Electronic Journal of Qualitative Theory of Differential Equations
Proc. 9th Coll. QTDE, 2012, No. **6** 1-22;
http://www.math.u-szeged.hu/ejqtde/

Notes on the basic notions in nonlinear numerical analysis*

I. Faragó[†], M. E. Mincsovics[‡], I. Fekete[§]

Abstract

In this paper we investigate the numerical solution of non-linear equations in an abstract (Banach space) setting. The main result is that the convergence can be guaranteed by two, directly checkable conditions (namely, by the consistency and the stability). We show that these conditions together are a sufficient, but not necessary condition for the convergence. Our theoretical results are demonstrated on the numerical solution of a Cauchy problem for ordinary differential equations by means of the explicit Euler method.

Keywords: numerical method, non-linear problems, Lax theory, convergence, stability.
Mathematical Subject Classification: 65J05, 65J15, 34G20.

1 Introduction

Many phenomena in nature can be described by mathematical models which consist of functions of a certain number of independent variables and parameters. In particular, these models often consist of equations, usually containing a large variety of derivatives with respect to the variables. Typically, we are not able to give the solution of the mathematical model in a closed (analytical) form, we construct some numerical and computer models that are useful for practical purposes. The

*This paper is in final form and no version of it is submitted for publication elsewhere.
[†]E-mail: faragois@cs.elte.hu, Institute of Mathematics, Eötvös Loránd University, Budapest, Hungary
[‡]E-mail: mincso@cs.elte.hu, Institute of Mathematics, Eötvös Loránd University, Budapest, Hungary
[§]E-mail:imrefekete1989@gmail.com, Institute of Mathematics, Eötvös Loránd University, Budapest, Hungary

ever-increasing advances in computer technology has enabled us to apply numerical methods to simulate plenty of physical and mechanical phenomena in science and engineering. As a result, numerical methods do not usually give the exact solution to the given problem, they can only provide approximations, getting closer and closer to the solution with each computational step. Numerical methods are generally useful only when they are implemented on computer using a computer programming language. Using a computer, it is possible to gain quantitative (and also qualitative) information with detailed and realistic mathematical models and numerical methods for a multitude of phenomena and processes in physics and technology. The application of computers and numerical methods has become ubiquitous. Computations are often cheaper than experiments; experiments can be expensive, dangerous or downright impossible. Real-life experiments can often be performed on a small scale only, and that makes their results less reliable.

The above modelling process of real-life phenomena can be illustrated as follows:

$$\boxed{\begin{array}{l}\text{real-life problem}\\ \text{+ physical model}\end{array}} \Rightarrow \boxed{\begin{array}{l}\text{mathematical}\\ \text{model}\end{array}} \Rightarrow \boxed{\text{numerical model}}$$

This means that the complete modelling process consists of three steps. In this paper we will analyze the step when we transform the mathematical (usually continuous) model into numerical (usually discrete) models. Our aim is to guarantee that this step does not cause any significant loss of the information.

The discrete model usually yields a sequence of (discrete) tasks. During the construction of the numerical models the basic requirements are the following.

- Each discrete problem in the numerical model is a well-posed problem.
- In the numerical model we can efficiently compute the numerical solution.
- The sequence of the numerical solutions is convergent.
- The limit of this sequence is the solution of the original problem.

The theory is more developed for linear problems, see [LR56, PS84a, PS84b, PS85], while the nonlinear theory is less elaborated. The main purpose of this work is to investigate the nonlinear theory.

The paper is organized as follows.

In Section 2 we give the mathematical formulation of the above formulated general description of the mathematical and numerical models. In Section 3 we define the basic numerical notions (convergence, consistency, stability), and we formulate the relation between them. In Section 4 we generalize these notions from a practical point of view. In Section 5, via some concrete examples, we examine the relation between consistency, stability and convergence. Finally, we give some remarks and conclusions.

2 Mathematical background

When we model some real-life phenomenon with a mathematical model, it results in a – usually nonlinear – problem of the form

$$F(u) = 0\,, \tag{1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are normed spaces, $\mathcal{D} \subset \mathcal{X}$ and $F : \mathcal{D} \to \mathcal{Y}$ is a (nonlinear) operator. In the theory of numerical analysis it is usually *assumed* that there exists a unique solution, which will be denoted by $\bar{u}$.

On the other side, for any concrete applied problems *we must prove* the existence of $\bar{u} \in \mathcal{D}$. In most cases the proof is not constructive, cf. [K75]. Even if it is possible to solve directly, the realization of the solving process is very difficult or even impossible. However, we need only a good approximation for the solution of problem (1), since our model is already a simplification of the real-life phenomenon. Therefore we construct numerical models by use of some discretization, which results in a sequence of simpler problems, i.e., a numerical method. The requirements from these simpler problems were formulated in the Introduction. With this approach we need to face the following difficulties:

- we need to compare the solution of the simpler problems with the solution of the original problem (1), which might be found in different spaces;
- this comparison seems to be impossible, since the solution of the original problem (1) is not known.

To get rid of the latter difficulty, the usual trick is to introduce the notions of consistency and stability, which are independent of the solution of the original problem (1) and are controllable. The convergence can be replaced with these two notions. Sometimes this popular "recipe" is summarized in the formula

$$\mathrm{Consistency} + \mathrm{Stability} = \mathrm{Convergence}\,. \tag{2}$$

In the following we introduce and investigate these notions in an abstract framework, and we try to shed some light on the formula (2). Namely:

- how to define consistency and stability to ensure the formula (2);

- is it consistency or/and stability that is necessary for the convergence (in the linear case the Lax-equivalence theorem deals with this question, too, see e.g. [LR56, PS84a]);

This paper is mainly devoted to these questions.

First, we start with some definitions and notations, by giving an example.

Definition 1. *Problem (1) can be given as a triplet* $\mathscr{P} = (\mathcal{X}, \mathcal{Y}, F)$. *We will refer to it as* problem $\mathscr{P}$.

Example 2. *Consider the following initial value problem:*

$$u'(t) = f(u(t)) \tag{3}$$

$$u(0) = u_0\,, \tag{4}$$

where $t \in [0,1],\ u_0 \in \mathbb{R}$ *and* $f \in C(\mathbb{R}, \mathbb{R})$ *is a Lipschitz continuous function.*

Then the operator F *and the spaces* $\mathcal{X}, \mathcal{Y}$ *are defined as follows.*

- $\mathcal{X} = C^1[0,1],\ \ \|u\|_{\mathcal{X}} = \max\limits_{t\in[0,1]} |u(t)|$
- $\mathcal{Y} = C[0,1] \times \mathbb{R},\ \ \left\| \begin{pmatrix} u \\ u_0 \end{pmatrix} \right\|_{\mathcal{Y}} = \max\limits_{t\in[0,1]} (|u(t)|) + |u_0|$
- $F(u) = \begin{pmatrix} u'(t) - f(u(t)) \\ u(0) - u_0 \end{pmatrix}.$

Definition 3. *We say that the sequence* $\mathscr{N} = (\mathcal{X}_n, \mathcal{Y}_n, F_n)_{n\in\mathbb{N}}$ *is a numerical method if it generates a sequence of problems*

$$F_n(u_n) = 0\,, \quad n = 1, 2, \dots\,, \tag{5}$$

where

- $\mathcal{X}_n, \mathcal{Y}_n$ *are normed spaces;*
- $\mathcal{D}_n \subset \mathcal{X}_n$ *and* $F_n : \mathcal{D}_n \to \mathcal{Y}_n\,.$

If there exists a unique solution of the (approximating) problems (5), it will be denoted by $\bar{u}_n\,.$

Example 4. *For* $n \in \mathbb{N}$ *we define the following sequence of triplets:*

- $\mathcal{X}_n = \mathbb{R}^{n+1},\ \ \mathbf{v}_n = (v_0, v_1, \dots, v_n) \in \mathcal{X}_n : \|\mathbf{v}_n\|_{\mathcal{X}_n} = \max\limits_{i=0,\dots,n} |v_i|$

- $\mathcal{Y}_n = \mathbb{R}^{n+1}$, $\mathbf{y}_n = (y_0, y_1, \ldots, y_n) \in \mathcal{Y}_n : \|\mathbf{y}_n\|_{\mathcal{Y}_n} = |y_0| + \max_{i=1,\ldots,n} |y_i|$.
- $F_n : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$*, and for any* $\mathbf{v}_n = (v_0, v_1, \ldots, v_n) \in \mathbb{R}^{n+1}$ *it acts as*

$$(F_n(\mathbf{v}_n))_i = \begin{cases} n\left(v_i - v_{i-1}\right) - g\left(v_{i-1}\right), & i = 1, \ldots, n, \\ \\ v_0 - c, & i = 0. \end{cases} \tag{6}$$

(Here $g : \mathbb{R} \to \mathbb{R}$ *and* $c \in \mathbb{R}$ *are arbitrary given data which define the numerical process.)*

Definition 5. *We say that the sequence* $\mathscr{D} = (\varphi_n, \psi_n, \Phi_n)_{n\in\mathbb{N}}$ *is a* discretization *if*

- *the* φ_n*-s (respectively* ψ_n*-s) are restriction operators from* $\mathcal{X}$ *into* $\mathcal{X}_n$ *(respectively from* $\mathcal{Y}$ *into* $\mathcal{Y}_n$*), where* $\mathcal{X}, \mathcal{X}_n, \mathcal{Y}, \mathcal{Y}_n$ *are normed spaces;*
- $\Phi_n : \{F : \mathcal{D} \to \mathcal{Y} \,|\, \mathcal{D} \subset \mathcal{X}\} \to \{F_n : \mathcal{D}_n \to \mathcal{Y}_n \,|\, \mathcal{D}_n \subset \mathcal{X}_n\}$.

Example 6. *Based on Examples 2 and 4, in Definition 5 we define* $\mathcal{X} = C^1[0,1]$, $\mathcal{Y} = C[0,1] \times \mathbb{R}$*, and* $\mathcal{X}_n = \mathcal{Y}_n = \mathbb{R}^{n+1}$. $\mathbb{G}_n := \{t_i = \frac{i}{n},\ i = 0, \ldots, n\}$*. Then, we define the triplet of the operators as follows.*

- *For any* $u \in \mathcal{X}$ *we put* $(\varphi_n u)_i = u(t_i)$, $i = 0, 1, \ldots, n$*,*
- *For any* $y \in \mathcal{Y}$ *we put*

$$(\psi_n y)_i = \begin{cases} y\left(t_{i-1}\right), & 1, \ldots, n, \\ \\ y(t_0), & i = 0. \end{cases}$$

- *In order to give* Φ_n*, we define the mapping* $\Phi_n : C^1[0,1] \to \mathbb{R}^{n+1}$ *in the following way:*

$$\left[\left(\Phi_n(F)\right) u\right]_i = \begin{cases} n\left(u(t_i) - u(t_{i-1})\right) - g(u(t_{i-1})), & i = 1, \ldots, n, \\ \\ u(t_0) - c, & i = 0. \end{cases} \tag{7}$$

We note that the introduced notions of problem and numerical methods are independent of each other. However, for our purposes only those numerical methods $\mathscr{N}$ are interesting which are obtained when some discretization method $\mathscr{D}$ is applied to some certain problem $\mathscr{P}$.

Remark 7. *Theoretically, the normed spaces* $\mathcal{X}$ *and* $\mathcal{Y}$ *in the definitions of the problem and of the discretization might be different. However the application of the discretization to the problem is possible only when these normed spaces are the same. In the sequel this will be always assumed.*

Example 8. *Let us define the numerical method $\mathscr{N}$ for the problem $\mathscr{P}$ from Example 2, and for the discretization $\mathscr{D}$ from Example 6. Then we solve the sequence of problems in the form* (5)*, where in the discretization for g and c we put f and u_0 from problem* (3)-(4)*, respectively. This yields that the mapping $F_n : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is defined as follows: for the vector $\mathbf{v}_n = (v_0, v_1, \ldots, v_n) \in \mathbb{R}^{n+1}$ we have*

$$(F_n(\mathbf{v}))_i = \begin{cases} n\,(v_i - v_{i-1}) - f(v_{i-1}), & i = 1, \ldots, n, \\ \\ v_0 - u_0, & i = 0. \end{cases} \tag{8}$$

Hence, using the notation $h = 1/n$, the equation (5) *for* (8) *results in the task: we seek the vector $\mathbf{v} = (v_0, v_1, \ldots, v_n) \in \mathbb{R}^{n+1}$ such that*

$$\begin{cases} \dfrac{v_i - v_{i-1}}{h} = f(v_{i-1}), & i = 1, \ldots, n, \\ \\ v_0 = u_0, & i = 0. \end{cases} \tag{9}$$

Hence, the obtained numerical method is the well-known explicit Euler method on the mesh $\mathbb{G}_n$ with step-size h.

In sequel for the discretization $\mathscr{D} = (\varphi_n, \psi_n, \Phi_n)_{n\in\mathbb{N}}$ we assume the validity of the following assumption.

Assumption 9. *The discretization $\mathscr{D}$ possesses the property $\psi_n(0) = 0$.*

Obviously, when ψ_n are linear operators, then this condition is automatically satisfied. We also list two further natural assumptions about the discretization, which will be used later.

Assumption 10. *The discretization $\mathscr{D}$ generates a numerical method $\mathscr{N}$ which possesses the property $\dim \mathcal{X}_n = \dim \mathcal{Y}_n < \infty$.*

Assumption 11. *Let us apply the discretization $\mathscr{D}$ to the problem $\mathscr{P}$. We assume that F_n is continuous on the ball $B_R(\varphi_n(\bar{u}))$.*

The general scheme of the above approach is illustrated in Figure 1.

3 Basic Theoretical Results

In this part we analyze the general framework of a numerical method (according to Figure 1). We apply a discretization $\mathscr{D}$ for some problem $\mathscr{P}$, then it results in a numerical method $\mathscr{N}$, which generates the sequence of problems (5). Our

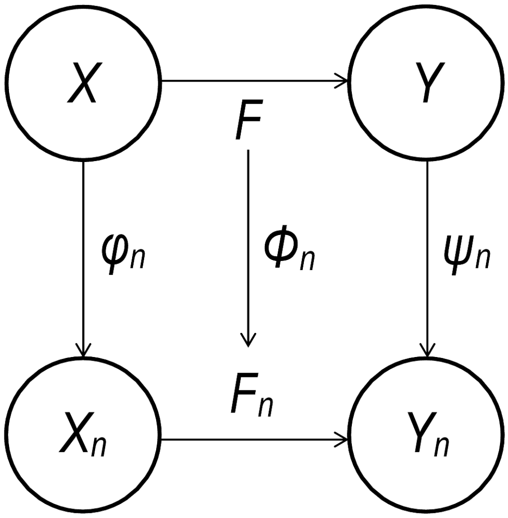

Figure 1: The general scheme of numerical methods.

aim is to guarantee the existence of the solutions $\bar{u}_n$ and the closeness of these to $\bar{u}$. To this aim we define the distance between these elements, which will be called global discretization error. (Since these elements belong to different spaces, this is not straightforward.) Independently of the form of the definition of the global error, it is hardly applicable in practice, because the knowledge of the exact solution $\bar{u}$ is assumed. Therefore, we introduce some further notions (consistency, stability), which help us in getting information about the behavior of the global discretization error.

3.1 Convergence

The usual approach for the characterization of the distance of the elements $\bar{u}$ and $\bar{u}_n$ is their comparison in $\mathcal{X}_n$ in the following way.

Definition 12. *The element* $e_n = \varphi_n(\bar{u}) - \bar{u}_n \in \mathcal{X}_n$ *is called global discretization error.*

Clearly, our aim is to guarantee that the global discretization error is arbitrary small, by increasing n. That is, we require the following property.

Definition 13. *The discretization* $\mathscr{D}$ *applied to the problem* $\mathscr{P}$ *is called* convergent *if*

$$\lim \|e_n\|_{\mathcal{X}_n} = 0 \tag{10}$$

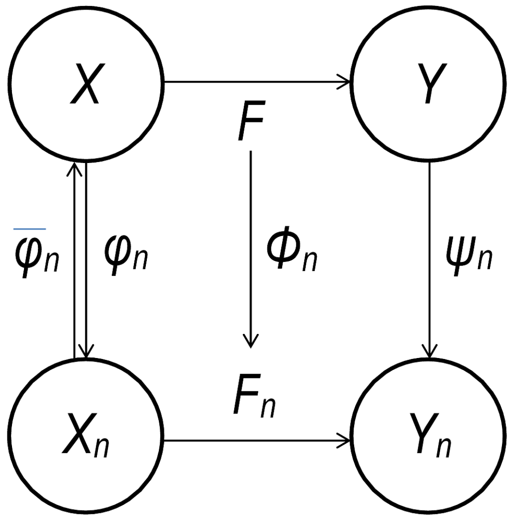

Figure 2: The general scheme of numerical methods with interpolation operator.

holds. When

$$\|e_n\|_{\mathcal{X}_n} = \mathcal{O}(n^{-p})$$

we say that the order of the convergence *is* p*.*

Remark 14. *It is possible to define the distance between the elements* $\bar{u}$ *and* $\bar{u}_n$ *in the space* $\mathcal{X}$*, with the help of an operator* $\bar{\varphi}_n : \mathcal{X}_n \to \mathcal{X}$*, by the quantity* $\|\bar{u} - \bar{\varphi}_n \bar{u}_n\|_{\mathcal{X}}$*. For such an approach see Figure 2.*

Here we assume that $\lim(\varphi_n \circ \bar{\varphi}_n)v = v$ *for any* $v \in \mathcal{X}$*. We note that this relation does not mean that* $\bar{\varphi}_n$ *is the inverse of* φ_n*, because* φ_n *is not invertible, typically it represents some interpolation. In this approach the convergence means that the numerical sequence* $\|\bar{u} - \bar{\varphi}_n \bar{u}_n\|_{\mathcal{X}}$ *tends to zero. Because this approach requires an additional interpolation, and the choice of the interpolation may influence the rate of the convergence, therefore this kind of convergence is less common.*

3.2 Consistency

Consistency is the notion which makes some connection between the problem $\mathscr{P}$ and the numerical method $\mathscr{N}$.

Definition 15. *The discretization* $\mathscr{D}$ *applied to problem* $\mathscr{P}$ *is called consistent at the element* $v \in D$ *if*

- $\varphi_n(v) \in \mathcal{D}_n$ *holds from some index,*

- *the relation*

$$\lim \|F_n(\varphi_n(v)) - \psi_n(F(v))\|_{\mathcal{Y}_n} = 0 \tag{11}$$

  holds.

The element $l_n(v) = F_n(\varphi_n(v)) - \psi_n(F(v)) \in \mathcal{Y}_n$ in (11) plays an important role in the numerical analysis. When we fix some element $v \in \mathcal{D}$, we can transform it into the space in two different ways: $\mathcal{X} \to \mathcal{Y} \to \mathcal{Y}_n$ and $\mathcal{X} \to \mathcal{X}_n \to \mathcal{Y}_n$ (c.f. Figure 1). The magnitude $l_n(v)$ characterizes the difference of this two directions for the element v. Hence, the consistency at the element v yields that in limit the diagram of Figure 1 is commutative. A special role is played by the behavior of $l_n(v)$ on the solution of the problem (1), that is the elements $l_n(\bar{u})$. Later on we will use the following notions.

Definition 16. *The element $l_n(v) = F_n(\varphi_n(v)) - \psi_n(F(v)) \in \mathcal{Y}_n$ is called local discretization error at the element v. The element $l_n(\bar{u}) = F_n(\varphi_n(\bar{u})) - \psi_n(F(\bar{u})) = F_n(\varphi_n(\bar{u}))$ is called* local discretization error. *When*

$$\|l_n(v)\|_{\mathcal{X}_n} = \mathcal{O}(n^{-p}),$$

we say that the order of the consistency at v is p.

Remark 17. *For simplicity, we will use the notation l_n for $l_n(\bar{u})$. In the sequel, the consistency on $\bar{u}$ and its order will be called consistency and order of consistency.*

One might ask whether consistency implies convergence. The following simple example shows that this is not true in general.

Example 18. *Let us consider the case $\mathcal{X} = \mathcal{X}_n = \mathcal{Y} = \mathcal{Y}_n = \mathbb{R}$, $\varphi_n = \psi_n = identity$. Our aim is to solve the scalar equation $F(x) = 0$, where we assume that it has a unique solution $\bar{x} = 0$. We define the numerical method $\mathcal{N}$ as $F_n(x) = (1 - x)/n$. Clearly, due to the linearity of φ_n and ψ_n, we have $l_n = F_n(0) - 0 = F_n(0)$. Since $F_n(0) \to 0$, therefore this discretization is consistent. However, it is not convergent, since the solution of each problem $F_n(x) = 0$ is $\bar{x}_n = 1$.*

Thus, convergence cannot be replaced by consistency in general.

3.3 Stability

As we have already seen, consistency in itself is not enough for convergence. Assuming the existence of the inverse operator F_n^{-1}, we can easily get to the relation

$$e_n = \varphi_n(\bar{u}) - \bar{u}_n = F_n^{-1}(F_n(\varphi_n(\bar{u}))) - F_n^{-1}(0) = F_n^{-1}(l_n) - F_n^{-1}(0),$$

which shows the connection between the global and local discretization errors. This relation suggests that the consistency (i.e., the convergence to of the local discretization error l_n to zero) can provide the convergence (i.e., the approach of e_n to zero) when $(F_n^{-1})_{n\in\mathbb{N}}$ has good behavior. Such a property is the Lipschitz continuity: it would be useful to assume that the functions F_n^{-1} uniformly satisfy the Lipschitz condition at the point $0 \in \mathcal{Y}_n$. However, generally at this point we have no guarantee even to the existence of F_n^{-1}, thus we provide this with some property of the functions F_n, without assuming their invertibility. The first step in this direction is done by introducing a simplified form of the notion of semistability in [LS88].

Definition 19. *The discretization $\mathscr{D}$ is called* semistable *on the problem $\mathscr{P}$ if there exist $S \in \mathbb{R}$, $R \in (0,\infty]$ such that*

- *$B_R(\varphi_n(\bar{u})) \subset \mathcal{D}_n$ holds from some index;*
- *$\forall (v_n)_{n\in\mathbb{N}}$ which satisfy $v_n \in B_R(\varphi_n(\bar{u}))$ from that index, the relation*

$$\|\varphi_n(\bar{u}) - v_n\|_{\mathcal{X}_n} \le S \, \|F_n(\varphi_n(\bar{u})) - F_n(v_n)\|_{\mathcal{Y}_n} \tag{12}$$

holds.

Semistability is a purely theoretical notion, which, similarly as the consistency, cannot be checked directly, due to the fact, that $\bar{u}$ is unknown. However, the following statement clearly shows the relation of the three important notions.

Lemma 20. *We assume that the discretization $\mathscr{D}$*

- *is consistent at $\bar{u}$ and semistable with stability threshold R on the problem $\mathscr{P}$;*
- *generates a numerical method $\mathscr{N}$ that Equation (5) has a solution in $B_R(\varphi_n(\bar{u}))$ from some index.*

Then the sequence of these solutions of Equation (5) converges to the solution of problem $\mathscr{P}$, and the order of convergence is not less than the order of consistency.

Proof. Having the relation $F_n(\bar{u}_n) = \psi_n(F(\bar{u})) = 0$, we get

$$\|\varphi_n(\bar{u}) - \bar{u}_n\|_{\mathcal{X}_n} \le S \, \|F_n(\varphi_n(\bar{u})) - F_n(\bar{u}_n)\|_{\mathcal{Y}_n} = S \, \|F_n(\varphi_n(\bar{u})) - \psi_n(F(\bar{u}))\|_{\mathcal{Y}_n} .$$

This yields that $\|e_n\|_{\mathcal{X}_n} \le S\|l_n\|_{\mathcal{Y}_n}$, which proves the statement. □

This lemma has some drawbacks. First, we cannot verify its conditions because this requires the knowledge of the solution. Secondly, we have no guarantee that equation (5) has a (possibly unique) solution in $B_R(\varphi_n(\bar{u}))$ from some index. The following modified stability notion, see [K75], gets rid of the second problem.

Definition 21. *The discretization* $\mathscr{D}$ *is called* stable *on problem* $\mathscr{P}$ *at the element* $v \in \mathcal{X}$ *if there exist* $S \in \mathbb{R}$, $R \in (0, \infty]$ *such that*

- $B_R(\varphi_n(v)) \subset \mathcal{D}_n$ *holds from some index;*
- $\forall (v_n^1)_{n\in\mathbb{N}}, (v_n^2)_{n\in\mathbb{N}}$ *which satisfy* $v_n^i \in B_R(\varphi_n(v))$, *the estimate*

$$\left\|v_n^1 - v_n^2\right\|_{\mathcal{X}_n} \leq S \left\|F_n(v_n^1) - F_n(v_n^2)\right\|_{\mathcal{Y}_n} \tag{13}$$

holds.

Remark 22. *Obviously, the stability on the solution of problem* (1) *(i.e., at the element* $\bar{u} \in \mathcal{X}$*) implies the semistability.*

The immediate profit of this definition is injectivity as it is formulated in the next statement.

Corollary 23. *If discretization* $\mathscr{D}$ *is stable on problem* $\mathscr{P}$ *at the element* $v \in \mathcal{X}$ *with stability threshold* R, *then* F_n *is injective on* $B_R(\varphi_n(v))$ *from some index.*

The following statements demonstrate the usefulness of the stability notion, given in Definition 21. (For more details we refer to [S73].)

Lemma 24. *We assume that*

- $\mathcal{V}$, $\mathcal{W}$ *are normed spaces with the property* $\dim \mathcal{V} = \dim \mathcal{W} < \infty$*;*
- $G : B_R(v) \to \mathcal{W}$ *is continuous, where* $B_R(v) \subset \mathcal{V}$ *is a ball for some* $v \in \mathcal{V}$ *and* $R \in (0, \infty]$*;*
- *for all* v^1, v^2 *which satisfy* $v^i \in B_R(v)$, *the stability estimate*

$$\left\|v^1 - v^2\right\|_{\mathcal{V}} \leq S \left\|G(v^1) - G(v^2)\right\|_{\mathcal{W}} \tag{14}$$

  holds.

Then

- G *is invertible, and* $G^{-1} : B_{R/S}(G(v)) \to B_R(v)$*;*
- G^{-1} *is Lipschitz continuous with the constant* S.

Proof. It is enough to show that $B_{R/S}(G(v)) \subset G(B_R(v))$, due to Corollary 23. We assume indirectly that there exists $w \in B_{R/S}(G(v))$ such that $w \notin G(B_R(v))$. We define the line $w(\lambda) = (1-\lambda)G(v) + \lambda w$ for $\lambda \geq 0$, and introduce the number $\hat{\lambda}$ as follows:

$$\hat{\lambda} := \begin{cases} \sup\left\{\lambda' > 0 \,|\, w(\lambda) \in G(B_R(v)) \,\forall \lambda \in [0, \lambda')\right\}\,, & \text{if it exists,} \\ 0\,, & \text{else.} \end{cases}$$

Then clearly the inequality $\hat{\lambda} \leq 1$ holds. We will show that $\hat{w} =: w(\hat{\lambda}) \in G(B_R(v))$.

For $\hat{\lambda} = 0$ this trivially holds. For $\hat{\lambda} > 0$ we observe that G is invertible on $w(\hat{\lambda} - \varepsilon)$, (i.e., the operators $G^{-1}(w(\hat{\lambda} - \varepsilon)) \in B_R(v)$ exist) for all $\varepsilon : \hat{\lambda} \geq \varepsilon > 0$. Thus, we can use the stability estimate (14)

$$\left\| G^{-1}(w(\hat{\lambda} - \varepsilon)) - v \right\|_{\mathcal{V}} \leq S \left\| w(\hat{\lambda} - \varepsilon) - G(v) \right\|_{\mathcal{W}} =$$
$$S(\hat{\lambda} - \varepsilon) \underbrace{\| w - G(v) \|_{\mathcal{W}}}_{= \frac{R}{S} - \frac{\delta}{S}} < \hat{\lambda}(R - \delta) \leq R - \delta \,,$$

for some $\delta > 0$, and using again the stability estimate we can conclude that the function $h(\varepsilon) = G^{-1}(w(\hat{\lambda} - \varepsilon))$ is uniformly continuous at $\varepsilon \in (0, \hat{\lambda}]$. Thus, there exists $\lim_{\varepsilon \searrow 0} h(\varepsilon) =: z \in B_R(v)$. Using the continuity of G, we get $G(z) = \hat{w}$.

Now we can choose a closed ball $\bar{B}_r(z) \subset B_R(v)$, $(r > 0)$ whose image $G(\bar{B}_r(z))$ contains a neighborhood of $\hat{w}$, due to the Brouwer's invariance domain theorem. This results in a contradiction.

Finally, the Lipschitz continuity with the constant S is a simple consequence of (14). $\square$

Lemma 25. *For the discretization $\mathscr{D}$ we assume that*

- *it is consistent and stable at $\bar{u}$ with stability threshold R and constant S on problem $\mathscr{P}$;*
- *Assumptions 10 and 11 are satisfied.*

Then the discretization $\mathscr{D}$ generates a numerical method $\mathscr{N}$ such that equation (5) has a unique solution in $B_R(\varphi_n(\bar{u}))$, from some index.

Proof. Due to Lemma 24, F_n is invertible, and $F_n^{-1} : B_{R/S}(F_n(\varphi_n(\bar{u}))) \to B_R(\varphi_n(\bar{u}))$. Note that $F_n(\varphi_n(\bar{u})) = l_n \to 0$, due to the consistency. This means that $0 \in B_{\frac{R}{S}}(F_n(\varphi_n(\bar{u})))$, from some index. This proves the statement. $\square$

Hence, we can formulate our main result.

Theorem 26. *We assume that*

- *the discretization $\mathscr{D}$ is consistent and stable at $\bar{u}$ with stability threshold R and constant S on problem $\mathscr{P}$;*
- *Assumptions 10 and 11 are true.*

Then the discretization $\mathscr{D}$ is convergent on problem $\mathscr{P}$, and the order of the convergence is not less than the order of consistency.

Proof. The statement is the consequence of Lemmas 25 and 20. $\square$

3.4 Some remarks on the stability notion

We finish this section with some remarks w.r.t. the stability notion by the Definition 21.

There are other definitions of the stability in the literature, these are mostly generalizations of the stability notion of Keller. We list two of them.

The first one of them is the following one, which is given by Stetter in [S73].

Definition 27. *The discretization $\mathscr{D}$ is called* stable in the sense of Stetter *on problem $\mathscr{P}$ if there exist $S \in \mathbb{R}$, $R \in (0, \infty]$ and $r \in (0, \infty]$ such that*

- $B_R(\varphi_n(\bar{u})) \subset \mathcal{D}_n$ *holds from some index;*
- *for all $(v_n^1)_{n\in\mathbb{N}}, (v_n^2)_{n\in\mathbb{N}}$ such that $v_n^i \in B_R(\varphi_n(\bar{u}))$, and the inclusion $F_n(v_n^i) \in B_r(F_n(\varphi_n(\bar{u})))$ is true, the estimate*

$$\left\|v_n^1 - v_n^2\right\|_{\mathcal{X}_n} \leq S \left\|F_n(v_n^1) - F_n(v_n^2)\right\|_{\mathcal{Y}_n}$$

holds.

Note that the stability notion by Stetter is less restrictive than the one given in Definition 21: if we put $r = \infty$ in Definition 27, then we re-obtain the stability definition by Keller, given in Definition 21.

The second one was given in the paper [LS88] by L**v**dż˝pez-Marcos and Sanz-Serna.

Definition 28. *The discretization $\mathscr{D}$ is called* stable in the sense of L**v**dż˝pez-Marcos and Sanz-Serna *on problem $\mathscr{P}$ if there exist $S \in \mathbb{R}$ and $(R_n)_{n\in\mathbb{N}}$, $R_n \in (0, \infty]$ such that*

- $B_{R_n}(\varphi_n(\bar{u})) \subset \mathcal{D}_n$ *holds from some index;*
- $\forall (v_n^1)_{n\in\mathbb{N}}, (v_n^2)_{n\in\mathbb{N}}$ *which satisfy $v_n^i \in B_{R_n}(\varphi_n(\bar{u}))$ from that index, the estimate*

$$\left\|v_n^1 - v_n^2\right\|_{\mathcal{X}_n} \leq S \left\|F_n(v_n^1) - F_n(v_n^2)\right\|_{\mathcal{Y}_n}$$

holds.

This stability notion allows us to vary the radius of the balls.

The third one is given in the book [T80].

Definition 29. *The discretization $\mathscr{D}$ is called* stable in the sense of Trenogin *if there exist a continuous, strictly monotonically increasing function $\omega(t)$ defined on $t \geq 0$ such that $\omega(0) = 0$ and $\omega(\infty) = \infty$, and*

$$\left\|F_n(v_n^1) - F_n(v_n^2)\right\|_{\mathcal{Y}_n} \geq \omega\left(\left\|v_n^1 - v_n^2\right\|_{\mathcal{X}_n}\right)$$

holds for all $v_n^1, v_n^2 \in \mathcal{D}_n$.

4 Basic Notions – Revisited from the Application Point of View

Theorem 26 is not yet suitable for our purposes: the condition requires to check the stability and the consistency on the unknown element $v = \bar{u}$. Therefore, this statement is not applicable for real problems. Since we are able to verify the above properties on *some set of points* (sometimes on the entire $\mathcal{D}$), we extend the previously given pointwise (local) definitions to the set (global) ones.

Definition 30. *The discretization $\mathscr{D}$ is called* consistent *on problem $\mathscr{P}$ if there exists a set $\mathcal{D}_0 \subset \mathcal{D}$ whose image $F(\mathcal{D}_0)$ is dense in some neighborhood of the point $0 \in \mathcal{Y}$, and it is consistent at each element $v \in \mathcal{D}_0$.*

The order of the consistency in $\mathcal{D}_0$ is defined as $\inf\{p_v : v \in \mathcal{D}_0\}$, *where p_v denotes the order of consistency at the point v.*

Example 31. *Let us consider the explicit Euler method, given in Examples 4, 6 and 8. We apply it to the Cauchy problem of Example 2, i.e., to the problem (3)-(4). We verify the consistency and its order on the set $\mathcal{D}_0 \subset \mathcal{D}$, where $\mathcal{D} := C^1[0,1]$ and $\mathcal{D}_0 := C^2[0,1]$. Then for the local discretization error we obtain*

$$[F_n(\varphi_n(v)) - \psi_n(F(v))](t_i) = \begin{cases} \frac{1}{2n} v''(\theta_i) & i = 1, \dots, n, \\ \\ 0, & i = 0, \end{cases} \tag{15}$$

where $\theta_i \in (t_{i-1}, t_i)$ are given numbers. Then $\|l_n(v)\|_{\mathcal{X}_n} = \mathcal{O}(n^{-1})$ from Definition 16.

Hence, for the class of problems (3)-(4) with Lipschitz continuous right-hand side f, the explicit Euler method is consistent, and the order of the consistency equals one.

In Section 3 (c.f. Example 18) we have shown that the pointwise consistency at the solution in itself is not enough for the convergence. One may think that the stronger notion of consistency, given by Definition 30, already ensures convergence. The following example shows that this is not true.

Example 32. *Let us choose the normed spaces as $\mathcal{X} = \mathcal{X}_n = \mathcal{Y} = \mathcal{Y}_n = \mathbb{R}$, $\varphi_n, = \psi_n = identity$. Our aim is to solve the scalar equation $F(x) = 0$, where the function $F \in C(\mathbb{R}, \mathbb{R})$ is given as follows*

$$F(x) = \begin{cases} |x| \,, \ \textit{if } x \in (-1,1)\,, \\ 1\,, \ \textit{if } x \in (-\infty, -1] \cup [1, \infty)\,. \end{cases}$$

Clearly this problem has a unique solution $\bar{x} = 0$. We define the numerical method $\mathscr{N}$ as

$$F_n(x) = \begin{cases} \frac{1}{n}\,, \text{ if } x \in \left[-\frac{1}{n}, \frac{1}{n}\right], \\ x\,, \text{ if } x \in \left(\frac{1}{n}, 1\right), \\ 1\,, \text{ if } x \in (-\infty, -1] \cup [1, n) \cup [n+2, \infty)\,, \\ -x\,, \text{ if } x \in \left(-1, -\frac{1}{n}\right), \\ |x - (n+1)|\,, \text{ if } x \in [n, n+2)\,. \end{cases}$$

For the given problem this discretization is consistent on the entire $\mathbb{R}$, however it is not convergent, since the solutions of the discrete problems $F_n(x) = 0$ are $\bar{x}_n = n+1$ and therefore $\bar{x}_n \nrightarrow \bar{x}$.

In the sequel, besides the Assumptions 10, 11, which we have already made, we assume the validity of the following new assumptions.

Assumption 33. *For the problem $\mathscr{P}$ we assume that F^{-1} is continuous at the point $0 \in \mathcal{Y}$.*

Assumption 34. *Let us apply the discretization $\mathscr{D}$ to problem $\mathscr{P}$. We assume that discretization $\mathscr{D}$ possesses the property: there exists $K_1 > 0$ such that for all $v \in \mathcal{D}$ the relation*

$$\|\varphi_n(\bar{u}) - \varphi_n(v)\|_{\mathcal{X}_n} \le K_1 \|\bar{u} - v\|_{\mathcal{X}}$$

holds for all $n \in \mathbb{N}$.

Assumption 35. *We assume that discretization $\mathscr{D}$ possesses the property: there exists $K_2 > 0$ such that for all $y \in \mathcal{Y}$ the relation*

$$\|\psi_n(y) - \psi_n(0)\|_{\mathcal{Y}_n} \le K_2 \|y - 0\|_{\mathcal{Y}}$$

holds for all $n \in \mathbb{N}$.

For the simplicity of the formulation, the collection of the Assumptions 9–11 and 33–35 will be called Assumption $A^\star$.

Lemma 36. *Besides Assumption $A^\star$ we assume that*

- *the discretization $\mathscr{D}$ on problem $\mathscr{P}$ is consistent,*
- *the discretization $\mathscr{D}$ on problem $\mathscr{P}$ at the element $\bar{u}$ is stable with stability threshold R and constant S.*

Then F_n is invertible at the point $\psi_n(0)$, i.e., there exists $F_n^{-1}(\psi_n(0))$ for sufficiently large indices n.

Proof. We can choose a sequence $(y^k)_{k\in\mathbb{N}}$ such that $y^k \to 0 \in \mathcal{Y}$ and $F^{-1}\left(y^k\right) =: u^k \to \bar{u}$, due to the continuity of F^{-1}. Then the discretization $\mathscr{D}$ on problem $\mathscr{P}$ at the element u^k is stable with stability threshold $R/2$ and constant S, for some sufficiently large indices k. Moreover, F_n is continuous on $B_{R/2}(\varphi_n(u^k))$. Thus, for these indices k and also for sufficiently large n there exists $F_n^{-1} : B_{R/2S}(F_n(\varphi_n(u^k))) \to B_{R/2}(\varphi_n(u^k))$ moreover, it is Lipschitz continuous with constant S, according to Lemma 24. Let us write a trivial upper estimate:

$$\left\|F_n(\varphi_n(u^k))\right\|_{\mathcal{Y}_n} \leq \left\|F_n(\varphi_n(u^k)) - \psi_n(F(u^k))\right\|_{\mathcal{Y}_n} + \left\|\psi_n(F(u^k))\right\|_{\mathcal{Y}_n}.$$

Here the first term tends to 0 as $n \to \infty$, due to the consistency. For the second term, based on (35) we have the estimate $\left\|\psi_n(y^k)\right\|_{\mathcal{Y}_n} \leq K_2 \left\|y^k\right\|_{\mathcal{X}_n}$. Since the right-hand side tends to zero as $k \to \infty$, this means that the centre of the ball $B_{R/2}(F_n(\varphi_n(u^k)))$ tends to $0 \in \mathcal{Y}_n$, which proves the statement. □

Corollary 37. *Under the conditions of Lemma 36, for sufficiently large indices k and n, the following results are true.*

- *There exists $F_n^{-1}(\psi_n(y^k))$, since $\psi_n(y^k) \in B_{R/2S}(F_n(\varphi_n(u^k)))$.*
- *$F_n^{-1}(\psi_n(y^k)), \varphi_n(F^{-1}(y^k)) \in B_{R/2}(\varphi_n(\bar{u}))$.*

Analogously to the consistency, the stability can also be defined on a set of points. (This makes it possible to avoid the direct knowledge of the usually unknown $\bar{u}$.)

Definition 38. *The discretization $\mathscr{D}$ is called* stable *on problem $\mathscr{P}$ if there exist $S \in \mathbb{R}$, $R \in (0, \infty]$ and a set $\mathcal{D}_1 \subset \mathcal{D}$ such that $\bar{u} \in \mathcal{D}_1$ and it is stable at each point $v \in \mathcal{D}_1$ with stability threshold R and constant S.*

Now we are in the position to formulate our basic result, in which the notion of convergence is ensured by the notions of consistency and stability on a set, which can usually be verified directly, without knowing the exact solution of problem $\mathcal{P}$.

Theorem 39. *Besides the Assumption $A^\star$ we suppose that the discretization $\mathscr{D}$ on problem $\mathscr{P}$ is*

- *consistent;*
- *stable with stability threshold R and constant S.*

Then the discretization $\mathscr{D}$ is convergent on problem $\mathscr{P}$, and the order of the convergence can be estimated from below by the order of consistency on the corresponding set $\mathcal{D}_0$.

Proof. By use of the triangle inequality, we have

$$
\begin{aligned}
&\|\varphi_n(\bar{u}) - \bar{u}_n\|_{\mathcal{X}_n} = \left\|\varphi_n(F^{-1}(0)) - F_n^{-1}(\psi_n(0))\right\|_{\mathcal{X}_n} \le \\
&\underbrace{\left\|\varphi_n(F^{-1}(0)) - \varphi_n(F^{-1}(y^k))\right\|_{\mathcal{X}_n}}_{I.} + \\
&\underbrace{\left\|\varphi_n(F^{-1}(y^k)) - F_n^{-1}(\psi_n(y^k))\right\|_{\mathcal{X}_n}}_{II.} + \\
&\underbrace{\left\|F_n^{-1}(\psi_n(y^k)) - F_n^{-1}(\psi_n(0))\right\|_{\mathcal{X}_n}}_{III.},
\end{aligned}
\tag{16}
$$

where the elements $y^k \in \mathcal{Y}$ are defined in the proof of Lemma 36.

In the next step we estimate the different terms on the left-hand side of (16).

I. For the first term, based on Assumption 34, we have the estimate

$$
\left\|\varphi_n(F^{-1}(0)) - \varphi_n(F^{-1}(y^k))\right\|_{\mathcal{X}_n} \le K_1 \left\|F^{-1}(0) - F^{-1}(y^k)\right\|_{\mathcal{X}} .
$$

Since $y^k \to 0$ as $k \to \infty$, and F^{-1} is continuous at the point $0 \in \mathcal{Y}$, therefore this term tends to zero, independently of n.

II. This term can be written as $\left\|F_n^{-1}(F_n(\varphi_n(F^{-1}(y^k)))) - F_n^{-1}(\psi_n(y^k))\right\|_{\mathcal{X}_n}$. Due to Corollary 37, we can use the stability estimate, therefore for this term we have the estimate

$$
\begin{aligned}
&\left\|\varphi_n(F^{-1}(y^k)) - F_n^{-1}(\psi_n(y^k))\right\|_{\mathcal{X}_n} \le \\
&S\left\|F_n(\varphi_n(F^{-1}(y^k))) - \psi_n(y^k)\right\|_{\mathcal{Y}_n} = S\left\|F_n(\varphi_n(u^k)) - \psi_n(F(u^k))\right\|_{\mathcal{Y}_n} .
\end{aligned}
$$

In this estimate the term on the right-hand side tends to zero because of the consistency at u^k.

III. For the estimation of the third term we can use the Lipschitz continuity of F_n^{-1}, due to Lemma 36 and Corollary 37. Hence, by using the Assumption 35, we have

$$
\left\|F_n^{-1}(\psi_n(y^k)) - F_n^{-1}(\psi_n(0))\right\|_{\mathcal{X}_n} \le S\left\|\psi_n(y^k) - \psi_n(0)\right\|_{\mathcal{Y}_n} \le SK_2\left\|y^k\right\|_{\mathcal{Y}} .
$$

The right-hand side of the above estimate tends to zero, independently of the index n.

These estimations complete the proof. ☐

Example 40. *Let us analyze the stability property of the explicit Euler method, given in Example 8.*

Let $\mathbf{v}^{(1)}, \mathbf{v}^{(2)} \in \mathcal{X}_n = \mathbb{R}^{n+1}$ *be two arbitrary vectors, and we use the notation* $\epsilon = \mathbf{v}^{(1)} - \mathbf{v}^{(2)} \in \mathbb{R}^{n+1}$. *We define the vector* $\delta = F_n\left(\mathbf{v}^{(1)}\right) - F_n\left(\mathbf{v}^{(2)}\right) \in \mathbb{R}^{n+1}$, *where* F_n *is defined in (6). (In the notation, for simplicity, we omit the use of the subscript* n *for the vectors. We recall that the coordinates of the vectors are numbered from* $i = 0$ *until* $i = n$.)

For the coordinates of the vector δ *we have the following relations.*

- *For the first coordinate (*$i = 0$*) we obtain:*

$$\delta_0 = \left(F_n\left(\mathbf{v}^{(1)}\right)\right)_0 - \left(F_n\left(\mathbf{v}^{(2)}\right)\right)_0 = \left(v_0^{(1)} - u_0\right) - \left(v_0^{(2)} - u_0\right) = \epsilon_0.$$

- *For the other coordinates* $i = 1, \ldots, n$ *we have*

$$\begin{gathered} \delta_i = v_i^{(1)} - v_i^{(2)} = \\ n(v_i^{(1)} - v_{i-1}^{(1)}) - f(v_{i-1}^{(1)}) - n(v_i^{(2)} - v_{i-1}^{(2)}) + f(v_{i-1}^{(2)}) = \\ n(v_i^{(1)} - v_i^{(2)}) - n(v_{i-1}^{(1)} - v_{i-1}^{(2)}) - (f(v_{i-1}^{(1)}) - f(v_{i-1}^{(2)})) = \\ n\epsilon_i - n\epsilon_{i-1} - (f(v_{i-1}^{(1)}) - f(v_{i-1}^{(2)})). \end{gathered}$$

We can express ϵ_i *from this relation as follows:*

$$\epsilon_i = \epsilon_{i-1} + \frac{1}{n}(f(v_{i-1}^{(1)}) - f(v_{i-1}^{(2)})) + \frac{1}{n}\delta_i.$$

Under our assumption $f \in C(\mathbb{R}, \mathbb{R})$ *is a Lipschitz continuous function, therefore we have the estimation* $|f(v_{i-1}^{(1)}) - f(v_{i-1}^{(2)})| \leq L|v_{i-1}^{(1)} - v_{i-1}^{(2)}|$. *Hence, we get*

$$|\epsilon_i| \leq |\epsilon_{i-1}| + \frac{1}{n}L|v_{i-1}^{(1)} - v_{i-1}^{(2)}| + \frac{1}{n}|\delta_i| = |\epsilon_{i-1}|\left(1 + \frac{L}{n}\right) + \frac{1}{n}|\delta_i|.$$

If we apply this estimate consecutively to $|\epsilon_{i-1}|$, $|\epsilon_{i-2}|$, *etc., we obtain:*

$$\begin{aligned} |\epsilon_i| \leq |\epsilon_{i-2}|\left(1 + \frac{L}{n}\right)^2 + \frac{1}{n}|\delta_i| + \left(1 + \frac{L}{n}\right)\frac{1}{n}|\delta_{i-1}| \leq \ldots \\ |\epsilon_0|\left(1 + \frac{L}{n}\right)^n + \frac{1}{n}\sum_{i=1}^{n}|\delta_i|\Big(1 + \frac{L}{n}\Big)^{n-i}. \end{aligned} \tag{17}$$

Since $\delta_0 = \epsilon_0$ *and* $\left\|\mathbf{v}^{(1)} - \mathbf{v}^{(2)}\right\|_{\mathcal{X}_n} = \max_{i=0,\ldots,n} |\epsilon_i|$, *hence we can write our estimation in the form*

$$\left\|\mathbf{v}^{(1)} - \mathbf{v}^{(2)}\right\|_{\mathcal{X}_n} \leq |\delta_0| \left(1 + \frac{L}{n}\right)^n + \frac{1}{n} \sum_{i=1}^{n} |\delta_i| \Big(1 + \frac{L}{n}\Big)^{n-i} \tag{18}$$

$$< e^L(\delta_0 + \max_{i=1,\dots,n} |\delta_i|) = e^L \left\|\delta\right\|_{\mathcal{Y}_n} = e^L \left\|F_n\left(\mathbf{v}^{(1)}\right) - F_n\left(\mathbf{v}^{(2)}\right)\right\|_{\mathcal{Y}_n}. \tag{19}$$

This shows us that the discretization (8), i.e., the explicit Euler method is stable on the whole set $\mathcal{X} = C^1[0,1]$ with $S = e^L$ and $R = \infty$.

Hence, based on Theorem 39, the results of this example and Example 31, we can conclude that the explicit Euler method is convergent, and the order of its convergence is one.

Remark 41. *The stability property of the explicit Euler method in the other stability senses can be proven in the same way. (E.g. the Trenogin's stability of the explicit Euler method is shown on [T80], and the proof is very similar to the proof in Example 40.)*

5 Relation between consistency, stability and convergence

Theorem 39 shows that, under the Assumption $A^\star$, the consistency and stability of discretization $\mathscr{D}$ on problem $\mathscr{P}$ result in the convergence, i.e., consistency and stability together are a sufficient condition for convergence. (Roughly speaking, this implication is shown in (2).) However, from this observation we cannot get an answer to the question of the necessity of these conditions.

In the sequel, we raise a more general question: What is the general relation between the above listed three basic notions? Since each of them can be true (T) or false (F), we have to consider eight different cases, listed in Table 1.

Before giving the answer, we consider some examples. In each examples $\mathcal{X} = \mathcal{X}_n = \mathcal{Y} = \mathcal{Y}_n = \mathbb{R}$, $\mathcal{D} = \mathcal{D}_n = [0,\infty)$, $\varphi_n = \psi_n = identity$. Our aim is to solve the scalar equation

$$F(x) \equiv x^2 = 0\,, \tag{20}$$

which has the unique solution $\bar{x} = 0$.

Example 42. *For solving equation* (20) *we choose the numerical method defined by the n-th Lagrangian interpolation, i.e., $F_n(x)$ is the Lagrangian interpolation polynomial of order n. Since the Lagrange interpolation is exact for $n \geq 2$, therefore $F_n(x) = x^2$ holds for all $n \geq 2$. Hence, clearly the numerical method is consistent and convergent. The operator F_n^{-1} can be defined easily, and it is $F_n^{-1}(x) = \sqrt{x}$. Hence its derivative is not bounded around the point $\bar{x} = 0$, therefore the numerical method is not stable.*

	consistency	stability	convergence
1	T	T	T
2	T	T	F
3	T	F	T
4	T	F	F
5	F	T	T
6	F	T	F
7	F	F	T
8	F	F	F

Table 1: The list of the different cases (T: true, F: false).

Example 43. *For solving equation* (20) *we choose now the numerical method* $F_n(x) = 1 - nx$. *The roots of the discrete equations* $F_n(x) = 0$ *are* $\bar{x}_n = 1/n$, *therefore* $\bar{x}_n \to \bar{x} = 0$ *as* $n \to \infty$. *This means that the numerical method is convergent. We observe that* $\varphi_n(F_n(0)) = \varphi_n(1) = 1$, *and* $\psi_n(F(0)) = \psi_n(0) = 0$. *Hence, for the local discretization error we have* $|l_n| = 1$, *for any index* n. *This means that the numerical method is not consistent. One can easily check that* F_n *is invertible, and* $F_n^{-1}(x) = -x/n + 1/n$. *Hence the derivative of the inverse operators are uniformly bounded on* $[0, \infty)$ *by* 1 *for any* n. *Therefore the numerical method is stable.*

Example 44. *For solving equation* (20) *we choose the following numerical method:* $F_n(x) = 1 - nx^2$. *Then* $\bar{x}_n = 1/\sqrt{n}$, *and hence* $\bar{x}_n \to \bar{x} = 0$ *as* $n \to \infty$. *This means that the numerical method is convergent. Due to the relations* $\varphi_n(F_n(0)) = \varphi_n(1) = 1$ *and* $\psi_n(F(0)) = \psi_n(0) = 0$, *this method is not consistent. Since for this numerical method* $F_n^{-1}(x) = \sqrt{(1-x)/n}$, *therefore the derivatives are not bounded. Therefore the numerical method is not stable.*

Now, we are in the position to answer the question, posed at beginning of this section. Using the numeration of the different cases in Table 1, the answers are included in Table 2. (We note that two cases (case 6 and 8 in Table 1) are uninteresting from a practical point of view, therefore we have neglected their investigation.) The results particularly show that neither consistency, nor stability is a necessary condition for the convergence.

6 Summary

We have considered the numerical solution of non-linear equations in an abstract (Banach space) setting. The main aim was to guarantee the convergence of the

number of the case	answer	reason
1	always true	Theorem 39
2	always false	Theorem 39
3	possible	Example 42
4	possible	Examples 18 and 32
5	possible	Example 43
6	n.a.	n.a.
7	possible	Example 44
8	n.a.	n.a.

Table 2: The possibility of the different cases.

numerical process. It was shown that, similarly to the linear case, this notion can be guaranteed by two notions: the consistency and the stability together ensure the convergence. In the linear case this result is well known as the Lax (or sometimes Lax-Richtmyer-Kantorovich) theory. From the formulation of the main theorem it turns out that these two, directly checkable conditions (i.e., the consistency and stability) serve together as a sufficient condition of the convergence. However, even in the linear theory, the necessity of these conditions is less investigated. By giving suitable examples we have shown that neither consistency, nor stability is necessary for the convergence, in general. As an example for the theory, we have investigated the numerical solution of a Cauchy problem for ordinary differential equations by means of the explicit Euler method. We have shown the first order consistency and the stability of this method, which, based on the basic theorem, yield first order convergence. (We note that, as opposed to the usual direct proof of the convergence of the explicit Euler method, the convergence in this example yields the convergence on the whole space-time domain, and not only at some fixed time level $t = t^*$.)

In the further works we plan to apply this developed theory to linear problems, and compare the results to the Lax theory. Moreover, our aim is to extend the non-linear theory by generalization of the stability notion. We also intend to apply the results of the non-linear theory to other, more complex problems, like boundary value problems of ordinary and partial differential equations, as well.

6.1 Acknowledgment

The paper is supported by the Hungarian Scientific Research Fund OTKA grant No. K67819. The European Union and the European Social Fund have provided financial support to the project under the grant agreement no. TÁMOP 4.2.1./B-09/1/KMR-2010-0003.

This work is connected to the scientific program of the ”Development of quality-oriented and harmonized R+D+I strategy and functional model at BME” project. This project is supported by the New Hungary Development Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002).

References

[K75] Keller, H. B.: Approximation Methods for Nonlinear Problems with Application to Two-Point Boundary Value Problems.
Math. Comput., 130, 464–474 (1975)

[LR56] Lax, P. D. and Richtmyer, R. D.: Survey of Stability of Linear Finite Difference Equations.
Comm. Pure Appl. Math., 9, 267–293 (1956)

[LS88] López-Marcos, J. C. and Sanz-Serna, J. M.: Stability and Convergence in Numerical Analysis III: Linear Investigation of Nonlinear Stability.
IMA J. Numer. Anal., 8, 71–84 (1988)

[PS84a] Palencia, C. and Sanz-Serna, J. M.: An Extension of the Lax-Richtmyer Theory.
Numer. Math., 44, 279–283 (1984)

[PS84b] Palencia, C. and Sanz-Serna, J. M.: Equivalence Theorems for Incomplete Spaces: an Appraisal.
IMA J. Numer. Anal., 4, 109–115 (1984)

[PS85] Palencia, C. and Sanz-Serna, J. M.: A General Equivalence Theorem in the Theory of Discretization Methods.
Math. of Comp., 45/171, 143–152 (1985)

[S73] Stetter, H. J.: Analysis of Discretization Methods for Ordinary Differential Equations.
Springer, Berlin, (1973)

[T80] Trenogin, V. A.: Functional Analysis.
Nauka, Moscow, (1980) (in Russian)

(Received July 31, 2011)