $t \neq 0$, $\|V_t\| = e^2$.) We shall show first that $\mathbb{H}$ cannot be renormed in such a way that both U_t, V_t are quasi-contractive. Indeed, if this were possible, we would have constants M, α such that, in the original norm,

$$\|V_{s_1} U_{t_1} V_{s_2} U_{t_2} \cdots V_{s_n} U_{t_n}\| \leqq M \exp\{\alpha(\textstyle\sum s_i + \sum t_i)\}$$

for any s_i, $t_i \geqq 0$. Indeed, we shall show that, for fixed t, no inequality of the form $\|(V_{t/2n} U_{t/n})^n\| \leqq M$ is possible. For this we compute

$$V_{t/2n} U_{t/n} f(x) = \left(\exp\left\{\sigma\left(x + \frac{t}{n}\right) - \sigma(x)\right\}\right) f(x) \Rightarrow (V_{t/2n} U_{t/n})^n f(x)$$
$$= \exp\left\{n\left[\sigma\left(x + \frac{t}{n}\right) - \sigma(x)\right]\right\} f(x) .$$

Hence

$$\log \|(V_{t/2n}U_{t/n})^n\| = \sup_x \{n[\sigma(x+t/n)-\sigma(x)]\}.$$

Since $n[\sigma(x+t/n)-\sigma(x)] \to t\sigma'(x)$ as $n \to \infty$, and $\sigma'(x)$ is unbounded above, there is *no* uniform estimate for $\log \|(V_{t/2n}U_{t/n})^n\|$.

Next, we shall show that $(V_{t/n}U_{t/n})^n$ converges strongly, uniformly on bounded t intervals to a semigroup W_t, and that the generator C of W_t is the closure of $A+B$. (Note that Theorem 3.1 applies here.) We find by computation that

$$(V_{t/n}U_{t/n})^n f(x) = \exp\left\{\sigma(x+t)+\sigma\left(x+t+\frac{t}{n}\right)-\sigma(x)-\sigma\left(x+\frac{t}{n}\right)\right\} f(x+t).$$

Since $\sigma(x)$ is bounded and continuous, the dominated convergence theorem shows that

$$(V_{t/n}U_{t/n})^n f(x) \to \exp\{2\sigma(x+t)-2\sigma(x)\} f(x+t)$$

$$= W_t f(x) \quad \text{in} \quad L^2 \text{ norm},$$

where $W_t = R^{-2}T_t R^2$. The generator of U_t is $A = -d/dx$ on its usual domain $H^1(\mathbf{R})$. Since $V_t = R^{-1}T_{2t}R$, the generator B of V_t is $B = -R^{-1} \cdot 2A \cdot R = 2(d/dx)+2\sigma'(n)$ with $D(B) = R^{-1}D(A)$ and the generator C of W_t is $R^{-2} \cdot (d/dx)R^2 = d/dx + 2\sigma'(x)$. Since each of V_t, U_t and W_t leaves C_0^∞ invariant, C_0^∞ is a core for A, B, and C (see, e.g. Chernoff–Marsden [7], pages 53–55). On C_0^∞, $A+B=C$, and because $D(A) \cap D(B) \subseteq D(C)$, it follows that $C = \overline{A+B}$.

The next example gives sufficient conditions for the convergence of the Crank–Nicholson scheme. The context is an abstract hyperbolic or parabolic equation (c.f. [7], page 35). For more detailed results, see the classic paper of Kreiss [35]. Nonlinear generalizations of this scheme are discussed in Section 6.

3.7. APPLICATION. Let $\mathbb{H}$ be a real Hilbert space and let A be an operator in $\mathbb{H}$ such that $\langle x, Ax\rangle \leqq 0$ for all $x \in D(A)$ and $\lambda - A$ is surjective for $\lambda > \beta$, so that A generates a C^0 contractive semigroup. Then

$$\exp\{tA\} = \underset{n\to\infty}{s\text{-lim}}\, K_{t/n}^n,$$

where

$$K_{\Delta t} = (I - \tfrac{1}{2}\Delta t A)^{-1} \circ (I + \tfrac{1}{2}\Delta t A). \tag{3.2}$$

Proof: From the dissipative assumption $\langle x, Ax\rangle \leqq 0$,

$$\|x-\lambda Ax\|^2 \geqq \|x\|^2+\lambda^2\|Ax\|^2 \geqq \|x+\lambda Ax\|^2 .$$

Therefore, $\|(I+\frac{1}{2}\epsilon A)(I-\frac{1}{2}\epsilon A)^{-1}y\| \leqq \|y\|$, so $\|K_\epsilon\| \leqq 1$, i.e., stability holds.[4] The identity

$$(\lambda-\epsilon^{-1}(K_\epsilon-I))^{-1}=(\lambda-(1-\tfrac{1}{2}\lambda\epsilon)A)^{-1}(I-\tfrac{1}{2}\epsilon A)$$

shows resolvent consistency. In fact, from

$$\epsilon^{-1}(K_\epsilon-I)=(1-\tfrac{1}{2}\epsilon A)^{-1}A$$

we see that consistency holds.

3.8. EXAMPLE (T. Kato). For an arbitrary one-parameter group of isometries in a Banach space, the Crank–Nicolson algorithm K_ϵ need not be stable (so cannot converge). For instance, let $\exp\{tA\}$ be the shift by t-units in $L^1(-\infty,\infty)$. Let

$$T_n=(1-(t/2n)A)^{-1}K_{t/n}^n=(1+(t/2n)A)^n(1-(t/2n)A)^{-n-1} .$$

Then

$$T_n u=f_n * u, \qquad u\in L^1(-\infty,\infty),$$

where f_n is the inverse Fourier transform of $\xi\mapsto(1-it\xi/2n)^n(1+it\xi/2n)^{-n-1}$. Hence

$$f_n(x)=\begin{cases}(2n/t)\varphi_n(2nx/t) & \text{for } x>0,\\ 0 & \text{for } x<0,\end{cases}$$

where the φ_n are the orthonormalized Laguerre functions on $[0,\infty)$ (that is, $\varphi_n(s)=\exp\{-\frac{1}{2}s\}L_n(s)$, where L_n are the Laguerre polynomials, such that $\int_0^\infty \varphi_n(s)\varphi_m(s)\,ds=\delta_{nm}$). Since $\|T_n\|=\|f_n\|_{L^1}=\|\varphi_n\|_{L^1}\geqq$ const. $n^{1/2}$ (see Askey and Wainger [1]), T_n is not uniformly bounded. Hence $K_{t/n}^n$ is not either.

4. Product Formulas for the Heat and Stokes Equation

1. Introduction. Algorithms for the heat equation, such as random walk with absorbing boundaries have played an important theoretical and computational role. For the Navier–Stokes or the Stokes equation, good computational algorithms are of obvious importance.

[4] In the hyperbolic case, i.e., if A is skew-adjoint, $\|K_\epsilon\|=1$, the well-known energy-preserving property of the step-forward operator for the Crank–Nicolson algorithm.

For the Stokes equation, and more generally for the Navier–Stokes equation, our purpose is to study vorticity production in boundary layers. A detailed investigation of the linear case is crucial in this regard.

2. The "random walk" algorithm for the heat equation. We shall begin with the heat equation. Let $\Omega \subset \mathbb{R}^n$ be an open region with smooth boundary $\partial\Omega$. Let Δ denote the free space Laplacian in $\mathbb{R}^n$, so that

$$(\exp\{t\Delta\}u_0)(x) = \int_{\mathbb{R}^n} \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}} u_0(y)\, dy.$$

Let Δ_0 be the Laplacian on Ω with zero boundary conditions; thus if $u_0 \in H^2(\Omega)$, $u_0 = 0$ on $\partial\Omega$, $(\exp\{t\Delta_0\}u_0)(x) = u(t, x)$ satisfies

$$\frac{\partial u}{\partial t} = \Delta u \quad \text{in} \quad \Omega,$$

$$u(t, x) = 0, \qquad x \in \partial\Omega,$$

$$u(0, x) = u_0(x).$$

Define an algorithm in $L_2(\Omega)$ as follows:

$$K_t : L_2(\Omega) \to L_2(\Omega),$$

$$K_t u = (\exp\{t\,\Delta\}u) \restriction \Omega \quad \text{(restriction to } \Omega),$$

where $\exp\{t\,\Delta\}u$ means $\exp\{t\,\Delta\}$ applied to u extended to be zero outside Ω.

This algorithm is similar to random walk with absorbing boundaries; hence one might conjecture that it converges to $\exp\{t\Delta_0\}$. In fact, in Courant–Friedrichs and Lewy [14] this is established for finite difference schemes. However there is a crucial distinction; for difference schemes, stability requires $\Delta t/(\Delta x)^2 < \frac{1}{2}$ (Δx = spatial discretization). Here effectively Δt is finite and $\Delta x = 0$. Another way of putting it is that infinitesimal 'blobs' of heat can re-enter through the boundary before being annihilated. This difference means that the present scheme is resolvent consistent, but is not consistent. To understand why the scheme ought to converge, imagine the domain Ω to be finite and immersed in an infinite bath which is at 0^0. To approximately solve the heat equation in Ω with zero boundary conditions, one would let heat flow in Ω, at the same time constantly stirring the fluid right outside Ω. The stirring would disperse the heat in the bath, making the temperature of the bath a constant, which would have to be 0^0 since only a finite amount of heat had been added to it. The iteration scheme amounts to stirring the bath every Δt seconds, which ought to be approximately the same for small Δt.

4.1. THEOREM (Kato). *For K_ϵ as defined above,*

$$s\text{-}\lim_{n\to\infty} K_{t/n}^n = \exp\{t\Delta_0\}$$

in $L_2(\Omega)$, uniformly on bounded t-intervals.

Proof: Since $\exp\{t\Delta\}$ is a contraction in $L_2(\mathbb{R}^n)$, $\|K_\epsilon\| \leqq 1$ and hence stability holds. We shall now show resolvent consistency. This is not easy to do by direct computation.[5] Instead one appeals to the following result.

4.2. THEOREM (Kato). *Let A be a non-negative selfadjoint operator in a Hilbert space H, H_0 a subspace of H, and let $P: H \to H_0$ be the orthogonal projection. Assume that $D(A^{1/2}) \cap H_0$ is dense in H_0. For $t>0$, set $A_t = t^{-1}(1-P\exp\{-tA\}) \in B(H_0)$. Then $(1+A_t)^{-1} \to (1+A_0)^{-1}$, $t \downarrow 0$ (strong convergence), where A_0 is the non-negative selfadjoint operator in H_0 associated with the closed quadratic form[6] $u \mapsto \|A^{1/2}u\|^2$, which is densely defined in H_0.*

Proof of Theorem 4.2: Let $f \in H_0$ and $(1+A_t)^{-1}f = u_t$. Then

$$u_t \in H_0 \quad \text{and} \quad f = u_t + A_t u_t = u_t + t^{-1}(1-P\exp\{-tA\})u_t \tag{2.1}$$

$$= u_t + t^{-1}P(1-\exp\{-tA\})u_t = u_t + PB_t u_t,$$

where $B_t = t^{-1}(1-\exp\{-tA\}) \in B(H)$. Relation (2.1) implies

$$(f, u_t) = \|u_t\|^2 + \|B_t^{1/2}u_t\|^2 \tag{2.2}$$

from which it easily follows that $\|u_t\| \leqq \|f\|$, $\|B_t^{1/2}u_t\| \leqq \|f\|$. Hence

$$u_t \rightharpoonup u \in H_0, \quad B_t^{1/2}u_t \rightharpoonup w \in H \quad \text{(weak convergence)} \tag{2.3}$$

along some subsequence $t_n \to 0$. We claim that

$$u \in D(A^{1/2}) \quad \text{and} \quad w = A^{1/2}u. \tag{2.4}$$

[5] For instance, if $\Omega = [0,\infty) \subset \mathbb{R}$, consistency (and hence resolvent consistency) holds on elements $g \in H^2(\Omega)$ such that $g(0)=0$, $g'(0)=0$. Taking orthogonal compliments, then resolvent consistency comes down to this: $(1-\Delta_\epsilon)^{-1}e^{-x} \to xe^{-x/2}$ in $L_2([0,\infty))$ where $\Delta_\epsilon = (K_\epsilon - 1)/\epsilon$. We do not know a direct way of doing this, but the simpler problem, replacing K_ϵ by $(1+\epsilon\Delta)^{-1}$ in $L_2([0,\infty))$, can be done by Wiener–Hopf techniques, as was pointed out by T. Kato.

[6] See Kato [30] and Simon [57] for the correspondences between selfadjoint operators and quadratic forms.

Indeed, let $v \in D(A^{1/2})$. Then $(w, v) = \lim (B_t^{1/2} u_t, v) = \lim (u_t, B_t^{1/2} v) = (u, A^{1/2} v)$, from which (2.4) follows. Next we prove that

$$(1 + A_0)u = f, \quad \text{i.e.,} \quad u = (1 + A_0)^{-1} f. \tag{2.5}$$

Indeed, let $v \in D(A_0^{1/2}) = D(A^{1/2}) \cap H_0$. Then

$$(f, v) = (u_t, v) + (B_t^{1/2} u_t, B_t^{1/2} v) \to (u, v) + (w, A^{1/2} v),$$

$$(f, v) = (u, v) + (A^{1/2} u, A^{1/2} v).$$

Thus the definition of A_0 gives (2.5), which implies that the weak limits in (2.3) are independent of the subsequence chosen. By a standard argument, it follows that (2.3) holds as $t \downarrow 0$, without taking any subsequence.

It remains to prove strong convergence. To this end, note that (2.2) implies

$$\|u_t\|^2 + \|B_t^{1/2} u_t\|^2 \to (f, u) = ((1 + A_0)u, u)$$
$$= \|u\|^2 + \|A_0^{1/2} u\|^2 = \|u\|^2 + \|A^{1/2} u\|^2.$$

Then we get, using (2.3),

$$\|u_t - u\|^2 + \|B_t^{1/2} u_t - A^{1/2} u\|^2 \to 0.$$

Hence $u_t \to u = (1 + A_0)^{-1} f$,

Remark. The existence of s-lim $(1 + A_t)^{-1}$ is trivial, since $A_t = PB_t P^*$ (where $P^* : H_0 \to H$ is the adjoint of P, i.e., P^* is the injection), and B_t is monotone increasing as $t \downarrow 0$.

Completion of the proof of Theorem 4.1: We apply Theorem 4.2 with $H = L_2(\mathbb{R}^n)$, $H_0 = L_2(\Omega)$ (regarded as a subspace of H by extending elements in $L_2(\Omega)$ to be zero outside Ω), and $A = -\Delta$. From the facts that (i) $\|A^{1/2} u\|^2 = \|\text{grad } u\|^2$ and (ii) $u \in D(A^{1/2}) \cap H_0$ is equivalent to $u \in H_0^1(\Omega)$ it follows that $A_0 = -\Delta_0$, and hence the resolvent consistency.

It is worth seeing why the algorithm in Theorem 4.1 is not consistent. The following argument is formal, but is easy to make precise.

For $t > 0$, u smooth and 0 on $\partial\Omega$,

$$\frac{d}{dt} K_t u = \Delta \exp\{t\Delta\} u \restriction \Omega.$$

For $x \in \Omega$, we can work out $(\Delta \exp\{t\Delta\}u)(x)$ explicitly; we get

$$\int_\Omega \Delta_y\left(\frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}}\right)u(y)\,dy$$
$$= \int_\Omega \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}}\Delta_y u(y)\,dy - \int_{\partial\Omega} \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}}\frac{\partial u}{\partial n}\,dy,$$

by Green's identity. As $t \downarrow 0$, this becomes

$$\Delta u - \delta_{\partial\Omega}\frac{\partial u}{\partial n}$$

and so the δ function term blows up in L_2.

3. Kato's conditions for Trotter's formula. The techniques used in Theorem 4.2 have been used by Kato [34] to obtain the following rather general result.

4.3. THEOREM (Kato). *Let A and B be non-positive selfadjoint operators in a Hilbert space H. Then*

$$\operatorname{s-lim}_{n\to\infty}(\exp\{tA/n\}\exp\{tB/n\})^n = \exp\{tC_0\}P,$$

where C_0 is the form sum of A and B and is a non-positive selfadjoint operator in the closed subspace H_0 of H spanned by $D_0 = D((-A)^{1/2}) \cap D((-B)^{1/2})$, and where P is the orthogonal projection of H onto H_0. In particular, the limit is zero if $D_0 = \{0\}$ (see Chernoff [7] for an example).

We refer to Kato's paper for generalizations and applications.

4. The modified heat algorithm. As we have remarked earlier, convergence may be undetectable numerically if only resolvent consistency holds. We shall now present a modification of the algorithm in Theorem 4.1 which will be important for the Navier–Stokes equations and which is consistent in the usual sense. Computationally, the algorithm in Theorem 4.1 has almost undetectable convergence since the convergence and error rates balance, whereas the modified algorithm converges quickly.

The modified algorithm is described as follows. Let $U \supset \partial\Omega$ be a tubular neighborhood of $\partial\Omega$ composed of straight line segments through $\partial\Omega$ and normal to it. We assume either Ω is bounded or $\partial\Omega$ has bounded curvature so these line segments extend a distance $\delta > 0$ away from $\partial\Omega$ without crossing.

Let $\varphi : U \to U$ be the map which reflects across the boundary relative to these lines.

Consider the map $\Phi : L_2(\Omega) \to L_2(\mathbb{R}^n)$, $u \mapsto \Phi(u) = \tilde{u}$, where $\tilde{u}$ equals u in Ω, $-u \circ \varphi$ in $U - \Omega$ and is zero outside $\Omega \cup U$, i.e., $\tilde{u}$ is u extended across $\partial\Omega$ to be odd.

Consider the algorithm whose step-forward operator is

$$K_{\Delta t}(u) = (\exp\{(\Delta t)\,\Delta)\tilde{u}) \restriction \Omega .$$

4.4. THEOREM. *The algorithm just defined is consistent and stable; thus*

$$s\text{-}\lim K_{t/n}^n = \exp\{t\Delta_0\}$$

uniformly on bounded t-intervals.

Proof: First we prove consistency. For $t > 0$, $x \in \Omega$, and $u \in H^2(\Omega)$, $u = 0$ on $\partial\Omega$, we have

$$\begin{aligned}
\left(\frac{d}{dt} K_t u\right)(x) &= (\Delta \exp\{t\Delta\}\tilde{u})(x) \\
&= \int_\Omega \Delta_x \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}} u(y)\, dy \\
&\quad + \int_{\mathbb{R}^n - \Omega} \Delta_x \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}} \tilde{u}(y)\, dy \\
&= \int_\Omega \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}} \Delta u(y)\, dy - \int_{\partial\Omega} \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}} \frac{\partial u}{\partial n}\, dy \\
&\quad - \int_{\partial(U \cup \Omega)} \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}} \frac{\partial \tilde{u}}{\partial n}\, dy \\
&\quad + \int_{U \cap (\mathbb{R}^n - \Omega)} \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}} \Delta\tilde{u}(y)\, dy \\
&\quad - \int_{\partial\Omega} \frac{\exp\{-|x-y|^2/4t\}}{(4\pi t)^{n/2}} \frac{\partial \tilde{u}}{\partial \tilde{n}}\, dy .
\end{aligned}$$

The two integrals over $\partial\Omega$ cancel exactly and so this converges strongly in $L_2(\Omega)$ *to* Δu, as $t \downarrow 0$. (The integrals over $U \cap (\mathbb{R}^3 - \Omega)$ and $\partial(U \cup \Omega)$ converge to zero since $x \in \Omega$.)[7]

[7] It can be checked that this algorithm is not, in general, second order accurate.

Secondly we must prove stability. The complication here is that K_ϵ is not a contraction and to demonstrate stability some cancellations must be taken into account. The following estimate suffices for stability:

$$\|K_\epsilon u\|_{L_2(\Omega)} \leqq (1+\gamma\epsilon)\,\|u\|_{L_2(\Omega)}$$

for a constant γ, depending on the curvature of $\partial\Omega$. Equivalently, we can obtain an estimate of the form

$$\|K_\epsilon u\|^2_{L_2(\Omega)} \leqq (1+\gamma\epsilon)\,\|u\|^2_{L_2(\Omega)}.$$

Write $K(x-y,t)=\exp\{-|x-y|^2/4t\}/(4\pi t)^{n/2}$. By the change of variables formula,

$$\begin{aligned}(K_\epsilon u)(x) &= \int_{\Omega-U} K(x-y,\epsilon)u(y)\,dy \\ &\quad + \int_{U\cap\Omega} [K(x-y,\epsilon)-K(x-\varphi(y),\epsilon)J_\varphi(y)]u(y)\,dy,\end{aligned} \tag{4.1}$$

where $J_\varphi(y)$ is the absolute value of the Jacobian of φ.

At the outset, let us choose U sufficiently small so that $J_\varphi(y)\leqq 2$ and points in U are a distance at most δ from $\partial\Omega$. This shrinking of U does not affect the estimates (δ depends only on the curvature of $\partial\Omega$ and is determined in the proof).

Let us first prove a special case. In fact, we shall show that K_ϵ is a contraction if Ω is convex. In this case, for $x,\ y\in U\cap\Omega$, $|x-y|\leqq|x-\varphi(y)|$, so

$$|K(x-y,t)-K(x-\varphi(y),t)J_\varphi(y)| \leqq K(x-y,t).$$

Here we used the inequality $|a-b|\leqq a$ if $a\geqq 0$, $b\geqq 0$ and $b\leqq 2a$. Thus, from (4.1), for each $x\in\Omega$,

$$|(K_\epsilon u)(x)| \leqq \int_\Omega K(x-y,\epsilon)\,|u(y)|\,dy.$$

Since the heat kernel induces a contraction, $\|K_\epsilon u\|_{L_2(\Omega)}\leqq\|u\|_{L_2(\Omega)}$. (Note that this works in a wide variety of function spaces.)

Now we turn to the general case. Let

$$L(x,y,t)=\begin{cases}K(x-y,t) & \text{if } \|x-y\|\leqq\|x-\varphi(y)\| \text{ or if } x\in\Omega-U,\\ K(x-y,t)-K(x-\varphi(y),t)J_\varphi(y) & \text{if } \|x-y\|>\|x-\varphi(y)\|.\end{cases} \tag{4.2}$$

so that (4.1) reads

$$(K_\epsilon u)(x) = \int_\Omega L(x, y, \epsilon) u(y)\, dy. \tag{4.3}$$

We can estimate the L_2 norm of (4.3) by the Schwarz inequality as follows:

$$\begin{aligned}
\|K_\epsilon u\|^2_{L_2(\Omega)} &\leqq \int_\Omega \left(\int_\Omega |L(x, y, \epsilon)|\, |u(y)|\, dy \right)^2 dx \\
&= \int_\Omega \left(\int_\Omega |L(x, y, \epsilon)|^{1/2} \left(|L(x, y, \epsilon)|^{1/2} |u(y)|\right) dy \right)^2 dx \\
&\leqq \int_\Omega \left(\int_\Omega |L(x, z, \epsilon)|\, dz \right) \left(\int_\Omega |L(x, y, \epsilon)|\, |u(y)|^2\, dy \right) dx \\
&\leqq \left(\sup_{x \in \Omega} \int_\Omega |L(x, z, \epsilon)|\, dz \right) \left(\sup_{y \in \Omega} \int_\Omega |L(x, y, \epsilon)|\, dx \right) \|u\|_{L_2(\Omega)}.
\end{aligned}$$

We shall prove the two estimates

$$\int_\Omega |L(x, y, t)|\, dy \leqq 1, \tag{4.4}$$

and

$$\int_\Omega |L(x, y, t)|\, dx \leqq 1 + Ct, \tag{4.5}$$

which will give us the result. (In what follows, C is a generic constant.)

The estimate (4.4) is proved by our earlier convex argument: for any $x \in \Omega$,

$$\int_\Omega |L(x, y, t)|\, dy = \int_{(\Omega - U) \cup \text{convex}} |L(x, y, t)|\, dy + \int_{\text{concave}} L(x, y, t)|\, dy,$$

where

$$\text{convex} = \{ y \in U \cap \Omega \mid \|x - y\| \leqq \|x - \varphi(y)\| \},$$

$$\text{concave} = \{ y \in U \cap \Omega \mid \|x - y\| > \|x - \varphi(y)\| \}.$$

Thus

$$\begin{aligned}
\int_\Omega |L(x, y, t)|\, dy &\leqq \int_\Omega K(x - y, t)\, dy + \int_{\text{concave}} |L(x, y, t)|\, dy \\
&\leqq \int_\Omega K(x - y, t)\, dy + \int_{\text{concave}} K(x - \varphi(y), t) J_\varphi(y)\, dy \\
&= \int_\Omega K(x - y, t)\, dy + \int_{\varphi(\text{concave})} K(x - z, t)\, dz \\
&\leqq \int_{R^n} K(x - y, t)\, dy = 1.
\end{aligned}$$

It remains to prove (4.5). We do this as follows: for each $y \in U \cap \Omega$ draw the line from y to $\varphi(y)$ and locate the origin 0 on the midpoint of this line, so that $0 \in \partial\Omega$. Let y sit on the x_n coordinate axis. Since Ω is compact, or $\partial\Omega$ has bounded curvature, there is a paraboloid $x_n = -C(x_1^2 + \cdots + x_{n-1}^2)$, $|x_i| \leqq 1$, $i = 1, \cdots, n-1$, containing $\partial\Omega$ between it and the $x_1, \cdots, x_{n-1}$-plane. This region, denoted D, is shaded in Figure 1. From estimate (4.4), it suffices to prove

$$\int_D |L(x, y, t)|\, dx \leqq Ct. \tag{4.6}$$

Notice that, $\varphi(y) = -y$ and that in D

$$|L(x, y, t)| = K(x+y, t) |\exp\{-(\|x-y\|^2 - \|x+y\|^2)/t\} - J_\varphi(y)|$$

$$= K(x+y, t)\, |\exp\{\langle x, y\rangle/t\} - J_\varphi(y)|\,.$$

Since $J_\varphi(y)$ is smooth and is one on $\partial\Omega$,

$$|J_\varphi(y) - 1| \leqq C\|y\|\,;$$

hence

$$|L(x, y, t)| \leqq K(x+y, t)\, |(\exp\{\langle x, y\rangle/t\} - 1) + C\|y\||\,.$$

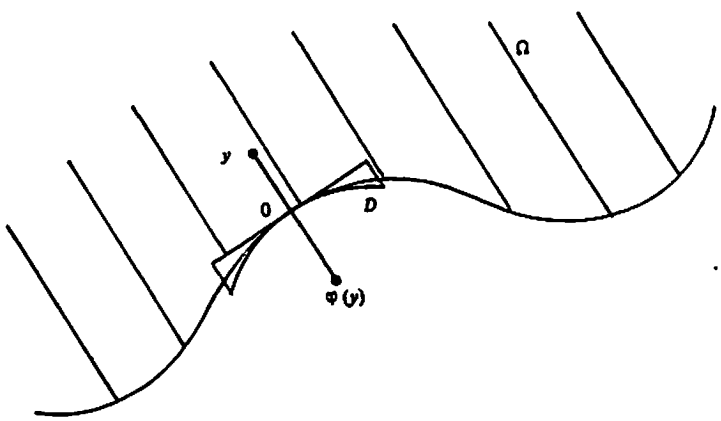

Figure 1

In D, $\langle x, y\rangle \leqq 0$ and thus

$$|L(x,y,t)| \leqq K(x+y,t)\left(\frac{|\langle x,y\rangle|}{t} + C\|y\|\right)$$

$$\leqq K(x+y,t)\left(\frac{|x_n|\,\|y\|}{t} + C\|y\|\right)$$

$$\leqq CK(x+y,t)\left(\frac{|x_n|}{t}+1\right)\|y\| \tag{4.7}$$

$$\leqq CK(x+y,t)\left(\frac{(x_1^2+\cdots+x_{n-1}^2)}{t}+1\right)\|y\|$$

$$\leqq C\frac{\exp\{-\|x\|^2/4t\}}{(4\pi t)^{n/2}}\exp\{-\langle x,y\rangle/4t\}\exp\{-\|y\|^2/4t\}$$

$$\times\left(\frac{x_1^2+\cdots+x_{n-1}^2}{t}+1\right)\|y\|$$

$$\leqq CK(x,t)\exp\{C(x_1^2+\cdots+x_{n-1}^2)\,\|y\|/4t\}(\exp\{-\|y\|^2/4t\}\,\|y\|)$$

$$\times\left(\frac{x_1^2+\cdots+x_{n-1}^2}{t}+1\right).$$

Now change variables: $x_i = \sqrt{t}\,w_i$, $y_i = \sqrt{t}\,z_i$; then D becomes the region $D_t : 0 \geqq w_n \geqq -\sqrt{t}\,C(w_1^2+\cdots+w_{n-1}^2)$. Hence, (4.7) becomes

$$\int_D |L(x,y,t)|\,dx \leqq C\int_{D_t}\left(\frac{\exp\{-\|w\|^2/4\}}{(4\pi)^{n/2}}\right)(\exp\{C(w_1^2+\cdots+w_{n-1}^2\,\|y\|/4\}) \tag{4.8}$$

$$\times(\exp\{-\|z\|^2/4\}\,\|z\|\sqrt{t})(w_1^2+\cdots+w_{n-1}^2+1)\,dw.$$

We chose δ so that $\frac{1}{4}C\|y\| < \frac{1}{8}$. (Here C is the constant for the parabolas.) Thus (4.8) gives

$$\int_D |L(x,y,t)|\,dx \leqq C\sqrt{t}\int_{D_t}\exp\{-\|w\|^2/8\}(\|w\|^2+1)\,dw, \tag{4.9}$$

since $\exp\{-\|z\|^2/4\}\,\|z\|$ is uniformly bounded. Finally, an elementary estimate gives

$$\int_{D_t}\exp\{-\|w\|^2/8\}(\|w\|^2+1)\,dw \leqq C\left(\sum_{n=0}^{\infty}\exp\{-n^2/8\}((n+1)^2+1)\right)\sqrt{t} \tag{4.10}$$

$$\leqq C\sqrt{t}.$$

Combining (4.9) and (4.10), we obtain

$$\int_D |L(x, y, t)|\, dx \leqq Ct,$$

our required estimate.

Remark. This algorithm works in other function spaces than L_2 using essentially the same proof. For example, one can use $C^\alpha(\Omega)$, $0<\alpha<1$. We shall come back to this point in Section 6. On the other hand, Theorem 4.1 depends heavily on the fact that one is dealing with positive selfadjoint operators on Hilbert space.

5. Algorithms for the Stokes equations. Next we turn to a discussion of the Stokes equations. First the notation. Let $J_2(\Omega)$ denote the L_2 closure of the C_0^∞ vector fields u in Ω with div $u = 0$. Formally, $u \in J_2(\Omega)$ if div $u = 0$ and u is parallel to the boundary. Let

$$\mathbb{P}\colon L_2(\Omega) \to J_2(\Omega)$$

be the L_2 orthogonal projection. (Then we have the well-known decomposition $L_2(\Omega, \mathbb{R}^n) = G_2(\Omega) \oplus J_2(\Omega)$, where $G_2(\Omega)$ consists of gradients of (locally) H^1 functions.) Then $\exp\{t\mathbb{P}\Delta_0\}$ is the contractive semigroup associated with the Stokes' equation. it is defined since $\mathbb{P}\Delta_0$ is a non-positive selfadjoint operator. Thus $\exp\{t\mathbb{P}\Delta\}u_0 = u$ solves

$$\frac{\partial u}{\partial t} = \Delta u - \operatorname{grad} p,$$

$$\operatorname{div} u = 0,$$

$$u = 0 \quad \text{on} \quad \partial\Omega \quad \text{and} \quad u(x, 0) = u_0(x),$$

i.e.,

$$\frac{\partial u}{\partial t} = \mathbb{P}\Delta u,$$

$$u = 0 \quad \text{on} \quad \partial\Omega \quad \text{and} \quad u(x, 0) = u_0(x).$$

4.5. THEOREM. *Let $K_s : J_2(\Omega) \to J_2(\Omega)$ be given by*

$$K_s u = \mathbb{P}(\exp\{t\Delta\}u \restriction \Omega). \tag{5.1}$$

Then

$$\operatorname*{s-lim}_{n\to\infty} K_{t/n}^n = \exp\{t\mathbb{P}\Delta_0\},$$

uniformly on bounded t-intervals.

This algorithm is stable (obviously K_t is a contraction) and resolvent consistent. This follows from Theorem 4.2 in exactly the same way as in the proof of Theorem 4.1. However, just as above, it is not consistent. We can rectify the latter situation exactly as in the heat equation. Let $u \mapsto \Phi(u) = \tilde{u}$ be the odd extention of u, this time for vector fields.

4.6. THEOREM. *Let* $K_t : J_2(\Omega) \to J_2(\Omega)$ *be given by*

$$K_t(u) = \mathbb{P}(\exp\{t\Delta\}\tilde{u} \restriction \Omega). \tag{5.2}$$

Then

$$\operatorname*{s-lim}_{n\to\infty} K_{t/n}^n = \exp\{t\mathbb{P}\Delta_0\}$$

uniformly on bounded t-intervals.

This time the scheme is consistent and stable (so, as above, converges faster). Theorem 4.6 follows directly from Theorem 4.4.

6. The creation of vorticity. Some comments on the physical meaning of the algorithm (4.2) are in order. This becomes most meaningful in the context of the Navier–Stokes equations, but is worth explaining here.

Given $u \in J_2(\Omega)$, the map $u \mapsto \tilde{u}$ creates a δ-layer of vorticity at $\partial\Omega$ (see Figure 2). The strength of the layer is twice the component of u parallel to the boundary. Following this, $\exp\{t\Delta\}\tilde{u}$ diffuses these vortices away from the

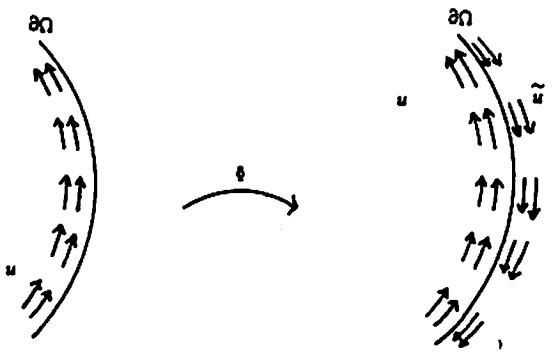

Figure 2. The vorticity creation operator.

boundary and finally $\mathbb{P}$ gives us back a divergence free vector field parallel to the boundary.

Remarks. 1. A characterization of the step forward operator (5.1) for the Stokes equation can be given in terms of the vorticity alone. In fact, given u (sufficiently smooth), $\operatorname{div} u = 0$ and $u|_{\partial\Omega} = 0$, let $\omega = \nabla \times u$ and $\omega_t = \exp\{t\Delta\}\omega|_\Omega$. Let u_t be the velocity field associated with ω_t; i.e., u_t solves

$$\nabla \times u_t = \omega_t ,$$

$$\operatorname{div} u_t = 0 ,$$

$$u_t \parallel \partial\Omega ,$$

Then

$$u_t = \mathbb{P}((\exp\{t\Delta\}u) \restriction \Omega) .$$

This may be verified by a straightforward calculation.

2. For the modified algorithm (5.2) a similar formulation in terms of the vorticity is possible. Here $\tilde{\omega} = \nabla \times \tilde{u}$ will have a δ-layer on $\partial\Omega$ (which can be smoothed out if desired—see Remark 4), but $\omega_t = \exp\{t\Delta\}\tilde{\omega}$ will be smooth.

3. The modified algorithm (5.2) might be of use in showing that the Stokes equation generates a C^0 semigroup in $L_p(\Omega)$.

4. The vorticity creation operator Φ can be changed to a large extent without affecting the validity of the above product formula. For example, the vorticity sheet can be smeared out to some extent so that if u is smooth, so is $\tilde{u}$. The width of the region of smoothing must decrease faster than $\sqrt{\Delta t}$ as $\Delta t \downarrow 0$. Also, the region U can depend on Δt; its width must go to zero slower than $\sqrt{\Delta t}$ as $\Delta t \downarrow 0$.[8]

5. Convergence of Nonlinear Algorithms

1. Introduction. Now we turn our attention to the nonlinear case. In this direction there are at least two approaches:

(i) develop a theory parallel to the linear case using monotone operators as generators,

(ii) develop a theory suitable for evolution problems in which solutions might exist only for a short time.

The general theory in case (i) is reasonably satisfactory, due to work of

[8] Without these conditions, one can show that consistency fails by example. (Details are available on request from Marjorie F. McCracken.)

Brezis and Pazy [3], [4]. See also Webb [65], Pazy [68] and Weissler [69]. A number of interesting questions still remain here however.

In structural dynamics, elasticity and fluid mechanics, it seems at present that a local (in time) approach is more fruitful. As in Theorem 2.1, one will have the validity of product formulas as long as the solution exists and is sufficiently smooth.

One can turn the problem around and use product formulas to get existence theorems for both classical and weak solutions. This is particularly promising for a nonlinear energy preserving algorithm (a nonlinear generalization of the Crank–Nicolson algorithm) presented in Section 7.

Here we shall present a general result suitable for many short time evolution problems. The idea follows Marsden [42], which was inspired in turn by Ebin–Marsden [16]. Sharper versions of these results in the context of, for example, Hughes, Kato and Marsden [29] ought to be possible.

2. A Convergence theorem. First some terminology. $Y \subset X$ will be Banach spaces with the inclusions dense and continuous. Let $A : Y \to X$ be a given nonlinear operator defined on an open set in Y. We are interested in the evolution problem

$$\frac{du}{dt} = A(u), \qquad u(0) = u_0, \tag{2.1}$$

for a curve $u(t) \in Y$ which is differentiable in X.

Let $K_\epsilon : Y \to Y$ be given maps with $\epsilon \geqq 0$ and $K_0 =$ identity (K_ϵ may be defined only on an open set).

We say K_ϵ is *consistent*[9] if, for all $y \in Y$,

$$\frac{d}{d\epsilon +} K_\epsilon(y)|_{\epsilon=0} = A(y).$$

The algorithm will be called *locally Y-stable* if, for all $y_0 \in Y$ and a neighborhood U of y_0, there is a $T > 0$ and a neighborhood V of y_0 such that $K^n_{t/n}(y)$ is defined and lies in U for all $y \in V$ and $0 \leqq t \leqq T$. (See proposition 5.2 below for some sufficient conditions.)

[9] Outside the context of monotone operators, the possibility of using resolvent consistency has not been adequately explored. Presumably it means that, at each $y \in Y$, $\left(\lambda - \frac{1}{\epsilon}(DK_\epsilon(y) - I)\right)^{-1} \exists (\lambda - DA(y))^{-1}$, where DK_ϵ is the Fréchet derivative.

Our first result is a straightforward generalization of the ordinary differential equations case. We make these assumptions:

(A_1) The algorithm is consistent, locally uniformly in Y; i.e.,

$$\frac{1}{\epsilon}\|K_\epsilon(y)-A(y)\|_X \to 0 \quad \text{as} \quad \epsilon \to 0+$$

uniformly for y in a Y-neighborhood of y_0.

(A_2) The algorithm is locally Y-stable.

In the case of nonlinear partial differential equations, consistency and stability do not suffice for convergence. We need another condition such as the following:

(A_3) The equation (2.1) defines a C^0-local semiflow $F_t : Y \to Y$ which is Y-locally X-Lipschitz, i.e., for y_1, y_2 in a Y-neighborhood of y_0,

$$\|F_t(y_1)-F_t(y_2)\|_X \leqq C\|y_1-y_2\|_X, \qquad 0 \leqq t \leqq T.$$

Remark. (A_3) does not say that F_t is locally Lipschitz in X. While the latter holds in parabolic and semi-linear hyperbolic problems, only (A_3) need hold in quasi-linear hyperbolic equations (see [29]).

5.1. THEOREM.[10] *Under the assumptions* (A_1)–(A_3), *for each* $y_0 \in Y$, $\lim_{n\to\infty} K^n_{t/n}(y_0) = F_t(y_0)$ *uniformly*, $0 \leqq t \leqq T$. *Moreover, the limit is defined and exists for* $t \in [0, T_1]$ *if and only if* $F_t(y_0)$ *exists for* $t \in [0, T_1]$.

Proof: As in Theorem 2.1 we write

$$\begin{aligned} F_t(y_0)-K^n_{t/n}(y_0) = {} & F^{n-1}_{t/n}F_{t/n}(y_0)-F^{n-1}_{t/n}K_{t/n}(y_0) \\ & + F^{n-2}_{t/n}F_{t/n}(y_1)-F^{n-2}_{t/n}K_{t/n}(y_1) \\ & \cdot \\ & \cdot \\ & \cdot \\ & + F_{t/n}(y_{n-1})-K_{t/n}(y_{n-1}), \end{aligned}$$

where $y_k = K^k_{t/n}(y_0)$. By assumption (A_2), the y_k lie in a y-neighborhood of y_0. From (A_3), and due to the fact that $F^{n-k}_{t/n} = F_{t((n-k)/n)}$,

$$\|F_t(y_0)-K^n_{t/n}(y_0)\|_X \leqq C \cdot \sum_{k=0}^{n-1} \|F_{t/n}(y_k)-K_{t/n}(y_k)\|_X.$$

Now, by (A_1), this is at most $C \cdot n \cdot o(1/n) \to 0$ as $n \to \infty$. The last part may be completed as in Theorem 2.1.

[10] This is implicit in the literature in a number of cases (e.g., [59], II).

3. Conditions for stability. The following gives sufficient conditions for Y-stability of an algorithm.

5.2. PROPOSITION. *Let the algorithm K_ϵ satisfy the following conditions:*
(S_1) *for a dense set $Z \subset Y$, $K_\epsilon(z)$ is Y-differentiable in ϵ,*
(S_2) *there is a metric d on Y equivalent to the given norm on Y such that locally in Y*

$$d(K_\epsilon(y_1), K_\epsilon(y_2)) \leqq \exp\{\beta\epsilon\}\, d(y_1, y_2).$$

Then K_ϵ is locally Y-stable.

Proof: First one shows that $K^n_{t/n}(z)$ remains defined and lies in a Y-ball by using

$$d(K_\epsilon(z), z) \leqq C\epsilon$$

and induction on the estimate

$$d(K^n_{t/n}(z), z) \leqq d(K^{n-1}_{t/n} K_{t/n}(z), K^{n-1}_{t/n}(z)) + d(K^{n-2}_{t/n} K_{t/n}(z), K^{n-1}_{t/n}(z))$$
$$+ \cdots + d(K_{t/n}(z), z)$$
$$\leqq n \exp\{\beta t\} C \frac{t}{n} = C \exp\{\beta t\} t.$$

Thus if t is sufficiently small, $K^n_{t/n}(z)$ lies in any given Y-ball about z. For $y \in Y$ and $d(z, y) < \delta$, one gets inductively

$$d(K^n_{t/n}(y), K^n_{t/n}(z)) \leqq \exp\{\beta t\}\, d(y, z)$$

and hence the result.

4. Product formulas used to prove an existence theorem. In some circumstances one can use product formulas to deduce properties of the flow F_t and indeed use them to prove an existence theorem. The next theorem gives an example of such a result.[11]

[11] For instance, this is the method used by Ebin-Marsden [16] to show that, on a boundaryless manifold, the time of existence in the Navier-Stokes equations is independent of the Reynolds number $1/\nu$ and that the solutions converge as $\nu \to 0$.

This time we assume there are three Banach spaces $Z \subset Y \subset X$ with inclusions continuous and dense. We make these hypotheses on K_ϵ:

(B_0) K_ϵ maps (an open set in) Y to Y and Z to Z; $K_0 =$ identity.

(B_1) For $y \in Y$, $\epsilon \mapsto K_\epsilon(y)$ from $[0, T]$ to X is C^1 and $A(y) \equiv \frac{d}{d\epsilon +} K_\epsilon(y)|_\epsilon = 0$ is continuous from Y to X.

(B_2) K_ϵ is locally Z and Y-stable.

(B_3) Locally in Y, K_ϵ satisfies the estimate

$$\|K^k_{t/n}(y_1) - K^k_{t/n}(y_2)\|_Y \leqq C \|y_1 - y_2\|_Y$$

for all n, $0 \leqq t \leqq T$, $0 \leqq k \leqq n$.

(B_4) Locally in Z we have the estimate

$$\|K_{\epsilon+\eta}(y) - K_\epsilon K_\eta(y)\|_Y \leqq C\epsilon\eta\,.$$

Remark. It would be of some interest to weaken (B_3) to "locally in Z".

5.3. THEOREM. *Assume* (B_0)–(B_4). *Then locally in* Y, $\lim_{n\to\infty} K^n_{t/n}(y) = H_t(y)$ *exists (in* Y*) and defines a* Y*-locally Lipschitz semi-flow* H_t *(in* Y*). Also,*

(i) *the semi-flow is generated by* A *in the sense that, for* $y \in Y$, $t \geqq 0$,

$$\frac{d}{dt} H_t(y) = A(H_t(y)) \quad \text{in} \quad X,$$

(ii) A *has unique integral curves,*

(iii) *the above limit holds on* $[0, T)$ *if and only if* $H_t(y)$ *exists on* $[0, T)$.

Proof: By local Y-stability and (B_3), it is enough to prove convergence at $y \in Z$; one can then extend convergence to $y \in Y$ by continuity. By Z-stability, the iterates will remain in a Z-neighborhood of $y \in Z$. If we write

$$\begin{aligned}\|K^n_{t/n}(y) - K_t(y)\|_Y \leqq{}& \|K_t(y) - K_{t/n} \circ K_{t((n-1)/n)}(y)\|_Y \\ &+ \|K_{t/n} \circ K_{t((n-1)/n)}(y) - K_{t/n} \circ K_{t/n} \circ K_{t((n-2)/n)}(y)\|_Y \\ &+ \cdots + \|K^{n-2}_{t/n} \circ K_{t\cdot 2/n}(y) - K^{n-2}_{t/n} \circ K_{t/n} \circ K_{t/n}(y)\|_Y\end{aligned}$$

and use (B_3), (B_4) and induction, we get the estimate (with a generic constant C)

$$\|K^n_{t/n}(y)-K_t(y)\|_Y \leqq C\frac{t^2}{n}\left\{\frac{n-1}{n}+\frac{n-2}{n}+\cdots+\frac{1}{n}\right\} \leqq Ct^2. \tag{4.1}$$

Using (4.1), we get

$$\|K^n_{t/n}(y)-K^{nl}_{t/nl}(y)\|_Y \leqq \sum_{j=0}^{n-1}\|K^{n-j-1}_{t/n}\circ K_{t/n}\circ K^{jl}_{t/nl}(y)-K^{n-j-1}_{t/n}K^{(j+1)l}_{t/nl}(y)\|_Y$$

$$\leqq C\sum_{j=0}^{n-1}\|K_{t/n}\circ K^{jl}_{t/nl}(y)-K^{l}_{t/nl}\circ K^{jl}_{t/nl}(y)\|_Y.$$

Since $K^{jl}_{t/nl}(y)$ lies in a Z-ball, by assumption, the previous estimate gives

$$\leqq C\sum_{j=0}^{n-1}\left(\frac{t}{n}\right)^2 \leqq \frac{Ct^2}{n}. \tag{4.2}$$

Therefore, by (4.2),

$$\|K^n_{t/n}(y)-K^m_{t/m}(y)\|_Y \leqq \|K^n_{t/n}(y)-K^{nm}_{t/nm}(y)\|_Y-\|K^{nm}_{t/nm}(y)-K^m_{t/m}(y)\|_Y$$

$$\leqq Ct^2\left(\frac{1}{n}+\frac{1}{m}\right).$$

Thus $K^n_{t/n}(y)$ is a Cauchy sequence, and therefore converges.

That H_t is a semiflow, i.e., $H_{t+s}=H_t\circ H_s$ follows readily from $\lim_{n\to\infty}K^n_{t/n}(y)=H_t(y)$ if t and s are rationally related, and by continuity for all t, s.

From (4.1), letting $n\to\infty$, we have

$$\|K_t(y)-H_t(y)\|_Y \leqq Ct^2.$$

From this and (B_1) we get

$$\frac{d}{dt+}H_t(y)|_{t=0}=A(y) \quad \text{in} \quad X.$$

From $H_{t+s} = H_t \circ H_s$ we see that $H_t(y)$ is right differentiable at t with derivative $A(H_t(y))$. Since this is continuous, we get (i) ([67], page 239). Part (ii) follows since H_t is Y-locally Y-Lipschitz (see [7]). Finally, the last part is proved as in Theorem 2.1.

Remarks. 1. If one has a family K_ϵ^ν of algorithms depending on a parameter ν, if the basic constants in Theorem 5.3 are independent of ν and $K_\epsilon^\nu \to K_\epsilon$ as $\nu \to 0$ in Y uniformly in ϵ, then the H_t^ν have times of existence independent of ν and $H_t^\nu \to H_t$ in Y.

2. For further information on the stability hypotheses, applications and questions of differentiability of $H_t(y)$ in y, see [43].

6. The Vorticity Algorithm for the Navier-Stokes Equations

1. Introduction and statement of the algorithm. Consider a region $\Omega \subset \mathbb{R}^n$ with smooth boundary $\partial\Omega$. For technical convenience, assume Ω is bounded. Here we are concerned with an algorithm for the Navier–Stokes equations; *viz*

$$\begin{aligned} &\frac{\partial u}{\partial t} = \nu \Delta u - u \cdot \nabla u - \nabla p, \\ &\operatorname{div} u = 0, \\ &u = 0 \quad \text{on} \quad \partial\Omega. \end{aligned} \tag{1.1}$$

Chorin [9] introduced a powerful numerical method for solving (1.1) which is based on a heuristic model of boundary layer mechanics and which explicitly includes a mechanism for vorticity production near the boundary. This method was written as a product formula in Marsden [43], although a number of important factors were then unclear.

Briefly, the vorticity algorithm is as follows. To solve an initial value problem with initial value u_0 in a domain $\Omega \subseteq R^2$, a grid is introduced. In the center of each box is placed a point vortex[12] whose strength is equal to the integral of the vorticity $\omega_0 = \nabla \times u_0$ over the box. Then the point vortices are moved around in such a way that the solution to the Euler equations is approximated. At this point, the velocity field associated with the point vortices is parallel to $\partial\Omega$, so that a layer of point vortices is added on $\partial\Omega$ whose associated velocity field exactly cancels the velocity field already there. Finally, each vortex is walked one random step, discarding those vortices

[12] A point vortex at x_0 is a delta function of vorticity at x_0. The associated velocity field has circular streamlines and the speed at x falls off like $1/r$, where $r = |x - x_0|$.

which cross $\partial\Omega$ into Ω^c. The whole procedure is then repeated, etc. The algorithm described below is a continuous analogue to the numerical method.

Let E_t be the local flow defined by the Euler equations (see subsection 2.4), so that E_t is a well-defined C^0 flow in Sobolev or Hölder function spaces (cf. [16]).

Let $\Phi(u)=\tilde{u}$ be the vorticity creation operator described in subsections 4.5 and 4.6, and let $H_t(\tilde{u})=\mathbb{P}(\exp\{t\nu\Delta\}\tilde{u}\restriction\Omega)$; then our earlier Stokes algorithm (formula (5.2) of subsection 4.5) reads:

$$u\mapsto H_t\circ\Phi(u)\,.$$

The algorithm for the Navier–Stokes equation now is defined as follows: for u a divergence-free vector field on Ω, $u=0$ on $\partial\Omega$, let

$$K_\epsilon(u)=H_\epsilon\circ\Phi\circ E_\epsilon(u)\,. \tag{1.2}$$

The operator Φ in (1.2) plays a more fundamental role than it did for the Stokes equation; here, for $u=0$ on $\partial\Omega$, $E_\epsilon(u)$ will only be parallel to the boundary; the operator Φ corrects for this by creating a vortex layer whose flow cancels $E_\epsilon(u)$ on $\partial\Omega$. Then as in subsection 4.6, H_ϵ diffuses this created vorticity. We shall refer to K_ϵ defined by (1.2) as the vorticity algorithm. For further intuition on this formula, see [44].

2. The consistency of the vorticity algorithm. From Section 5 we know that much more than consistency is needed in the nonlinear case to ensure convergence. However at first sight, one might not think the algorithm (1.2) is consistent because of the more or less "ad hoc" introduction of the vortex sheet. However, from our work in Section 4 we know this *is* consistent. In considering these matters, the choice of function spaces poses a problem. One would like to choose $X=L_2(\Omega)$, $Y=H_0^2(\Omega)\cong D(\Delta_0)$. However, the Euler equations might not be nicely behaved on Y.[13] Instead we shall let $n=2$ or 3 and choose Y as divergence-free vector fields in $W^{2,p}$ and zero on $\partial\Omega$ and $X=J_p$, $p>n$. The Euler operator E_ϵ then maps Y to $W_\sigma^{2,p}$ the space of divergence-free vector fields parallel to the boundary (cf. [16]) and

$$\frac{d}{d\epsilon}E_\epsilon(u)\big|_{\epsilon=0}=-\mathbb{P}(u\cdot\nabla u)\,. \tag{2.1}$$

Then Φ operates on $E_\epsilon(u)$ and H_ϵ maps $\Phi\circ E_\epsilon(u)$ back to Y. Thus K_ϵ maps Y to $W_\sigma^{2,p}$ (for ϵ small).

[13] In two dimensions, the situation might be salvaged in H^2 using the fact that elements of H^2 are quasi-Lipschitz; cf. T. Kato, Arch. Rat. Mech. An. 25, 1967, pp. 188–200. The situation seems rather complicated, however, even here.

6.1. THEOREM. *With this choice of X and Y, the algorithm* (1.2) *is consistent with the Navier–Stokes equations.*

Proof: In Section 4 we saw that

$$\frac{d}{d\epsilon} H_\epsilon \Phi(u)|_{\epsilon=0+} = \mathbb{P}\Delta u,$$

which also holds with X, Y as presently chosen. This together with (2.1) proves consistency.

We have shown the non-trivial fact that the vorticity algorithm is formally correct and that the introduction of the vortex sheet (or vorticity creation operator) is essential for consistency in the usual sense.

3. Stability and convergence of the vorticity algorithm. Stability and the other conditions of Theorem 5.1 (or 5.3) are much more difficult than consistency. If we choose the function space Y to be $\{u \in W^{2,p}(\Omega) \mid u|_{\partial\Omega} = 0, p > n, n = 2 \text{ or } 3\}$ in order that the Euler equations have solutions, then the Stokes part of the algorithm presents problems. The norm of the Hodge projection $\mathbb{P}$ is not one in L_p, $p \neq 2$. It seems unlikely that the Stokes equations generate a quasi-contractive semigroup in L_p Sobolev spaces if $p \neq 2$. Thus, it is imporbable that $\|H_t \circ \Phi(u)\|_{W^{2,p}(\Omega)} \leqq (1+ct)\|u\|_{W^{2,p}(\Omega)}$. It appears, therefore, that in order to prove

$$\|(H_{t/n} \circ \Phi \circ E_{t/n})^n u\|_{W^{2,p}(\Omega)} \leqq C\|u\|_{W^{2,p}}$$

one would not be able to estimate $\|H_{t/n} \circ \Phi \circ E_{t/n} u\|$ and then take the n-th power. Hence, one would have to look directly at the iterates. This seems to be very difficult.

On the other hand, one might consider using a subspace of $H^4(\Omega)$ for Y. This, however, would entail dealing with higher order boundary conditions, which also seems difficult.

7. Stability of Algorithms

1. Introduction. Step-by-step marching methods, such as those previously described, are heavily relied upon in the computer analysis of large-scale, time-dependent systems. Frequently, the original problem to be solved consists of a system of time-dependent, partial differential equations (PDE's) subject to certain initial and boundary conditions. A commonly used technique for spatially discretizing the PDE's is the finite element method, a projection technique involving basis functions of local support. The spatial

discretization leads to a system of ordinary differential equations in the time variable, and it is this system which is approximately integrated by a step-by-step method.

Often, the number of equations involved is in the thousands, and much effort has been expended to develop efficient and reliable techniques. The computer plays a valuable role as an experimental tool in this effort, and there have been many incidences when "peculiarities" of algorithms, which have gone undetected in analysis, have been discovered on the computer, and have ultimately lead to new analytical criteria for evaluating algorithms (see subsection 7.2 below).

Perhaps the most important aspect of a step-by-step method, from a practical standpoint, is its stability characteristics. We take it for granted that any method under consideration is convergent. Unfortunately there are many notions of stability prevalent, which often lead to confusion.

In the remainder of this section we shall discuss what we feel are important stability ideas in some fields of practical interest.

2. Linear structural dynamics. The time-dependent, matrix equation of a linear, elastic structure is

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{F}, \tag{2.1}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{K}$ is the stiffness matrix, $\mathbf{F}$ is the vector of applied forces (a given function of time), $\mathbf{u}$ is the displacement vector, and superposed dots indicate time differentiation. $\mathbf{M}$ and $\mathbf{K}$ are symmetric, $\mathbf{M}$ is positive definite, and $\mathbf{K}$ is positive semi-definite. (Frequently $\mathbf{K}$ is positive definite also.) We wish to think of (2.1) as arising from a finite-element spatial discretization of an elastic continuum or a structural model.

The initial value problem consists of finding a function $\mathbf{u} = \mathbf{u}(t)$ which satisfies (2.1) for all $t \in [0, T]$, $T > 0$, such that

$$\mathbf{u}(0) = \mathbf{d}, \tag{2.2}$$

$$\mathbf{u}(0) = \mathbf{v}, \tag{2.3}$$

where $\mathbf{d}$ and $\mathbf{v}$ are given vectors of initial data. It is well known that the initial value problem is well posed, and furthermore that solutions satisfy the

following energy identity:

$$E(\mathbf{u}(t), \dot{\mathbf{u}}(t)) = E_0 + \int_0^t \dot{\mathbf{u}}^T(\tau)\mathbf{F}(\tau)\, d\tau, \tag{2.4}$$

[14]

where

$$E(\mathbf{u}, \dot{\mathbf{u}}) = T(\dot{\mathbf{u}}) + U(\mathbf{u}), \quad \text{the total energy},$$

$$T(\dot{\mathbf{u}}) = \tfrac{1}{2}\dot{\mathbf{u}}^T\mathbf{M}\dot{\mathbf{u}}, \quad \text{the kinetic energy},$$

$$U(\mathbf{u}) = \tfrac{1}{2}\mathbf{u}^T\mathbf{K}\mathbf{u}, \quad \text{the strain energy},$$

and

$$E_0 = E(\mathbf{d}, \mathbf{v}).$$

If $\mathbf{F} \equiv \mathbf{0}$, total energy is conserved, *viz.*

$$E(\mathbf{u}(t), \dot{\mathbf{u}}(t)) = E_0. \tag{2.5}$$

Roughly speaking, a numerical solution of (2.1) is stable if the rate-of-growth of total energy "approximates" the rate-of-growth indicated by (2.4). An unstable solution is one in which the energy grows too fast, or "blows up".[15]

A concept of stability prevalent in the engineering literature and pertinent to equation (2.1) is "unconditional stability". An unconditionally stable algorithm is one in which the stability condition in question is satisfied independent of the size of the time step taken. On the other hand, a conditionally stable algorithm requires that a time step be taken which is less than a constant times the smallest period of the structure. In complicated structural models, containing slender members exhibiting bending effects, this restriction is a stringent one and often entails using time steps which are much smaller than those needed for accuracy, especially when only low-mode response is of interest. In these cases unconditionally stable algorithms are generally preferred.

To make our ideas more precise we shall consider a particular family of step-by-step methods for solving (2.1), called the Newmark methods [49]:

[14] A superscript T indicates transpose.

[15] We treat stability from scratch in this section. Stability in the previous sections was used only on finite time intervals; here we are concerned with all $t \geq 0$.

Find $\mathbf{d}_n$, $\mathbf{v}_n$ and $\mathbf{a}_n$, $n \in \{0, 1, \cdots, N\}$, such that

$$\mathbf{M}\mathbf{a}_{n+1} + \mathbf{K}\mathbf{d}_{n+1} = \mathbf{F}_{n+1}, \tag{2.6}$$

$$\mathbf{d}_{n+1} = \mathbf{d}_n + \Delta t \mathbf{v}_n + \Delta t^2[(\tfrac{1}{2} - \beta)\mathbf{a}_n + \beta \mathbf{a}_{n+1}], \tag{2.7}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \Delta t[(1-\gamma)\mathbf{a}_n + \gamma \mathbf{a}_{n+1}], \tag{2.8}$$

$$\mathbf{d}_0 = \mathbf{d}, \tag{2.9}$$

$$\mathbf{v}_0 = \mathbf{v}, \tag{2.10}$$

$$\mathbf{a}_0 = \mathbf{M}^{-1}(\mathbf{F}_0 - \mathbf{K}\mathbf{d}_0), \tag{2.11}$$

where N is the number of time steps, $\Delta t = T/N$, $\mathbf{d}_n$, $\mathbf{v}_n$ and $\mathbf{a}_n$ are the approximations to $\mathbf{u}(t_n)$, $\dot{\mathbf{u}}(t_n)$ and $\ddot{\mathbf{u}}(t_n)$, respectively, in which $t_n = n\,\Delta t$, $\mathbf{F}_n = \mathbf{F}(t_n)$, and β and γ are free parameters which govern the stability and accuracy of the methods.

The stability of the preceding algorithms can be ascertained by considering the model equation

$$M\ddot{u} + Ku = 0. \tag{2.12}$$

In this case, (2.6)–(2.11) can be written as

$$\mathbf{X}_{n+1} = \mathbf{A}\mathbf{X}_n, \tag{2.13}$$

where

$$\mathbf{X}_n = (d_n, \Delta t v_n, \Delta t^2 a_n)^T,$$

and $\mathbf{A}$ is called the amplification matrix. Many important properties of an algorithm can be determined by studying the spectral properties of its amplification matrix.

We shall call a matrix such as $\mathbf{A}$ *spectrally stable*, or simply, *stable* if (i) the spectral radius $\rho \leqq 1$ and (ii) eigenvalues of multiplicity greater than one satisfy $|\lambda| < 1$.

The condition of stability that we shall require of the family of algorithms (2.6)–(2.11) is that the associated amplification matrix $\mathbf{A}$ be stable. If $\mathbf{A}$ is stable for all $\Delta t \in [0, \Delta t_c]$, where t_c is a positive constant, stability is said to be *conditional*, whereas if $\mathbf{A}$ is stable for all $\Delta t \in [0, \infty)$, stability is said to be *unconditional*.

The notion of unconditional stability is closely related to so-called "A-stability" [21]. An algorithm is said to be *A-stable* if when applied to $\dot{y} = -\lambda y$,

$\mathcal{R}e\,\lambda > 0$, the solution y_n goes to zero as $n \to \infty$, for all $\Delta t > 0$. Many algorithms which are employed on second-order systems such as (2.1) cannot be directly applied to first-order systems, and thus we retain the use of the terminology "unconditional stability" when speaking of (2.1).

The conditions for stability of the Newmark methods are well known:

$$2\beta \geqq \gamma \geqq \tfrac{1}{2} \quad \text{(unconditional stability)},$$

(2.15)

$$\gamma \geqq \tfrac{1}{2}, \quad 2\beta < \gamma, \quad \tfrac{1}{2}\Delta t^2 < \frac{(M/K)}{(\gamma - 2\beta)} \quad \text{(conditional stability)}.$$

To see what the condition of spectral stability means, let us assume that the eigenvalues of $\mathbf{A}$ are distinct. In this case, $\mathbf{A}$ admits the representation

$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^{-1}, \tag{2.16}$$

where $\mathbf{P}$ is a matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues. Combining (2.16) with (2.13) results in

$$\mathbf{X}_n = \mathbf{A}^n \mathbf{X}_0 = \mathbf{P}\Lambda^n \mathbf{P}^{-1}\mathbf{X}_0. \tag{2.17}$$

Let $\| \ \|$ denote any norm such that $\|\Lambda\| = \rho$. Since we may scale $\mathbf{P}$ in any way we like, let $\|\mathbf{P}\| = 1$. Multiplying (2.17) by $\mathbf{P}^{-1}$ and taking the norm of both sides results in

$$\|\mathbf{P}^{-1}\mathbf{X}_n\| \leqq \rho^n \|\mathbf{P}^{-1}\mathbf{X}_0\|. \tag{2.18}$$

Note that $\mathbf{P}^{-1}\mathbf{X}_n$ are the eigencomponents of $\mathbf{X}_n$. Thus if $\rho = 1$, $\|\mathbf{P}^{-1}\mathbf{X}_n\|$ is uniformly bounded by its initial value; if $\rho < 1$, $\|\mathbf{P}^{-1}\mathbf{X}_n\| \to 0$ as $n \to \infty$. To determine a bound on $\mathbf{X}_n$ we take the norm of (2.17), *viz.*

$$\|\mathbf{X}_n\| \leqq \rho^n \|\mathbf{P}^{-1}\| \|\mathbf{X}_0\|, \tag{2.19}$$

which reveals that $\|\mathbf{P}^{-1}\|$ must be considered when comparing $\|\mathbf{X}_n\|$ with $\|\mathbf{X}_0\|$. Here, if $\rho = 1$, $\|\mathbf{X}_n\|$ is uniformly bounded by $\|\mathbf{P}^{-1}\| \|\mathbf{X}_0\|$, but if $\rho < 1$, then $\|\mathbf{X}_n\| \to 0$ as $n \to \infty$.

A spectrally stable amplification matrix also implies that a conservation law, or growth inequality, exists for the algorithm in question. Let $\mathbf{N}$ denote the symmetric real part of $(\mathbf{P}^{-1})^*\mathbf{P}^{-1}$, so $\mathbf{N}$ is positive definite. Then the spectral stability of $\mathbf{A}$ implies

$$\|\mathbf{X}_{n+1}\|_N \leqq \|\mathbf{X}_n\|_N, \tag{2.20}$$

where

$$\|\mathbf{X}\|_N = (\mathbf{X}^T\mathbf{N}\mathbf{X})^{1/2}.$$

If all the eigenvalues of $\mathbf{A}$ fall on the unit circle, then we have equality in (2.20) (i.e., a conservation law). On the other hand, if $\rho < 1$, then (2.20) is a strict inequality.

Assuming conditions necessary for spectral stability hold (i.e., either of (2.15)), the conservation law/growth inequality for the Newmark methods is

$$E(d_{n+1}, v_{n+1}) + \tfrac{1}{2}\Delta t^2(2\beta - \gamma)T(a_{n+1}) \leqq E(d_n, v_n) + \tfrac{1}{2}\Delta t^2(2\beta - \gamma)T(a_n). \tag{2.21}$$

If $\gamma = \frac{1}{2}$, we have equality in (2.21).

Spectral stability implies the uniform boundedness of $\mathbf{X}_n$, but the bound is dependent upon $\|\mathbf{P}^{-1}\|$. We shall now give an example of a matrix which is spectrally stable, but admits virtually unbounded early growth. Let

$$\mathbf{A} = \begin{bmatrix} \epsilon & k \\ 0 & \epsilon \end{bmatrix},$$

where $0 < \epsilon < 1$ and $k \gg 1$. The spectral radius of $\mathbf{A}$ is ϵ. The effect of the spectral radius is evident from

$$\mathbf{A}^n = \begin{bmatrix} \epsilon^n & n\epsilon^{n-1}k \\ 0 & \epsilon^n \end{bmatrix},$$

i.e., all terms go to zero as $n \to \infty$. However, due to the presence of k, the term $n\epsilon^{n-1}k$ will be very large for small n.

From this example, and the previous discussion, we draw the following conclusions:

1. The long term, or asymptotic behavior of $\mathbf{X}_n$ is governed by the spectral properties of $\mathbf{A}$.
2. The short term behavior of $\mathbf{X}_n$ may be independent of the spectral properties of $\mathbf{A}$.
3. A stable amplification matrix may permit arbitrarily large growth for small n.

In fact there are algorithms in existence which are spectrally stable and accurate (in the technical sense), but exhibit pathological high-frequency growth in the early response. For examples of this behavior, called "overshoot", see [24]. Likewise, there are algorithms which have virtually identical stability and accuracy properties, but preclude any pathological early growth.

The trapezoidal rule is such a method. This algorithm corresponds to $\beta = \frac{1}{4}$ and $\gamma = \frac{1}{2}$ in the Newmark family. The conservation law in this case reduces to

$$E(d_n, v_n) = E_0, \tag{2.22}$$

which makes it clear why no early pathological growth is possible. Spectrally stable algorithms which overshoot possess conservation law/growth inequalities which explicitly involve Δt, see (2.21).

3. Stability in nonlinear problems. In the nonlinear regime the issue of stability is more complicated. Here it is important to study both the stability of solutions with respect to perturbations, and the growth/decay properties of solutions in appropriate norms. In the linear case both of the above considerations are governed by the same equation, and no such distinction need be made.

An example will illustrate these ideas. Consider the nonlinear model equation

$$\ddot{u} + K(u) = 0, \tag{3.1}$$

where

$$K(u) = \begin{cases} 100u & \text{for } |u| \leqq 2, \\ 200 \operatorname{sgn}(u) & \text{for } |u| > 2. \end{cases} \tag{3.2}$$

This is the equation of a nonlinear spring with zero tangent beyond $|u| = 2$. As in the linear case, solutions of (3.1) conserve total energy, i.e.,

$$T[\dot{u}(t)] + \int_d^{u(t)} K(w)\, dw = T(v). \tag{3.3}$$

Let us consider the initial value problem for (3.1)–(3.2) in which $d = 0$ and $v = 25$. The algorithm to be employed is the trapezoidal rule; namely,

$$a_{n+1} + K(d_{n+1}) = 0, \tag{3.4}$$

$$d_{n+1} = d_n + \tfrac{1}{2}\Delta t(v_n + v_{n+1}), \tag{3.5}$$

$$v_{n+1} = v_n + \tfrac{1}{2}\Delta t(a_n + a_{n+1}), \tag{3.6}$$

$$d_0 = d, \tag{3.7}$$

$$v_0 = v, \tag{3.8}$$

$$a_0 = -K(d_0). \tag{3.9}$$

Let us take $\Delta t = 0.2$. Under these circumstances the solution of (3.4)–(3.9) may be easily computed by hand and shown to be periodic, reproducing itself every 24 steps. However, solution on a computer, in which round-off error exists, reveals that the periodic solution eventually breaks down (after about 100 steps) and thereafter linear growth in total energy manifests itself. The growth is quite striking. After 1000 steps the error in energy is anywhere from a factor of 100 to 300 times E_0 depending on the processor and precision involved. (See [26] for further details.) Thus here we have a case in which an *a priori* bound can be given for the solution. However, the solution is unstable with respect to perturbations. This can be ascertained in the usual way, by studying the locally linearized system about the solution of the given equation.

In the linear case, spectral stability enabled us to deduce global conservation laws, or growth inequalities. If we write a nonlinear algorithm in amplification matrix form

$$\mathbf{X}_{n+1} = \mathbf{A}_{n+1}\mathbf{X}_n, \tag{3.10}$$

we can get a *local* conservation law/growth inequality, which depends on the eigenvectors of $\mathbf{A}_{n+1}$ (which, in turn, is a function of $\mathbf{X}_n$ and $\mathbf{X}_{n+1}$). Unfortunately, the global result which manifests the changing of $\mathbf{A}_{n+1}$ from step to step is so crude as to be virtually useless.

However, the spectral stability of $\mathbf{A}_{n+1}$ does tell us one important fact: An instability characteristic of the linear case (i.e., one involving exponential feedback) is impossible. Nevertheless, a weaker instability formed by the biased accumulation of local truncation errors due to nonlinear terms *is* possible. The solution to the initial value problem previously described is such an example.

Because of pathological behavior as indicated above some writers have asserted that there are no unconditionally stable algorithms in the nonlinear regime. However, well-known examples of unconditionally stable algorithms, for particular nonlinear problems, exist. For example, consider the following discrete problem arising in nonlinear heat conduction:

$$\mathbf{C}\dot{\boldsymbol{\theta}} + \mathbf{K}(\boldsymbol{\theta}, t)\boldsymbol{\theta} = \mathbf{0}, \tag{3.11}$$

$$\boldsymbol{\theta}(0) = \mathbf{T}, \tag{3.12}$$

where $\mathbf{C}$ is the constant capacity matrix, $\mathbf{K}(\boldsymbol{\theta}, t)$ is the conductivity matrix, $\boldsymbol{\theta}$ is the temperature vector, and $\mathbf{T}$ is the given initial value. $\mathbf{C}$ and $\mathbf{K}(\boldsymbol{\theta}, t)$ are assumed symmetric and positive definite. It can be easily shown (see [15], [27]) that the midpoint rule, i.e.,

$$\begin{aligned} \mathbf{C}(\mathbf{T}_{n+1} - \mathbf{T}_n) + \Delta t\mathbf{K}(\mathbf{T}_{n+1/2}, t_{n+1/2})\mathbf{T}_{n+1/2} &= \mathbf{0}, \\ \mathbf{T}_0 &= \mathbf{T}, \end{aligned} \tag{3.13}$$

satisfies the following growth inequality

$$\|\mathbf{T}_{n+1}\|_C < \|\mathbf{T}_n\|_C, \tag{3.14}$$

where

$$\|\mathbf{T}\|_C = (\mathbf{T}^T \mathbf{C}\mathbf{T})^{1/2}.$$

Thus the midpoint rule is unconditionally stable for this problem. This result also indicates that small perturbations are not a problem since we have a strict inequality.

The midpoint rule can be applied to the first order form of (3.1). When written in the form (3.10), the spectral stability of the algorithm can be verified. However, solutions which are unbounded in energy exist. To see this consider the initial data $d = -2.5$ and $v = 25$, and $\Delta t = 0.2$. The solution is the same as for the trapezoidal rule with the previously described initial data. (The equivalence of midpoint and trapezoidal rules under a change of data is due to Lindberg and Dahlquist [15].)

It must be mentioned that spectrally stable algorithms seem to work quite well in a majority of large scale computations and, because of this, people frequently speak of a spectrally stable scheme as a stable scheme. Nevertheless, there is some cause for concern due to certain pathological occurrences such as that indicated above. The fact is that in general, we cannot obtain a useful global norm condition on the discrete solution given spectral stability.

4. Energy preserving algorithms. Since it is the global norm condition that we are ultimately after, it has been suggested in [28] that one modify the standard algorithms, such as trapezoidal rule or midpoint rule, so that the global conservation law/growth inequality is satisfied *ab initio*, thereby achieving unconditional stability automatically. For example, the following implicit algorithm may be employed:

$$\mathbf{M}\mathbf{a}_{n+1} + \mathbf{K}(\mathbf{d}_{n+1}) = \mathbf{F}_n, \tag{3.15}$$

$$\mathbf{d}_{n+1} = \mathbf{d}_n + \tfrac{1}{2}\Delta t(\mathbf{v}_n + \mathbf{v}_{n+1}), \tag{3.16}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \tfrac{1}{2}\Delta t\lambda(\mathbf{a}_n + \mathbf{a}_{n+1}), \tag{3.17}$$

$$\lambda = \frac{2[U(\mathbf{d}_{n+1}) - U(\mathbf{d}_n)]}{(\mathbf{d}_{n+1} - \mathbf{d}_n)^T[\mathbf{K}(\mathbf{d}_{n+1}) + \mathbf{K}(\mathbf{d}_n)]}, \tag{3.18}$$

$$\mathbf{d}_0 = \mathbf{d}, \tag{3.19}$$

$$\mathbf{v}_0 = \mathbf{v}, \tag{3.20}$$

where U is the potential which generates $\mathbf{K}$, i.e., $DU=\mathbf{K}$. This algorithm obeys the identity

$$E(\mathbf{d}_{n+1},\mathbf{v}_{n+1})=E(\mathbf{d}_n,\mathbf{v}_n)+\tfrac{1}{2}(\mathbf{d}_{n+1}-\mathbf{d}_n)^T(\mathbf{F}_{n+1}+\mathbf{F}_n), \tag{3.21}$$

as is easily checked (see (3.27) below). Thus when $\mathbf{F}\equiv\mathbf{0}$, total energy is conserved, just as for the exact solution. The algorithm (3.15)–(3.20) is second order accurate, unconditionally stable and reduces to the trapezoidal rule in the linear case. We view it as the appropriate generalization of trapezoidal rule for nonlinear elastodynamics because of the above conservation law. (For related finite difference ideas involving conservative discrete algorithms see Labudde and Greenspan [36] and references therein.)

The energy-preserving algorithm (3.15)–(3.20) can be defined for a general Hamiltonian system (finite or infinite-dimensional) as well,

$$\dot{\mathbf{q}}=\frac{\partial H}{\partial \mathbf{p}},\qquad \dot{\mathbf{p}}=-\frac{\partial H}{\partial \mathbf{q}}, \tag{3.22}$$

by the following implicit scheme

$$\mathbf{q}_{n+1}=\mathbf{q}_n+\Delta t\frac{(H(\mathbf{q}_{n+1},\mathbf{p}_{n+1})-H(\mathbf{q}_{n+1},\mathbf{p}_n))}{\boldsymbol{\lambda}^T(\mathbf{p}_{n+1}-\mathbf{p}_n)}\boldsymbol{\lambda}, \tag{3.23}$$

$$\mathbf{p}_{n+1}=\mathbf{p}_n-\Delta t\frac{(H(\mathbf{q}_{n+1},\mathbf{p}_n)-H(\mathbf{q}_n,\mathbf{p}_n))}{\boldsymbol{\mu}^T(\mathbf{q}_{n+1}-\mathbf{q}_n)}\boldsymbol{\mu}, \tag{3.24}$$

$$\boldsymbol{\lambda}=\frac{\partial H}{\partial \mathbf{p}}(\alpha\mathbf{q}_{n+1}+(1-\alpha)\mathbf{q}_n,\beta\mathbf{p}_{n+1}+(1-\beta)\mathbf{p}_n), \tag{3.25}$$

$$\boldsymbol{\mu}=\frac{\partial H}{\partial \mathbf{q}}(\gamma\mathbf{q}_{n+1}+(1-\gamma)\mathbf{q}_n,\delta\mathbf{p}_{n+1}+(1-\delta)\mathbf{p}_n), \tag{3.26}$$

where α, β, γ, δ are arbitrarily chosen in $[0,1]$.

The proof of conservation of energy is simple: From (3.23), we have

$$(\mathbf{q}_{n+1}-\mathbf{q}_n)^T(\mathbf{p}_{n+1}-\mathbf{p}_n)=\Delta t(H(\mathbf{q}_{n+1},\mathbf{p}_{n+1})-H(\mathbf{q}_{n+1},\mathbf{p}_n)), \tag{3.27}$$

and from (3.24),

$$(\mathbf{p}_{n+1}-\mathbf{p}_n)^T(\mathbf{q}_{n+1}-\mathbf{q}_n)=-\Delta t(H(\mathbf{q}_{n+1},\mathbf{p}_n)-H(\mathbf{q}_n,\mathbf{p}_n)). \tag{3.28}$$

Subtracting (3.28) from (3.27), we obtain

$$H(\mathbf{q}_{n+1},\mathbf{p}_{n+1})=H(\mathbf{q}_n,\mathbf{p}_n). \tag{3.29}$$

The algorithm (3.23)–(3.26) is easily checked to be consistent. Thus the energy preserving property (3.29), if H is related to a norm, will give us stability. In many cases, the methods of Section 5 then give, as a bonus, a unified "algorithmic" approach to existence theorems for weak and strong solutions of Hamiltonian or Hamiltonian-dissipative systems. (For instance, see [7], page 35, [44], §8.)

Bibliography

[1] Askey, R., and Wainger, S., *Mean convergence of expansions in Laguerre and Hermite series*, Am. J. Math. 87, 1965, pp. 695–708.

[2] Batchelor, G. K., *An Introduction to Fluid Mechanics*, Cambridge University Press, 1970.

[3] Brezis, H., *Operateurs Maximaux Monotones et Semi-Groupes de Contractions dan les Espaces de Hilbert*, North Holland, New York, 1973.

[4] Brezis, H., and Pazy, A., *Semigroups of nonlinear contractions on convex sets*, J. Funct. Anal. 6, 1970, pp. 237–281; *Convergence and approximation of semigroups of nonlinear operators in Banach spaces*, J. Funct. Anal. 9, 1972, pp. 63–74.

[5] Chernoff, P., *Note on product formulas for operator semigroups*, J. Funct. Anal. 2, 1968, pp. 238–242.

[6] Chernoff, P., *Product Formulas, Nonlinear Semigroups and Addition of Unbounded Operators*, Memoirs of AMS #140, 1974, (see also Bull. Amer. Math. Soc. 76, 1970, pp. 395–398.)

[7] Chernoff, P., and Marsden, J., *Properties of Infinite Dimensional Hamiltonian Systems*, Springer Lecture Notes, Springer, New York, #425, 1974.

[8] Chorin, A. J., *On the convergence of discrete approximations to the Navier–Stokes equations*, Math. of Comp. 23, 1969, pp. 341–353.

[9] Chorin, A. J., *Numerical Study of Slightly Viscous Flow*, Journal of Fluid Mechanics, 57, 1973, pp. 785–796.

[10] Chorin, A. J., *Lectures on Turbulence Theory*, Publish/Perish, Boston, 1975.

[11] Chorin, A. J., *Numerical Methods in Statistical Hydrodynamics*, University of Montreal Press, 1976.

[12] Chorin, A. J., *Crude numerical approximation of turbulent flow*, pp. 165–175 of *Numerical Solution of Partial Differential Equations*, III, Academic Press, New York, 1976.

[13] Ciarlet, P. G., and Raviart, P. A., *General Lagrange and Hermite interpolation in R^n with applications to finite element methods*, Arch. Rat. Mech. Anal. 46, 1972, pp. 177–199.

[14] Courant, R., Friedrichs, K., and Lewy, H., *Über die partiellen Differenzengleichungen der Mathematischen Physik*, Math. Ann. 100, 1928, pp. 32–74.

[15] Dahlquist, G., and Lindberg, B., *On some implicit one-step methods for stiff differential equations*, Report No. TRITA-NA7302, Dept. of Information Processing, The Royal Institute of Technology, Stockholm.

[16] Ebin, D., and Marsden, J., *Groups of diffeomorphisms and the motion of an incompressible fluid*, Ann. of Math., 92, 1970, pp. 102–163.

[17] Faris, W., *The product formula for semi-groups defined by Friedrichs extensions*, Pacific J. Math. 21, 1967, pp. 47–70.

[18] Faris, W., *Product formulas for perturbations of linear propagators*, J. Funct. Anal. 1, 1967, pp. 93–180.

[19] Faris, W., *Self-Adjoint Operators*, Springer Lecture Notes, #433, Springer, New York, 1975.

[20] Friedman, C. N., *Semigroup product formulas, compressions and continual observations in quantum mechanics*, Indiana Math. J. 21, 1972, pp. 1001–1011.

[21] Gear, C. W., *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice Hall, Englewood Cliffs, N.J., 1971.

[22] Goldstein, J. A., *A Lie product formula for one parameter groups of isometries on Banach spaces*, Math. Ann. 186, 1970, pp. 299–306.
[23] Herring, J. R., Orszag, S. A., Kraichnan, R. H., and Fox, D. A., *Decay of two-dimensional homogeneous turbulence*, J. Fluid Mech. 66, 1974, pp. 417–444.
[24] Hilber, H. M., and Hughes, T. J. R., *Collocation, dissipation and 'Overshoot' for time integration schemes in structural dynamics*, Earthquake Engineering and Structural Dynamics (in press).
[25] Hille, E., and Philips, R. S., *Functional Analysis and Semi-Groups*, AMS Colloquium Publ. 31, Providence, R.I., 1957.
[26] Hughes, T. J. R., *Stability, convergence and growth and decay of energy of the average acceleration method in nonlinear structural dynamics*, Computers and Structures 6, 1976, pp. 313–324.
[27] Hughes, T. J. R., *Unconditionally stable algorithms for nonlinear heat conduction*, Computational Methods in Applied Mechanics and Engineering 10, 1977, pp. 135–139.
[28] Hughes, T. J. R., and Caughey, T. K., *Finite element methods for nonlinear elastodynamics which conserve energy* (to appear).
[29] Hughes, T. J. R., Kato, T., and Marsden, J., *Well-posed quasi-linear second-order hyperbolic systems with applications to elastodynamics and general relativity*, Arch. Rat. Mech. Anal. 63, 1977, pp. 273–294.
[30] Kato, T., *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, New York, 1966.
[31] Kato, T., *Linear evolution equations of "hyperbolic type"*, I, J. Fac. Sci. Univ. Tokyo 17, 1970, pp. 241–258.
[32] Kato, T., *Linear evolution equations of "hyperbolic type"*, *II*. J. Math. Soc. Japan 25, 1973, pp. 648–666.
[33] Kato, T., *On the Trotter-Lie product formula*, Proc. Japan Academy 50, 1974, pp. 694–698.
[34] Kato, T., *Trotter's product formula for an arbitrary pair of selfadjoint contraction semigroups* (preprint).
[35] Kreiss, H.-O., *On difference approximations of the dissipative type for hyperbolic differential equations*, Comm. Pure Appl. Math. 17, 1964, pp. 335–353.
[36] LaBudde, R. A., and Greenspan, D., *Energy and momentum conserving methods of arbitrary order for the numerical integration of equations of motion*, Part 1; *Motion of a single particle*, Part II; *Motion of a system of particles*, Numer. Math. 25, 1976, pp. 323–346; 26, pp. 1–16.
[37] Lax, P. D., and Richtmyer, R. D., *Survey of the stability of linear finite difference equations*, Comm. Pure Appl. Math., 9, 1956, pp. 267–293.
[38] Lie, S., and Engel, F., *Theorie der Transformationsgruppen*, 3 Vols., Teubner, Leipzig, 1888.
[39] Lighthill, M. J., *Introduction to boundary layer theory*, in *Laminar Boundary Layers*, L. Rosenhead (ed.), Chapter II, Oxford Univ. Press, 1963.
[40] Lions, J. L., *Quelques Méthodes de Résolution des Problemes aux Limites non Linéaires*, Dunod, Gauthier-Villars, Paris, 1969.
[41] Lusternik, L. A., and Sobolev, V. J., *Elements of Functional Analysis*, Halsted Press, Delhi, Hindustan, 1974.
[42] Marsden, J., *On product formulas for nonlinear semi-groups*, J. Funct. Anal. 13, 1973, pp. 51–72.
[43] Marsden, J., *A formula for the solution of the Navier-Stokes equations, based on a method of Chorin*, Bull. of Am. Math. Soc. 80, 1974, pp. 154–158.
[44] Marsden, J., *Applications of Global Analysis in Mathematical Physics*, Publish/Perish, Boston, 1974.
[45] Marsden, J. Ebin, D., and Fischer, A., *Diffeomorphism groups, hydrodynamics and relativity*, Proc. 13th Bienniel Seminar of Canadian Math. Congress, 1972, pp. 139–279.
[46] Nelson, E., *Feynmann integrals and Schrödinger equation*, J. Math. Phys. 5, 1964, pp. 332–343.

[47] Nelson, E., *Topics in Dynamics, Flows*, I, Princeton University Press, Princeton, N.J., 1969.
[48] von Neumann, J., and Richtmyer, R. D., *A method for the numerical calculations of hydrodynamical shocks*, J. Appl. Phys. 21, 1950, p. 232.
[49] Newmark, N. M., *A method of computation for structural dynamics*, J. of the Engineering Mechanics Division, ASCE, 1959, pp. 67–94.
[50] Oden, J. T., and Reddy, J. N., *The Mathematical Theory of Finite Elements*, Wiley-Interscience, New York, 1976.
[51] Orszag, S. A., *Numerical simulation of the Taylor-Green vortex, Computing Methods in Applied Science and Engineering*, ed. R. Glowinski and J. L. Lions, Springer, New York, 1974.
[52] Raviart, P. A., *Sur l'approximation de certaines équations d'évolution linéares et non linéarires*, J. Math. Pures et Appliquees, 46, 1967, pp. 11–107; 109–183.
[53] Reed, M., and Simon, B., *Methods of Modern Mathematical Physics, Vol. I, Functional Analysis, Vol. II, Fourier Analysis, Self-Adjointness*, Academic Press, New York, 1972, 1975.
[54] Richtmyer, R. D., andM orton, K. W., *Difference Methods for Initial Value Problems*, 2nd Ed., Wiley-Interscience, New York, 1967.
[55] Segal, I., *Nonlinear semigroups*, Ann. of Math. 27, 1963, pp. 339–364.
[56] Segal, I., *The global Cauchy problem for a relativistic scalar field with power interaction*, Bull. Soc. Math. France, 91, 1963, p. 129.
[57] Simon, B., *Quantum Mechanics for Hamiltonians Defined As Quadratic Forms*, Princeton Series in Physics, Princeton University Press, Princeton, N.J., 1971.
[58] Semenov, Yu. A., *The product formula for semigroups and its application to the Schrödinger equation with singular potentials*, preprint, ITP-72-36E, Inst. for Theor. Phys. Acad. Sci. Ukranian S.S.R., Kiev, 1972.
[59] Strang, G., *Accurate partial difference methods, I, Linear Cauchy problems*, Arch. Rat. Mech. Anal. 12, 1963, p. 392 and *II, Nonlinear problems*, Numer. Math. 6, 1964, p. 37.
[60] Strang, G., *Approximating semigroups and the consistency of difference schemes*, Proc. Am. Math. Soc. 20, 1969, pp. 1–7.
[61] Strang, G., and Fix, G. J., *An Analysis of the Finite Element Method*, Prentice Hall, Englewood Cliffs, N.J., 1973.
[62] Taylor, G. I., and Green, A. E., *Mechanism of the production of small eddies from large one*, Proc. Roy. Soc. A158, 1937, pp. 499–521.
[63] Trotter, H. F., *Approximation of semi-groups of operators*, Pacific J. Math. 8, 1958, pp. 887–919.
[64] Trotter, H. F., *On the product of semi-groups of operators*, Proc. Am. Math. Soc. 10, 1959, pp. 545–551.
[65] Webb, G. F., *Continuous nonlinear perturbations of linear accretive operators in Banach spaces*, J. Funct. Anal. 10, 1972, pp. 191–203.
[66] Wellford, L. C., and Oden, J. T., *On some finite element methods for certain nonlinear second-order hyperbolic equations*, TICOM Report No. 74, 1974.
[67] Yosida, K., *Functional Analysis*, 4th Ed., Springer, New York, 1974.
[68] Pazy, A., *A Trotter-Lie formula for compactly generated semigroups of nonlinear operators*, J. Funct. Anal. 23, 1976, pp. 353–361.
[69] Weissler, F., *Construction of nonlinear semi-groups using product formulas* (preprint).
[70] Crandall, M., and Pazy, A., *Nonlinear evolution equations in Banach spaces*, Israel J. Math. 11, 1972, pp. 57–94.
[71] Douglis, A., *An approximate layering method for multi-dimensional nonlinear parabolic systems of a certain type*, University of Maryland Technical Report #TR 77–38.
[72] Evans, L. C., *Nonlinear evolution equations in an arbitrary Banach space*, Israel J. Math., to appear.
[73] Kato, T., and Masuda, K., *Trotter's product formula for nonlinear semigroups generated by the subdifferentials of convex functionals* (to appear).

Received May, 1977.