IMA Journal of Numerical Analysis (1984) **4**, 109–115

Equivalence Theorems for Incomplete Spaces: An Appraisal

C. PALENCIA AND J. M. SANZ-SERNA
Departamento de Ecuaciones Funcionales, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain

[Received 24 May 1983]

We present a practical appraisal of the functional analysis ideas involved in some recent equivalence theorems. We also derive some new results which may be of independent interest.

1. Introduction

THE CELEBRATED Lax Equivalence Theorem (Lax & Richtmyer, 1956) provided, together with Dahlquist's (1956) theory on linear multistep methods, one of the first cornerstones of the analysis of numerical methods. The results of Lax & Richtmyer have been generalized in several directions (cf. Ansorge, 1977, 1978; Palencia & Sanz-Serna, 1983). In particular, and not surprisingly, much effort has been expended in extending the theory so as to cover *non-linear* problems (Ansorge, 1977, 1978; Rosinger, 1980, 1982). Unfortunately, the definitions employed in the most recent of these general theories rely heavily on functional analysis ideas, thus making it difficult for the average numerical analyst to appreciate their scope.

The aim of this note is to help the interested reader in grasping the significance of these recent contributions. We also present some new results and compare some existing equivalence theorems, as applied to *linear* problems. (Clearly, a self-respecting *non-linear* theory must be able to cope with *linear* problems.) One of our conclusions will be that much of the material in Rosinger, 1980, 1982, although useful in the treatment of non-linear ordinary differential or integral equations, suffers from serious drawbacks when dealing with partial differential equations.

In the interest of clarity, we have postponed the proofs to an appendix.

2. Linear Cauchy Problems

We consider linear Cauchy problems

$$\begin{aligned} du/dt &= Au, \quad 0 \leqslant t \leqslant T < \infty \\ u(0) &= u_0. \end{aligned} \tag{1}$$

Here A is a linear operator mapping a subspace $D(A)$ of a normed space X into X, and u_0 an element in X. In applications to initial-value problems in partial differential equations, X is a space consisting of scalar or vector valued functions of

0272–4979/84/010109+07 $03.00/0

one or several independent variables $x_1, x_2, \ldots$ These are called space variables to distinguish them from the "temporal" variable t. The norm in X is typically one of the familiar L^p norms or, at least, is related to the L^p norms. The operator A is a linear differential operator involving the derivatives $\partial/\partial x_1, \partial/\partial x_2, \ldots$ Usually Au is not defined for every function u in X, as it is required that u be smooth enough to guarantee that, after performing the differentiations involved in A, the result Au lies in X. Therefore, the domain $D(A)$ is often much smaller than X. However, in practice, $D(A)$ is dense in X, i.e. every element in X is the limit of a sequence of elements in A. Hereafter, we always assume that $D(A)$ is dense in X.

In this paper X is not necessarily a complete or Banach space. Frequently, in practice, X is an L^p space and therefore complete. Now, functions in L^p may include (i) "smooth" functions, (ii) some non-smooth functions with a physical meaning (such as step-functions), (iii) garbage without any conceivable physical significance which is there just in order to make the space complete and therefore easier to handle mathematically. Thus we have preferred to include the possibility of an incomplete X, which would arise if the attention were restricted to smooth functions. Nevertheless, every normed space X can be embedded in a complete normed space $\hat{X}$, whose elements are limits of sequences in X. In practice $\hat{X}$ is usually an L^p space.

We have not assumed that A is *bounded*. In fact in applications to PDEs, A is *never* bounded. We recall that if A were bounded and X a Banach space, then (1) would have a unique solution for every initial datum u_0 in $D(A)$, given by the formula $u(t) = \exp(At)u_0$, with $\exp(At) = I + At + \frac{1}{2}A^2t^2 + \ldots$ (see Aubin, 1979, p. 333). However, for unbounded A, the well-posedness of (1) is not guaranteed. We *assume* that (1) has a unique solution $u(t)$ for each u_0 in $D(A)$. (For a precise definition of the term solution, see Richtmyer & Morton, 1967, p. 40.) As a consequence of the linearity of the problem the operator $E_0(t)$ which maps the initial datum u_0 into the value $u(t)$ of the corresponding solution is linear. We next *assume* that, for $0 \leqslant t \leqslant T$, the operator $E_0(t)$ is bounded, i.e. small changes of u_0 induce small changes in $u(t)$ and, furthermore, we *suppose* that the family of operators $\{E_0(t) : 0 \leqslant t \leqslant T\}$ is equicontinuous, i.e. there exists a positive constant C, so that $\|E_0(t)\| \leqslant C$, uniformly in t. The last hypothesis is identical with the requirement that changes in the solution $u(t)$ at time t can be bounded, *uniformly in t*, in terms of the corresponding changes in the initial datum u_0.

So far we have been dealing with initial data in $D(A)$. Often u_0 does not lie in $D(A)$. (In practical applications u_0 may fail to fulfil the smoothness requirements for Au_0 to be defined or to lie in X.) It is obvious that if u_0 does not belong to $D(A)$, then (1) has no (genuine) solution. Nevertheless, each u_0 is the limit of a sequence $u_0^{(n)}$ in $D(A)$. If the corresponding sequence of solutions $u^{(n)}(t)$ has a limit $u(t)$ in X, then it is natural to refer to $u(t)$ as a *generalized* solution to (1). Had X been assumed complete, then the existence of the limit $u(t)$ (i.e. of the generalized solution to (1)), would have been guaranteed (Richtmyer & Morton, 1967, p. 41). Here X is not in general complete and we make the final hypothesis that (1) has a unique generalized solution $u(t)$ for each u_0 in X. It is then possible to show that if we denote by $E(t)$ the operator which maps the initial datum u_0 into the corresponding (generalized) solution $u(t)$, then the family $\{E(t) : 0 \leqslant t \leqslant T\}$ is also equicontinuous. We emphasize that for u_0 in X, $E(t)u_0$ denotes the value at time t of the generalized solution to (1).

When u_0 is smooth enough to lie in $D(A)$, then $E(t)u_0 = E_0(t)u_0$ is a genuine solution of (1). Finally, it should be pointed out that under the hypotheses above, (1) has also a generalized solution $u(t)$ in $\hat{X}$ for each u_0 in the completion $\hat{X}$.

3. Difference Methods

A discretization of (1) replaces the function $u(t)$ by a sequence of points $u^0, u^1, \ldots$, where u^n is supposed to approximate $u(nh)$, $h = \Delta t$ being a small increment. The points u^n are recursively computed from the finite-difference equations

$$u^{n+1} = C(h)u^n. \tag{2}$$

Here $C(h)$ is, for each h in an appropriate interval $0 < h < \hat{h} \leqslant T$, a bounded linear operator in X. It is assumed that the following (mild) requirement holds

$$\sup \{\|C(h)\| : 0 < h < \hat{h}\} < +\infty. \tag{3}$$

The formulation (2) includes both explicit and implicit methods, provided that the latter are formally written as if the equations for u^{n+1} had been solved. However, for simplicity, we shall only be concerned with one-step methods such as (2).

The extension theorem (Richtmyer & Morton, 1967) guarantees that the finite-difference scheme $C(h)$ may also be applied to elements u in the completion $\hat{X}$, i.e. to non-smooth, functions. We now write down, for future reference, the familiar definitions of consistency, stability and convergence (Richtmyer & Morton, 1967).

The method (2) is said to be consistent if there exists a dense subset Y of $D(A)$ such that for each u_0 in Y, and each t, $0 \leqslant t \leqslant T$

$$\lim_{h \to 0} h^{-1}\|E(t+h)u_0 - C(h)E(t)u_0\| = 0. \tag{4}$$

Note that the expression in the norm is the difference between the theoretical solution $E(t+h)u_0$ at time $t+h$ and the result of one step of the numerical procedure taken from the theoretical solution at the previous time-level, i.e. the local truncation error. We stress that (4) is only supposed to hold in a *subset* Y of $D(A)$. Typically, consistency is investigated by a Taylor expansion of the truncation error and this requires u_0 to be rather smooth; then Y is the set of functions which satisfy the appropriate smoothness requirements. In fact, it is easy to see that in practical situations (4) does *not* hold if u_0 is, say, a step-function.

The method is stable if the set of operators $C(h)^n$, $0 < h < \hat{h}$, $0 \leqslant nh \leqslant T$ is uniformly bounded. Rosinger (1980) uses the following alternative definition.

Definition 3.1. The difference scheme (2) is stable if for any compact set $K \subset X$, there exists $L(K)$ such that for $0 < h < \hat{h}$, $0 \leqslant nh \leqslant T$, $u \in K$

$$\|C(h)^n u\| \leqslant L(K)\|u\|.$$

This definition is equivalent to the usual one. (See the Appendix.)

The method is convergent if for each t, $0 \leqslant t \leqslant T$, each t-admissible pair of sequences $n_1, n_2, \ldots, n_j, \ldots, h_1, h_2, \ldots, h_j, \ldots$ and each u_0 in X

$$\lim_{j \to \infty} \|C(h_j)^{n_j}u_0 - E(t)u_0\| = 0. \tag{5}$$

In this paper a pair of sequences (n_j), (h_j) is said to be t-admissible if

$$0 < h_j < \hat{h}, \quad 0 \leqslant n_j h_j \leqslant T, \lim_j n_j h_j = t,$$

and each n_j is an integer.

We emphasize that convergence, unlike consistency, refers to all u_0 in X. If X is an L^p space this means that convergent methods can cope even with pathological initial data.

4. Equivalence Theorems

We begin with the classical Lax theorem.

THEOREM 4.1 (Lax). *Let X be a Banach space (i.e. $X = \hat{X}$), and the method* (2) *consistent. Then* (2) *is convergent if and only if it is stable.*

We remark (Richtmyer & Morton, 1967) that the proof of the implication stable $\Rightarrow$ convergent is rather elementary and does not require the completeness of X. On the other hand, the implication convergent $\Rightarrow$ stable invokes the principle of uniform boundedness, a deep result needing the completeness of X. Then it is fair to say that in Lax's theory, stability is necessary for convergence only because the latter is demanded for every u_0 in $\hat{X}$, i.e. even for pathological initial data. We shall address later the study of equivalence theorems that do not require the completeness of X.

We now note that, upon using some functional analysis (Schaefer, 1971, Chapter 3, Section 4), if a method is stable and convergent, then the limit (5) is uniform for u_0 ranging in a compact $K \subset X$. We then have the following result, which appears to be new.

THEOREM 4.2 *Let X be a Banach space (i.e. $X = \hat{X}$) and the method* (2) *consistent. Then* (2) *is convergent* uniformly in compact sets *if and only if it is stable.*

Compact sets in functional spaces are rather elusive creatures. In particular, balls are *never* compact. Thus Theorem 4.2, as it stands, is not too useful in practice. However, (5) holds uniformly in compact sets if and only if the following property holds: for each t, $0 \leqslant t \leqslant T$, each t-admissible pair of sequences (n_j), (h_j), each u_0 in X and each sequence $u_0^{(j)}$ converging to u_0,

$$\lim_{j \to \infty} \|C(h_j)^{n_j} u_0^{(j)} - E(t) u_0\| = 0. \tag{6}$$

Property (6) has been called L-convergence by Ansorge (1977). Whilst it is equivalent to uniform convergence in compact sets (this equivalence is proved in the Appendix), it has a richer numerical meaning. It demands that numerical solutions approach the theoretical solutions even if the former start from perturbed initial data. Clearly, a convergent method may fail to have interest if it were not L-convergent.

We then rephrase Theorem 4.2 as follows.

THEOREM 4.2′ *Let X be a Banach space (i.e. $X = \hat{X}$) and the method* (2) *consistent. Then* (2) *is L-convergent if and only if it is stable.*

This theorem is proved in Richtmyer & Morton (1967, ch. 7).

We now turn to the question of whether the theorems above hold when X is not complete. Lax's theorem does not: a counter-example is provided in Thomee (1969, Theorem 3.1). This means that if attention is restricted to suitably smooth initial data one may have convergence without stability. In fact, the literature (Thomee, 1969; Richtmyer & Morton, 1967) contains several relaxations of the concept of stability which still lead to convergence for smooth functions. Instances of convergence without stability are sometimes regarded as deprived of practical interest. Such a view is often backed by considering the effect of round-off errors. However, we shall see next that convergence without stability implies convergence without L-convergence. In fact, Theorems 4.2, 4.2′ hold for incomplete X, as in the following.

THEOREM 4.3 *Let X be a normed space. Then for a consistent method, stability, L-convergence, and uniform convergence in compact sets are equivalent.*

The proof of this result is analogous to that of Theorem 4.2′. (See the Appendix.) The equivalence between stability and L-convergence was first noted by Spijker (1968). Rosinger (1980, 1982) has investigated the possibility of deriving equivalence theorems without the assumption of completeness. Although his results hold even for *non-linear* problems, it is useful to investigate their scope as applied to *linear situations*, where his best theorem (Rosinger, 1982, ch. 3) reads as follows.

THEOREM 4.4 (Rosinger). *Let X be a normed space and suppose that the method* (2) *is consistent uniformly in compact sets* (*i.e. the limit* (4) *is uniform for u_0 in a compact subset of Y*). *Then* (2) *is convergent uniformly in compact sets if and only if it is stable.*

While the conclusions in Theorem 4.4 are the same as those in Theorem 4.3, the hypotheses in Theorem 4.4 are much stronger than those in Theorem 4.3. In fact, we prove in the Appendix that:

(i) Uniform consistency in compact sets implies *stability*. Therefore, Theorem 4.4 applying only to methods consistent uniformly in compact sets does not deserve the title of "equivalence" theorem: one of its implications is tautological.

(ii) Uniform consistency in compact sets implies that (4) holds for every u_0 in the completion $\hat{X}$. Thus, none of the usual discretizations of PDEs can satisfy the hypotheses of Theorem 4.4

An equivalence theorem similar to Theorem 4.3 has been given by Hass (Ansorge, 1978).

All the results quoted in this section can be extended to cover the case where the theoretical solution $u(t)$ and the numerical approximation u^n lie in different spaces (for instance, if u^n is a grid function) (see Palencia & Sanz-Serna, 1983).

REFERENCES

ANSORGE, R. 1977 Survey of equivalence theorems in the theory of difference approximations for partial differential equations. In *Topics in Numerical Analysis* III (Miller, J. J. H., Ed.). London: Academic Press.

ANSORGE, R. 1978 *Differenzenapproximationen partieller Anfangwertaufgaben*. Stuttgart: Teubner.

AUBIN, J. P. 1979 *Applied Functional Analysis*. New York: John Wiley.

DAHLQUIST, G. 1956 Convergence and stability in the numerical integration of ordinary differential equations. *Math. Scand.* **4,** 33–53.
LAX, P. D. & RICHTMYER, R. D. 1956 Survey of stability of linear finite difference equations. *Comm. Pure appl. Math.* **9,** 267–293.
PALENCIA, C. & SANZ-SERNA, J. M. 1983 An extension of the Lax–Richtmyer theory. (Submitted).
RICHTMYER, R. D. & MORTON, K. W. 1967 *Difference Methods for Initial Value Problems.* New York: Interscience.
ROSINGER, E. E. 1980 Stability and convergence for nonlinear difference schemes are equivalent. *J. Inst. Maths Applics* **26,** 143–149.
ROSINGER, E. E. 1982 *Nonlinear Equivalence Reduction for PDEs and Fast Convergence Numerical Methods.* Boston: Pitman.
SCHAEFER, H. H. 1971 *Topological Vector Spaces,* 3rd Edition. Berlin: Springer-Verlag.
SPIJKER, M. N. 1968 *Stability and Convergence of Finite-difference Methods.* Leiden: Rijkuniversiteit.
THOMEE, V. 1969 Stability of partial difference operators. *SIAM Rev.* **11,** 152–195.

Appendix: Some Proofs

(1) We first prove the equivalence between the usual definition of stability and that given in Definition 3.1. Clearly, a Lax-stable method is stable in the sense of Definition 3.1. If a method is not Lax-stable, there exist sequences n_j, h_j and u_j with $\|u_j\| \leqslant 1$ such that $n_j h_j \leqslant T$ and

$$d_j = \|C(h_j)^{n_j} u_j\| \to \infty\cdot$$

Then the elements $d_j^{-\frac{1}{2}} u_j$ together with the zero constitute a compact set for which the bound in Definition 3.1 cannot hold.

(2) It is easy to prove that uniform convergence in compact sets implies L-convergence since a convergent sequence together with its limit is a compact set. The proof of the converse uses *reductio ad absurdum.*

(3) Turning now to Theorem 4.3, the implication stability $\Rightarrow$ L-convergence is proved as in Richtmyer & Morton (1967), Section 7.3. If the method is not stable, then there exist sequences n_j, h_j, u_j as in (1) above. Then $d_j^{-\frac{1}{2}} u_j$ approaches zero and one concludes that L-convergence cannot take place just as in Richtmyer & Morton (1967, Section 7.3).

(4) If (4) holds uniformly in compacts of Y, then the operators $S(h) = h^{-1}(C(h) - E(h))$ are convergent to zero uniformly in compacts of Y, and therefore an *equicontinuous* family. Then $S(h) \to 0$ uniformly in compact sets of $\hat{X}$ (Schaefer, 1971, ch. 3, Section 4) and, in particular, (5) holds for every u_0 in $\hat{X}$. Furthermore, let C be a uniform bound for the norms $\|E(t)\|$, $0 \leqslant t \leqslant T$ and B a uniform bound for the norms $\|S(h)\|$, $0 < h < \hat{h}$. We also introduce the notation

$$\alpha_n(h) = \max\{|C(h)^k\| : k = 0, 1, \ldots, n\}.$$

Consideration of the identity

$$C(h)^n = E(nh) + \sum_{k=0}^{n-1} C(h)^k [C(h) - E(h)] E((n-1-k)h)$$

leads to

$$\|C(h)^n\| \leqslant C + n\alpha_n(h) BhC$$

which, in turn, implies ($\alpha_n(h)$ is monotonic with respect to n)

$$\alpha_n(h) \leqslant C + n\alpha_n(h)BhC.$$

Thus, for n, h fulfilling the condition $nh \leqslant 1/(2BC)$

$$\alpha_n(h) \leqslant 2C,$$

or, in other words, the method is stable for $0 \leqslant t \leqslant 1/(2BC)$. The interval $0 \leqslant t \leqslant T$ can be covered by a finite number of intervals of length $1/(2BC)$, and therefore the method is stable. We note that the semigroup property $E(t+s) = E(t)E(s)$ has been used. This property follows easily from the assumption that (1) has a unique solution for each u_0 in $D(A)$.