RAIRO
ANALYSE NUMÉRIQUE

VLASTIMIL PTÁK

What should be a rate of convergence ?

RAIRO – Analyse numérique, tome 11, nº 3 (1977), p. 279-286.
<http://www.numdam.org/item?id=M2AN_1977__11_3_279_0>

R.A.I.R.O. Analyse Numérique/Numerical Analysis
(vol. 11, n° 3, 1977, p. 279 à 286)

WHAT SHOULD BE A RATE OF CONVERGENCE ? (1)

Vlastimil PTÁK (2)

Communiqué par F. ROBERT

Abstract. — *An attempt to set up a method for estimating convergence of iterative processes using inequalities of the type*

$$d(x_{n+2}, x_{n+1}) \leqslant \omega(d(x_{n+1}, x_n)),$$

where ω *is a "small" function as defined by the author. This method is compared with the classical notion of a rate of convergence; for convex functions* ω *inequalities using the distance of* x_n *from the solution and those using the distance of two consecutive steps are shown to be equivalent. Some advantages of this new approach are pointed out.*

1. HEURISTICS

The classical notion of the order of convergence or rate of convergence which reputedly goes back to the last century is defined as follows. Given an iterative process which yields a sequence x_n of elements of a complete metric space (E, d) converging to an element $x \in E$ we say that the convergence is of order p if there exists a constant α such that

$$d(x_{n+1}, x) \leqslant \alpha(d(x_n, x))^p$$

Clearly it is immaterial whether we require this for all n or only asymptotically. Let us point out two difficulties which seem to arise if this point of view is adopted.

1° If $p > 1$ then the above inequality contains a certain amount of information about the process; the information, however, is more of a qualitative nature since it relates quantities which we are not able to measure at any finite stage of the process. The obvious meaning of the above inequality seems to consist rather in the fact that, at each stage of the process, the following step of the iteration yields a significant improvement of the estimate.

(1) Text of a lecture (unpublished) given at the Dublin Conference on Numerical Analysis. August 1976. Manuscrit reçu le 10 novembre 1976.

(2) Vlastimil Pták, Ceskoslovenska Akademie Ved, Matematicky Ustav, Praha, Tchécoslovaquie.

2° Theoretical considerations enable us, in many cases, to establish an inequality of the above type for certain constants α and p; however, usually this is only possible if we assume n to be larger than a certain bound. We might want, however, to stop the process before this bound is reached — in this case the inequality cannot be used. Of course, it is possible to extend the validity of the estimate to all n by making α sufficiently large — this may invalidate its practical applicability for the initial steps.

It seems therefore reasonable to look for another method of estimating the convergence of iterative processes, one which would satisfy the following requirements.

1° it should relate quantities which may be measured or estimated during the actual process

2° it should describe accurately in particular the initial stage of the process, not only its asymptotic behaviour since, after all, we are interested in keeping the number of steps necessary to obtain a good estimate as low as possible.

We would like to call the attention of the specialists to a method proposed by the author with the aim of satisfying the above postulates.

It is obvious that wa cannot expect to have an adequate description of both the beginning and the tail end of the process by any formula as simple as the one we discussed above. In our opinion, a description which fits the whole process, not only an asymptotic one, is only possible by means of suitable functions, not just numbers.

We therefore propose a method based on looking for positive functions ω (defined for small positive arguments) which relate two consecutive increments of the process by an inequality of the following type

$$d(x_{n+1}, x_n) \leqslant \omega(d(x_n, x_{n-1}))$$

By allowing a larger class of functions than just those of the type $t \to \alpha t^p$ we have a better chance of getting a closer fit of the estimates even at the beginning of the process.

At the same time this approach measures the rate of convergence at finite stages of the process using only data available at that particular stage of the process; in fact, instead of comparing the two unknown quantities $d(x_n, x)$ and $d(x_{n+1}, x)$ it is based on the relation between $d(x_n, x_{n-1})$ and $d(x_{n+1}, x_n)$.

Suppose we have a sequence of inequalities

$$d(x_{n+k}, x_{n+k-1}) \leqslant \omega^{(k-1)}(d(x_{n+1}, x_n))$$

for $k = 1, 2, \ldots$ (where $\omega^{(j)}$ stands for the $j - th$ iteration of the function ω) and that the series $\sum_0^\infty \omega^{(j)}(d(x_{n+1}, x_n))$ is convergent. Such a sequence of inequalities may be deduced from the above inequality if ω is an increasing function. Then the sequence $x_n, x_{n+1}, \ldots$ is a fundamental sequence and, the space (E, d) being complete, converges to a limit x for which

$$d(x_n, x) \leqslant d(x_{n+1}, x_n) + d(x_{n+2}, x_{n+1}) + \ldots \leqslant \sum_0^\infty \omega^{(j)}(d(x_{n+1}, x_n))$$

As an example, let us mention the rate of convergence of Newton's process recently established by the author. There we have

$$\omega(t) = \frac{t^2}{2(t^2 + d)^{1/2}}$$

where d is a positive constant depending on the data of the problem. A closer inspection of this formula shows that, for very small t, the function assumes approximately the form $\frac{t^2}{2d^{1/2}}$ whereas, for large t, the summand t^2 predominates in the denominator to that the function is approximatively linear, $\frac{1}{2}t$.

Since ω relates the consecutive steps of Newton's process by the inequality $d(x_{n+1}, x_n) \leqslant \omega(d(x_n, x_{n-1}))$ this shows first that, asymptotically – in other words for small $d(x_n, x_{n-1})$ – the next increment is approximately $\frac{1}{2d^{1/2}}(d(x_n, x_{n-1}))^2$. This phenomenon is usually described by saying that the convergence is quadratic.

However, in the initial stages of the process $d(x_n, x_{n-1})$ is still large so that ω is almost linear. Since it may be shown that the estimates for Newton's process process based on ω are sharp at each step, it follows that accurate estimates valid for the whole process – including the initial steps – cannot be based on any simple quadratic monomial.

Having explained the motivation, let us pass now to precise formulations.

(1.1) Definition : Let T be an interval of the form $T = \{ t ; 0 < t < t_0 \}$ for some positive t_0. A rate convergence on T is a function ω defined on T with the following properties

1. ω maps T into itself
2. for each $t \in T$ series $t + \omega(t) + \omega^{(2)}(t) + \ldots$ is convergent.

We use the abbreviation $\omega^{(n)}$ for the *n-th* iterate of the function ω, so that $\omega^{(2)}(t) = \omega(\omega(t))$ and so on. The sum of the above series will be denoted by σ. The function σ satisfies the following functional equation

$$\sigma(t) - t = \sigma(\omega(t));$$

one of the consequences of this fact is the possibility of recovering ω if σ is given. Indeed,

$$\omega(t) = \sigma^{-1}(\sigma(t) - t).$$

This notion plays a fundamental role in the method of non-discrete mathematical induction, a new approach to iterative existence theorems.

The method of nondiscrete mathematical induction, inaugurated by the author in [3] and [4], was intended to give a general, abstract model for iterative constructions in mathematical analysis and numerical analysis. It turned out that such a model may be based on a simple result about families of sets which represents a quantitative refinement of the closed graph theorem. We call this result the Induction Theorem.

Let us restate here for the convenience of the reader the induction theorem and explain the notation. Given a metric space (E, d) with distance function d, a point $x \in E$ and a positive number r, we denote by $U(x, r)$ the open spherical neighbourhood of x with radius r, $U(x, r) = \{ y \in E; d(y, x) < r \}$. Similarly, if $M \subset E$, we denote by $U(M, r)$ the set of all $y \in E$ for which $d(y, M) < r$. If we are given, for each sufficiently small positive r, a set $A(r) \subset E$, we define the limit $A(0)$ of the family $A(.)$ as follows

$$A(0) = \bigcap_{s>0} \left(\bigcup_{r \leq s} A(r) \right).$$

Now we may state the Induction Theorem.

(1.2) Theorem : *Let (E, d) be a complete metric space, let T be an interval $\{ t; 0 < t < t_0 \}$ and ω a rate of convergence on T. For each $t \in T$ let $Z(t)$ be a subset of E; denote by $Z(0)$ the limit of the family $Z(.)$. Suppose that*

$$Z(t) \subset U(Z(\omega(t)), t)$$

for each $t \in T$. Then

$$Z(t) \subset U(Z(0), \sigma(t))$$

for each $t \in T$.

The theorem is closely related to the closed graph theorem in functional analysis which is nothing more than a limit case of the induction theorem. The proof of the induction theorem is very simple and, moreover, is analogous to the proof of the closed graph theorem. The proof may be found in [4] where the relation of these two theorems is discussed or in the Gatlinburg Lecture [5] where the principles governing its applications are expounded.

The method of nondiscrete mathematical induction has been applied successfully to obtain improvements of selection theorems [4], transitivity theorems in the theory of C^*-algebras [4], [10], factorization theorems in

Banach algebras [3], [1] and existence theorems in the theory of partial differential equations [9], [2].

In the Gatlinburg Lecture [5] the method of nondiscrete mathematical induction was illustrated by means of the example of an iteration splitting off an eigenvalue of an almost decomposable operator. Also, this method makes it possible to obtain estimates sharp at each step for the case of Newton's process [6].

The most obvious example of a rate of convergence is that of a linear contraction, the function $\omega(t) = \alpha t$ with $0 < \alpha < 1$ on the whole positive axis. Since $\omega^{(n)}(t) = \alpha^n t$, explicit formulas for $\sigma(t)$ and $\sigma_n(t)$ are immediate.

Some existence probleme require, however, in a natural manner, more complicated rates of convergence.

Let us mention two examples.

(1.3) The function $\omega(t) = \dfrac{t^2}{2(x^2 + d)^{1/2}}$ where d is a positive number in a rate of convergence on the whole positive axis. In the author's paper [6] it is shown that function measures the convergence of Newton's process. The corresponding σ-function is computed in [6] and the finite sums σ_n in [8].

(1.4) If γ and β are positive numbers such that $\gamma^2 > 4\beta$ then

$$\omega(t) = t\frac{\gamma + t - ((\gamma + t)^2 - 4\beta)^{1/2}}{\gamma - t + ((\gamma + t)^2 - 4\beta)^{1/2}}$$

is a rate of convergence on the whole positive axis. It has been used in [5] to obtain result on the spectrum of an almost decomposable operator. The corresponding σ-function is computed in [5] and the finite sums σ_n in [7].

2. CONVEX RATES OF CONVERGENCE

Let us turn now to the problem of comparing this new method of measuring convergence with the classical notion described at the beginning.

The new method is based on comparing consecutive terms in the sequence

$$d(x_n, x_{n+1})$$

while the classical one compares consecutive terms in the sequence

$$d(x_n, x)$$

It is thus natural to ask whether estimates using consecutive distances $d(x_n, x_{n+1})$ imply similar estimates for the distances $d(x_n, x)$. More precisely, if $e_{n,n+1}$ stands for an estimate of $d(x_{n+1}, x_n)$ and e_n for an estimate of $d(x_n, x)$ we can ask whether estimates of the form $e_{n+1,n+2} \leqslant \omega(e_{n,n+1})$ imply estimates of the classical type $e_{n+1} \leqslant \omega(e_n)$. We intend to show that this is indeed so at least in the case where ω convex.

To see that, suppose we have a sequence x_n for which the estimate

$$d(x_{n+1}, x_n) \leqslant \omega(d(x_n, x_{n-1}))$$

holds. Hence

$$\begin{aligned} d(x_n, x) &\leqslant d(x_{n+1}, x_n) + d(x_{n+2}, x_{n+1}) + \dots \\ &\leqslant \sum_0^\infty \omega^{(k)}(d(x_{n+1}, x_n)) = \sigma(d(x_{n+1}, x_n)). \end{aligned}$$

Here we have used the fact that ω is nondecreassing; this is a simple consequence of the convexity of ω.

Similarly, $d(x_{n-1}, x) \leqslant \sigma(\omega(d(x_{n-1}, x_n)))$; it follows that the estimates

$$e_{n+p,n+p+1} = \omega^{(p)}(d(x_{n+1,x_n})) \qquad p = 0, 1, 2, \dots$$

and

$$e_n = \sigma(e_{n,n+1})$$

satisfy the inequalities $e_{n+1} \leqslant \sigma'(\omega(e_{n,n+1}))$.

To obtain the desirable estimate $e_{n+1} \leqslant \omega(e_n)$ it would be sufficient to have the inequality $\sigma \circ \omega \leqslant \omega \circ \sigma$ since this yields the following estimates

$$e_{n+1} \leqslant \sigma(\omega(e_{n,n+1})) \leqslant \omega(\sigma(e_{n,n+1})) = \omega(e_n).$$

This heuristic ressoning should be sufficient to explain the importance of the inequality $\sigma \circ \omega \leqslant \omega \circ \sigma$. We now proceed to a formal proof of this inequality for convex rates of convergence.

(2.1) Suppose ω is a rate of convergence on the interval T. If ω is convex, then

$$\omega \circ \sigma \geqslant \sigma \circ \omega$$

Proof : We intend to show that $\omega(\sigma(t)) \geqslant \sigma(\omega(t))$ for each $t \in T$ such that $\sigma(t)$ again belongs to T. First of all, we make the following observation.

If $0 < x \leqslant y \in T$ then

$$\omega(y)x \geqslant \omega(x)y.$$

This is an immediate consequence of the convexity of ω. We include a formal proof although the inequality is evident from a simple picture. Consider a third point z, $0 < z < x$. We have then

$$\omega(x) = \omega\left(\frac{y-x}{y-z} z + \frac{x-z}{y-z} y\right) \leqslant \frac{y-x}{y-z}\omega(z) + \frac{x-z}{y-z}\omega(y).$$

Since $\omega(z)$ tends to zero for $z \to 0$, it follows from the above inequality that $\omega(x) \leqslant \frac{x}{y}\,\omega(y)$; this establishes the inequality $\omega(y)x \geqslant \omega(x)y$. In particular, it follows that ω is nondecreasing.

Suppose now that $t \in T$ is such that $\sigma(t) \in T$. Since $\omega^{(n)}(t) \leqslant \sigma(t)$ for $n = 0, 1, 2, \ldots$ it follows from the above inequality that

$$\omega(\sigma(t))\omega^{(n)}(t) \geqslant \sigma(t)\omega^{(n+1)}(t)$$

for all $n = 0, 1, 2, \ldots$. Upon summing these inequalities and dividing by $\sigma(t)$ we obtain the desired inequality.

It is possible to obtain a somewhat sharper inequality.

(2.2) Let ω be a rate of convergence on the interval T. If ω is convex then, for each natural number n (and each $t \in T$ for which $\sigma_n(t) \in T$)

$$\omega \circ \sigma_n \geqslant \frac{\sigma_n}{\sigma_{n-1}}\,\sigma_{n-1} \circ \omega$$

Proof: For $n = 1$ this inequality is an immediate consequence of the inequality

$$\frac{\omega(t + \omega(t))}{t + \omega(t)} \geqslant \frac{\omega(t)}{t}$$

Now suppose that n is a natural number and that the $n - th$ inequality holds. We have thus

$$\omega(\sigma_n(t))\sigma_{n-1}(t) \geqslant \sigma_n(t)\sigma_{n-1}(\omega(t)) \tag{1_n}$$

At the same time,

$$\omega(\sigma_{n+1}(t))\sigma_n(t) \geqslant \omega(\sigma_n(t))\sigma_{n+1}(t) \tag{2}$$

Using (2) and (1_n) we obtain

$$\begin{aligned}\omega(\sigma_{n+1})\sigma_{n-1} &= \frac{\sigma_{n-1}}{\sigma_n}\,\omega(\sigma_{n+1})\sigma_n \geqslant \frac{\sigma_{n-1}}{\sigma_n}\,\omega(\sigma_n)\sigma_{n+1} \\ &= \frac{\sigma_{n+1}}{\sigma_n}\,\omega(\sigma_n)\sigma_{n-1} \geqslant \sigma_{n-1}(\omega)\sigma_{n+1}\end{aligned}$$

in other words

$$\omega(\sigma_{n+1})\sigma_{n-1} \geqslant \sigma_{n-1}(\omega)\sigma_{n+1}.$$

At the same time

$$\omega(\sigma_{n+1})\omega^{n} \geqslant \omega^{n+1}\,\sigma_{n+1} \tag{3}$$

Upon adding these two inequalities we obtain

$$\omega(\sigma_{n+1})\sigma_n \geqslant \sigma_{n+1}\sigma_n(\omega) \tag{1_{n+1}}$$

This completes the induction.

The preceding discussion seems to indicate that convex rates of convergence form a natural generalization of the classical notion.

It should be noted though that convexity is not a consequence of the conditions in Definition (1.1). In fact, there exist even concave rates of convergence. However, at this early stage of our investigations, we do not know of any problem in analysis which would require in a natural manner the use of a non-convex rate of convergence.

REFERENCES

1. J. KŘÍŽKOVÁ and P. VRBOVÁ, *A remark on a factorization theorem*, Com. Math. Univ. Car. 15, 1974, p. 611-614
2. H. PETZELTOVÁ and P. VRBOVÁ, *A remark on small divisors problems* (in print).
3. V. PTÁK, *Deux théorèmes de factorisation*, Comptes Rendus Ac. Sci. Paris 278, 1974, p. 1091-1094.
4. V. PTÁK, *A theorem of the closed graph type*, Manuscripta math. 13, 1974, p. 109-130.
5. V. PTÁK, *Mondiscrete mathematical induction and iterative existence proofs*, Linear algebra and its applications 13, 1976, p. 223-238.
6. V. PTÁK, *The rate of convergence of Newton's process*, Num. Mathem. 25, 1976, p. 279-285.
7. V. PTÁK, *A rate of convergence*, Abhandlungen aus dem mathematischen Seminar Hamburg (in print).
8. V. PTÁK, *Concerning the rate of convergence of Newton's process*, Comm. Math. Univ. Carolinae 16, 1975, p. 699-705.
9. V. PTÁK, *A modification of Newton's Method*, Čas. pěst. mat. 101, 1976, p. 188-194.
10. J. ZEMÁNEK, *A remark on transitivity of operator algebras*, Čas. pěst. mat. 100, 1975, p. 176-178.