

A General Equivalence Theorem in the Theory of Discretization Methods

By J. M. Sanz-Serna and C. Palencia

Abstract. The Lax-Richtmyer theorem is extended to work in the framework of Stetter's theory of discretizations. The new result applies to both initial and boundary value problems discretized by finite elements, finite differences, etc. Several examples are given, together with a comparison with other available equivalence theorems. The proof relies on a generalized Banach-Steinhaus theorem.

1. Introduction. In this paper we extend the classical Lax-Richtmyer equivalence theorem [6], so as to cover in a simple way not only initial value problems, but also boundary value problems, mixed problems, etc. Our theory relies on a generalized Banach-Steinhaus theorem [9] and works (essentially) in the framework of Stetter [13]. This set-up employs restriction operators to compare the true and discretized solutions, as distinct to those theories which use prolongation operators. (One of the oldest prolongation theories is probably that of Aubin, summarized in [16].) Our main result is given in Section 2. Sections 3 and 4 are devoted to examples and counterexamples. The former are meant to show the scope of our result and include the Galerkin method for boundary value elliptic problems and semidiscrete and fully discrete schemes for initial value problems. The counterexamples prove that the present hypotheses cannot be dispensed with. In particular, we show that a method which is consistent and convergent for all data in a Banach space may be unstable. The final section contains a comparison with other available equivalence theorems.

2. An Equivalence Theorem.

2.1. *The True Problem.* Let X (the space of solutions) and Y (the space of data) be normed spaces, both real or both complex. We consider a linear operator A with domain $D \subset X$ and range $R \subset Y$. The problems to be solved are of the form

$$(2.1) \quad Au = f, \quad f \in Y.$$

Here A is not assumed to be bounded, so that unbounded differential operators are included. We suppose that problem (2.1) is well-posed in the following sense: The range R of A is dense in Y , and there exists a *bounded* linear operator $E \in B(Y, X)$ such that the composition EA is the identity in D . Note that this implies that, for $f \in R$, Eq. (2.1) has the unique solution $u = Ef$ and that solutions depend continuously on the data. When $f \in Y, f \notin R$, Eq. (2.1) has no solution, and Ef can be

Received March 15, 1984.

1980 *Mathematics Subject Classification.* Primary 65J10; Secondary 65L07, 65M10, 65N10.

©1985 American Mathematical Society
0025-5718/85 \$1.00 + \$.25 per page

regarded as a generalized solution, since E is the unique bounded extension to Y of $A^{-1}: R \rightarrow D$ (see [8] for a discussion).

2.2 *The Approximate Problems.* Let H be a set of positive numbers such that 0 is the unique limit point of H . For each $h \in H$, let X_h, Y_h be normed spaces and consider the approximate or discretized problem

$$(2.2) \quad A_h u_h = f_h, \quad f_h \in Y_h,$$

where A_h is a linear operator $A_h: X_h \rightarrow Y_h$. We assume that for each $h \in H$, problem (2.2) is well-posed in the sense of the previous paragraph, with solution operator $E_h = A_h^{-1}$. In practice X_h, Y_h are subspaces of X, Y or spaces of grid functions, etc. (see [9] for a discussion of the various possibilities). In order to relate the true solutions u and data f , which lie in X, Y , with the approximate solutions u_h and data f_h , which lie in X_h, Y_h , we introduce restriction operators r_h, s_h as follows [9]. For each $h \in H$, r_h (resp. s_h) is a bounded, linear operator from X (resp. Y) into X_h (resp. Y_h). We assume that the operator norms can be bounded

$$(2.3) \quad \|r_h\| \leq C_1, \quad \|s_h\| \leq C_2,$$

with C_1, C_2 independent of h . We shall compare the true solution $u = Ef$ with the discrete solution $u_h = E_h s_h f$ corresponding to the discretized datum f . This comparison is achieved by measuring the distance in X_h between u_h and the restriction $r_h u$. (Some authors prefer to measure in X the distance between u and some sort of prolongation of u_h [9].)

The family $(X_h, Y_h, A_h, r_h, s_h)$ defines a *method* for the solution of (2.1).

2.3. *Convergence, Stability, Consistency.* Let f be a given element in Y . We say that the method $(X_h, Y_h, A_h, r_h, s_h)$ is *convergent for the problem* (2.1) if

$$(2.4) \quad \lim_h \|r_h E f - E_h s_h f\|_{X_h} = 0.$$

We say that the method is *convergent* if it is convergent for each problem (2.1) as f ranges in Y .

Let u be a given element in D . We say that the method is *consistent at u* if

$$(2.5) \quad \lim_h \|A_h r_h u - s_h A u\|_{Y_h} = 0.$$

A method is *consistent* if it is consistent at each u in a set D_0 such that the image $A(D_0)$ is dense in Y . (We recall that it is not appropriate to demand consistency at each u in the domain of A ; cf. [8].)

Finally, the method is *stable* if a constant K exists such that

$$(2.6) \quad \|E_h\|_{B(X_h, Y_h)} \leq K.$$

Note that stability depends only on X_h, Y_h, A_h and does not relate to (2.1) or to r_h, s_h .

The quantities within the norms in (2.4), (2.5) are, respectively, the global and local discretization errors.

2.4. *The Equivalence Theorem.*

THEOREM 1. *Let $X, Y, A, X_h, Y_h, A_h, r_h, s_h$ be as above.*

(i) *If the method is consistent and stable, then it is convergent.*

(ii) *If the method is convergent, then it is stable provided that Y is a Banach space and that the following condition holds:*

(P) *There exists a constant L such that, for each $h \in H$ and each $g \in Y_h$ with $\|g\| \leq 1$, there exists an element $f \in Y$ such that $\|f\| \leq L$ and $s_h f = g$.*

Proof. (i) Let $f \in A(D_0)$. The convergence for the problem $Au = f$ follows upon using (2.5) in the bound

$$\|r_h E f - E_h s_h f\| = \|E_h(A_h r_h u - s_h A u)\| \leq K \|A_h r_h u - s_h A u\|.$$

If $f \in Y, f \notin A(D_0)$, we can choose a sequence (f_n) , with $f_n \in A(D_0), \lim f_n = f$. Then

$$\|r_h E f - E_h s_h f\| \leq \|r_h E f - r_h E f_n\| + \|r_h E f_n - E_h s_h f_n\| + \|E_h s_h f_n - E_h s_h f\|.$$

Since E, E_h, r_h, s_h can be bounded independently of h , the first and third terms of the right-hand side can be made arbitrarily small, uniformly in h , by taking n large, while the second term tends to zero with h .

(ii) Let $f \in Y$. The norms $\|r_h E f\|$ are bounded as $h \rightarrow 0$, because (2.3) holds. From (2.4) we conclude that the norms $\|E_h s_h f\|$ are also bounded, since H has no limit points other than 0. The generalized Banach-Steinhaus lemma of [9] then shows that there exists a constant K_1 such that $\|E_h s_h\| \leq K_1$. If $g \in Y_h$, with $\|g\| \leq 1$, we can write (cf. condition (P))

$$\|E_h g\| = \|E_h s_h f\| \leq K_1 L,$$

whence $\|E_h\| \leq K_1 L$.

Remark 2.1. It has been shown in [9] that condition (P) holds in most practical applications.

Remark 2.2. We emphasize that while implication (i) has been proved by elementary means, implication (ii) requires the use of a deep result from functional analysis. In this regard we note that while the convergence or otherwise of a method depends on the norms in X_h , but *not* on the norms in Y_h , the concept of stability depends on the norms of X_h and Y_h . Therefore, one may argue that by changing the norms in Y_h one could turn a stable method into an unstable one without altering the convergence. From this line of thought one may be led to believe that implication (ii) cannot hold in general (cf. [13, p. 14]). This paradox is explained as follows. The equiboundedness of s_h together with property (P) establish a link between the norms in Y_h and the norm in Y . (If a *finer* norm were introduced in Y_h , the equiboundedness would be likely to disappear. The introduction of a *coarser* norm in Y_h would threaten the validity of (P).) Thus, in our framework the norms in Y_h cannot be significantly altered without altering the norm in Y . But, as a consequence of the closed graph theorem, the norm of the Banach space Y cannot be weakened or strengthened.

Remark 2.3. The considerations above suggest that the completeness of Y and condition (P) are essential if (ii) is to hold. The necessity of these conditions is shown in Section 4 by means of counterexamples.

Remark 2.4. It is obvious from the above proof that (P) can be relaxed to read (P') There exists a constant L and subspaces $S_h \subset Y_h$ such that, for $h \in H$,

$$\sup\{\|E_h g\| : g \in S_h, \|g\| \leq 1\} = \sup\{\|E_h g\| : g \in Y_h, \|g\| \leq 1\},$$

and to each $g \in S_h$, with $\|g\| \leq 1$, there corresponds an element $f \in Y$, with $\|f\| \leq L, s_h f = g$.

In other words, it suffices to check (P) for g ranging in a subspace S_h such that E_h "attains its norm" in S_h .

Remark 2.5. The hypotheses that H has no limit point other than 0 is not essential. The theorem holds for general H such that $\inf H = 0$, provided that $\|E_h\|$ is bounded for h bounded away from zero. This supplementary condition is invariably verified in the applications.

3. Examples. In this section we present four examples of applications of the previous theory. These examples show the way to further generality.

3.1. *The Classical Lax-Richtmyer Theory* [6]. We considered a well-posed Cauchy problem [10, p. 39]

$$(3.1) \quad du/dt = \mathcal{A}u, \quad 0 \leq t \leq T, \quad u(0) = u_0,$$

where \mathcal{A} is the generator of a strongly continuous semigroup in a Banach space \mathcal{B} . This problem is cast in the form of Section 2.1 by choosing X to be the space of continuous mappings from $[0, T]$ into \mathcal{B} with the supremum norm, $Y = \mathcal{B}$, and A the operator

$$u(\cdot) \rightarrow Au(\cdot) = u(0),$$

with domain

$$D(A) = \{u(\cdot) \in X \mid du/dt \text{ exists, } du/dt = \mathcal{A}u, 0 < t \leq T\}.$$

A difference scheme is a recursion [6]

$$(3.2) \quad u_{n+1} = C(h)u_n, \quad n = 0, 1, 2, \dots, [T/h] - 1,$$

where h ranges in a set H as in Section 2.2, $C(h)$ is a bounded linear operator in \mathcal{B} , and u_n is meant to approximate $u(nh)$. This is accommodated in the present formalism as follows. We define X_h to be the product of $N + 1 = [T/h] + 1$ copies of the space \mathcal{B} endowed with the supremum norm. The restriction r_h is the natural point restriction

$$r_h u(\cdot) = [u(0), u(h), \dots, u(Nh)].$$

The space Y_h is taken as the product of $N + 1$ copies of \mathcal{B} with norm

$$\|[f_0, f_1, \dots, f_N]\| = \sum_{i=0}^N \|f_i\|.$$

The restriction s_h is defined as

$$s_h u = [u, 0, 0, \dots, 0], \quad u \in \mathcal{B}.$$

It is clear that (2.3) holds with $C_1 = C_2 = 1$. Now the recursion (3.2) can be written

$$(3.3) \quad \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} I & & & & & \\ -C & I & & & & 0 \\ & -C & I & & & \\ & & \cdot & \cdot & & \\ & 0 & & \cdot & \cdot & \\ & & & & \cdot & \\ & & & & -C & I \end{bmatrix}^{-1} \quad s_h u_0 = A_h^{-1} s_h u_0,$$

which is of the required form. (We have omitted the dependence of C on h .)

We shall show that Theorem 1, as applied to this choice of $X, Y, A, X_h, Y_h, A_h, r_h, s_h$, is precisely the Lax-Richtmyer equivalence theorem. In fact, the matrix operator A_h has an inverse

$$E_h = \begin{bmatrix} I & & & & \\ C & I & & & 0 \\ C^2 & C & I & & \\ \cdot & \cdot & \cdot & \cdot & \\ C^N & C^{N-1} & C^{N-2} & \cdot & I \end{bmatrix}$$

with norm $\sup\{\|C^n\|, 0 \leq n \leq N\}$. (We recall that the norm of a matrix operator from L^1 into L^∞ is given by the supremum of the norms of its entries.) Therefore our definition of stability reduces in this case to

$$(3.4) \quad \sup\{\|C(h)^n\|: 0 \leq nh \leq T, h \in H\} < \infty,$$

i.e., the usual definition of Lax stability. Furthermore, E_h attains its norm on $s_h \mathcal{B}$, so that condition (P') holds with $S_h = s_h \mathcal{B}$. Finally, it is obvious that our definitions of convergence and consistency are essentially those of the Lax-Richtmyer theory [10]. (Actually, the present requirement of consistency is less demanding than that of [10].)

Remark. In order to follow more closely the conventions of [13, p. 6], we may divide by h each row of A_h except the first, in order to write the difference equations as approximations to the differential problem (i.e., Euler's method may be written as $(u_{n+1} - u_n)/h = Au_n$, rather than $u_{n+1} = u_n + hAu_n$). Simultaneously one must change the norm in Y_h into the normalized form $\|f_0\| + h\sum_{i=1}^N \|f_i\|$. We emphasize that these changes are merely a matter of notation. See [13, p. 75] on the practical advisability of choosing the norm in X_h to be of supremum type and that of Y_h to be of L^1 type.

3.2. *The L^2 -Inhomogeneous Case* [7]. We now consider the inhomogeneous problem

$$du/dt = \mathcal{A}u + f(t), \quad 0 < t \leq T, \quad u(0) = u_0,$$

where \mathcal{A} is as in the previous paragraph and $f \in L^2([0, T], \mathcal{B})$. We define $Y = \mathcal{B} \times L^2([0, T], \mathcal{B})$, $X = \mathcal{C}([0, T], \mathcal{B})$, and A the operator

$$u(\cdot) \rightarrow Au(\cdot) = (u(0), du/dt - \mathcal{A}u)$$

with domain

$$D(A) = \{u \in X \mid du/dt \text{ exists, } du/dt - \mathcal{A}u \in L^2([0, T], \mathcal{B})\}.$$

The problem is well-posed with solution operator

$$E(u_0, f)(t) = S(t)u_0 + \int_0^t S(t-s)f(s) ds,$$

where S is the semigroup generated by \mathcal{A} . Following Mountain [7], we consider the method

$$u_{n+1} = C(h)u_n + hf_n, \quad f_n = h^{-1} \int_{nh}^{(n+1)h} f(t) dt$$

where $C(h)$ is as before. Note that averages must be used for f_n since point values are meaningless for L^2 -functions. This is cast in the form (2.2) by choosing X_h, Y_h, A_h, r_h as before but changing s_h into

$$s_h(v, f) = [v; hf_0, hf_1, \dots, hf_{N-1}].$$

It is readily shown that the s_n are uniformly bounded. The condition of stability is still given by (3.4), since X_h, Y_h, A_h have not been altered. Thus, the convergence of those methods which are stable and consistent follows from Theorem 1 (i). Conversely, a method which converges for all data in Y is stable, since it was shown before that stability follows from the weaker requirement of convergence for problems having $f \equiv 0$. Also note that L^2 may be replaced by any $L^p, 1 \leq p \leq \infty$, or by the space of continuous functions, thus generalizing the results of [7]. When $Y = \mathcal{B} \times \mathcal{C}([0, T], \mathcal{B})$, one can use point values $f_n = f(nh)$ rather than averages.

3.3. *A Generalized Lax-Richtmyer Theory* [9]. In the Lax-Richtmyer theory the true solution $u(nh)$ and its approximation u_n are supposed to lie in the same space \mathcal{B} , while in practical applications the former is a function of the space variables and u_n is only a grid function. In [9] a simple proof was given of the validity of the Lax equivalence theorem even if $u(nh), u_n$ are allowed to lie in different spaces. It is easy to show that this generalization of the Lax theorem is also a particular case of our Theorem 1.

3.4. *Elliptic Boundary Value Problems*. We consider a homogeneous Dirichlet problem [3] in a bounded domain with smooth boundary

$$\sum_{i,j} \frac{\partial}{\partial x_i} \left(a_{ij}(x) \frac{\partial}{\partial x_j} u(x) \right) = f(x), \quad x \in \Omega \subset R^d. \quad u(x) = 0, \quad x \in \partial\Omega,$$

for a strongly elliptic operator, with smooth coefficients $a_{ij} = a_{ji}$.

We take the domain of the operator to be $H_0^1(\Omega) \cap H^2(\Omega)$ with the energy norm and assume that f ranges in $L^2(\Omega)$. Thus there is a continuous solution operator $E: L^2 \rightarrow H_0^1$, which is characterized by the variational formulation

$$a(Ef, \psi) = (f, \psi), \quad \forall \psi \in H_0^1.$$

Here $a(\cdot, \cdot)$ is the bilinear form associated with the differential operator (i.e., the energy inner product), and (\cdot, \cdot) is the usual L^2 inner product. If Z_h is a sequence of finite-dimensional subspaces of H_0^1 , we consider the Galerkin solutions $u_h \in Z_h$:

$$a(u_h, \psi) = (f, \psi), \quad \forall \psi \in Z_h.$$

We choose X_h to be Z_h with the energy norm, and Y_h to be Z_h with the L^2 -norm. If we take for the roles of r_h, s_h the $a(\cdot, \cdot)$ - and (\cdot, \cdot) -orthogonal projections of H_0^1 and L^2 onto Z_h , respectively, then the conditions (P) and (2.3) are trivially satisfied.

provided that the perturbations g_h satisfy $\lim \|g_h\|_{Y_h} = 0$. The method is L -convergent if it is L -convergent for every f in Y .

The name L -convergent originates from the theory of initial value problems in PDEs (Ansoerge [1], [2]). Similar concepts have often been used in the literature: cf. stable convergence (Dahlquist [4]) and convergence under perturbations (Spijker [12]). The notion of convergence of linear multistep methods in Henrici [5] is in fact an L -convergence concept, since it is required there that convergence take place under arbitrary consistent starting procedures (in particular, method (4.2) is not convergent in the sense of Henrici).

Note that both the norms of X_h and Y_h enter in the definition of L -convergence and that the idea of stability is implied in the idea of L -convergence. Therefore, it is not surprising that the following theorem can be easily proved by elementary means (cf. Remark 2.2).

THEOREM 2. *Let $(X_h, Y_h, A_h, r_h, s_h)$ be a method for the solution of (1.1). Then*

(i) *If the method is stable and convergent for f in a dense subset of Y , then it is L -convergent.*

(ii) *If the method is L -convergent for $f \equiv 0$, then it is stable.*

A proof can be found in Stummel [14, p. 53]. (Stummel uses “consistency” where we use “convergence”, and “convergence” where we use “ L -convergence”.) From Theorems 1 and 2 we conclude

COROLLARY 1. *A method is L -convergent if and only if it is stable and convergent.*

COROLLARY 2. *A consistent method is L -convergent if and only if it is stable.*

Note that the equivalence result in Corollary 2 requires neither the completeness of Y nor the condition (P). Corollary 1 holds—when L -convergence, stability, and convergence are generalized in the obvious way—even in nonlinear situations [15, Theorem 1]. For initial value problems the equivalence between L -convergence and stability was first noted by Spijker [11]. Corollary 2 can also be extended to nonlinear situations [12]. An appraisal of some recent work on equivalence theorems may be found in [8].

Acknowledgment. The authors are thankful to J. C. López-Marcos for the discussion which started this research. They have benefited from financial support from “Comisión Asesora”.

Departamento Ecuaciones Funcionales
Universidad de Valladolid
Valladolid, Spain

1. R. ANSORGE, “Survey of equivalence theorems in the theory of difference approximations for partial differential equations”, *Topics in Numerical Analysis III* (J. J. M. Miller, ed.), Academic Press, London, 1977, pp. 1–16.

2. R. ANSORGE, *Differenzenapproximationen partieller Anfangswertaufgaben*, Teubner, Stuttgart, 1978.

3. P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

4. G. DAHLQUIST, “Convergence and stability in the numerical integration of ordinary differential equations,” *Math. Scand.*, v. 4, 1956, pp. 33–53.

5. P. HENRICI, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York, 1962.

6. P. D. LAX & R. D. RICHTMYER, "Survey of stability of linear finite difference equations," *Comm. Pure Appl. Math.*, v. 9, 1956, pp. 267–293.
7. J. K. MOUNTAIN, "The Lax equivalence theorem for linear inhomogeneous equations in L^2 spaces," *J. Approx. Theory*, v. 33, 1981, pp. 126–130.
8. C. PALENCIA & J. M. SANZ-SERNA, "Equivalence theorems for incomplete spaces: an appraisal," *IMA J. Numer. Anal.*, v. 4, 1984, pp. 109–115.
9. C. PALENCIA & J. M. SANZ-SERNA, "An extension of the Lax-Richtmyer theory," *Numer. Math.*, v. 44, 1984, pp. 279–283.
10. R. D. RICHTMYER & K. W. MORTON, *Difference Methods for Initial Value Problems*, Interscience, New York, 1967.
11. M. N. SPIJKER, *Stability and Convergence of Finite-Difference Methods*, Thesis, Leiden, Rijksuniversiteit, 1968.
12. M. N. SPIJKER, "Equivalence theorems for nonlinear finite difference methods," *Numerische Behandlung Nichtlinearer Integrodifferential und Differentialgleichungen* (R. Ansorge and W. Tornig, eds.), Lecture Notes in Math., Vol. 395, Springer-Verlag, Berlin, 1974, pp. 109–122.
13. H. J. STETTER, *Analysis of Discretization Methods for Ordinary Differential Equations*, Springer, Berlin, 1973.
14. F. STUMMEL, "Diskrete Konvergenz linearer Operatoren. I," *Math. Ann.*, v. 190, 1970, pp. 45–92.
15. F. STUMMEL, "Weak stability and weak discrete convergence of continuous mappings," *Numer. Math.*, v. 26, 1976, pp. 301–315.
16. R. TEMAM, *Navier-Stokes Equations, Theory and Numerical Analysis*, North-Holland, Amsterdam, 1977.