Stanislav Maloň

On nonlinear numerical iteration processes

Commentationes Mathematicae Universitatis Carolinae, Vol. 3 (1962), No. 3, 14--22

Persistent URL: `http://dml.cz/dmlcz/104914`

Commentationes Mathematicae Universitatis Carolinae

3, 3 (1962)

ON NONLINEAR NUMERICAL ITERATION PROCESSES

Stanislav MALOŇ , Praha

I

Let Y be a Banach space and F its closed subset. Let K be a Lipschitz operator mapping F into Y , i.e. there exists a positive constant β such that

(1) $$\| Ku - Kv \| \leq \beta \| u - v \|$$

holds for any $u, v \in F$.

Let us consider the iteration process

(2) $$y_0 \in F, \quad y_{n+1} = Ky_n , \quad n = 0, 1, 2, \dots .$$

This process is convergent if the well known conditions given in the Banach theorem (see e.g.[2]) are satisfied, i.e.

(3a) $$\beta < 1 ,$$

(3b) $$y_0 \in F \Longrightarrow y_1 = Ky_0 \in F ,$$

(3c) $$S(y_1, r) \subset F,$$

where S is a closed sphere whose centre is y_1 and radius

$$r = \frac{\beta}{1-\beta} \| y_1 - y_0 \| .$$

We shall suppose that the sequence (2) converges to the limit y^*.

In practice, if the digital computation technique be used, there often will be necessary to transfer the problem of realising the sequence $\{y_n\}$, defined by (2), into a space different from the original one, so that the elements y_n might be numerically interpreted [1]. To this effect it is necessary to replace the original process (2) by another subsidiary process which is easy to be realised as to using the numerical technique; moreover, we must naturally desire the original process (2) to be approximated sufficiently accurately by the subsidiary process.

In agreement with Kantorovič [1], let us transfer the problem into space $\bar{Y}$ isomorphe with Y, the isomorphism being realised by the linear bounded operation $\wp$; it is natural to assume that the elements $\bar{y} \in \bar{F}$ can be numerically interpreted. The construction of the subsidiary iteration process will be done by a suitable operator $\bar{K}$ approximating the operator K. Let us assign the analogical process

$$\text{(4)} \qquad \begin{aligned} \bar{y}_0 &= \varphi y_0 \\ \bar{y}_{n+1} &= \bar{K}\bar{y}_n, \quad n = 0, 1, 2, \dots \end{aligned}$$

to the process (2).

In this paper we study the approximative solution of the equation $y = Ky$ when the iteration process (2) is replaced by the process (4).

First, we shall deal with the question what sufficient conditions are to be required from the operator $\bar{K}$ to get the limit $\bar{y}^*$ of the process (4) sufficiently near to the limit y^* of the sequence (2) in terms of the definition

$$\text{(5)} \qquad \begin{aligned} &\rho(y^*, \bar{y}^*) < \varepsilon, \qquad \text{where} \\ &\rho(u, \bar{u}) = \|\varphi u - \bar{u}\|_{\bar{Y}}, \quad u \in Y, \ \bar{u} \in \bar{Y} \end{aligned}$$

and ε is a given positive number.

Next, we consider the question of the influence of the rounding-off errors.

II

We assume that the approximating operator $\bar{K}$ is Lipschitz bounded, i.e. for $\bar{u}, \bar{v}$ being two arbitrary elements belonging to $\bar{F}$,

(6) $$\| \bar{K}\bar{u} - \bar{K}\bar{v} \|_{\bar{Y}} \leq \bar{\beta} \| \bar{u} - \bar{v} \|_{\bar{Y}}$$

holds.

The process (4) will converge to a certain limit $\bar{y}^*$, if analogical conditions, the same as for the process (2), i.e.

(7a) $$\bar{\beta} < 1 ,$$

(7b) $$\bar{y}_0 \in \bar{F} \Longrightarrow \bar{y}_1 = \bar{K}\bar{y}_0 \in \bar{F},$$

(7c) $$\bar{S}_1(\bar{y}_1, \bar{r}_1) \subset \bar{F}, \quad \bar{r}_1 = \frac{\bar{\beta}}{1-\bar{\beta}} \| \bar{y}_1 - \bar{y}_0 \|_{\bar{Y}}$$

hold.

What other conditions are to be satisfied by the approximation $\bar{K}$, so that the limits y^* and $\bar{y}^*$ may be sufficiently near in terms of definition (5)? The answer to this question is given by the following theorem:

<u>Theorem 1</u>: Let the following assumptions be fulfilled:

1) The conditions (1), (3a,b,c), (7a,b,c) and (6) are satisfied.

2) Approximating operator $\bar{K}$ is such that for any element $u \in F$ the inequality

(8) $$\| \varphi K u - \bar{K} \varphi u \|_{\bar{Y}} \leq \alpha \| u \|_{Y} \quad \text{x)}$$

x) The condition of L.V.Kantorovič, [1], p.107

holds.

3) $\bar{S}(\varphi y_1, \varkappa) \subset \bar{S}_1(\bar{y}_1, \bar{\varkappa}_1)$

holds, where y_1, $\varkappa$ are defined by the process (2) and by the formula (3c),

4) $\|\varphi\| \leqq 1$.

Then

(9) $$\lim_{n\to\infty} \|\varphi y^* - \bar{y}_n\|_{\bar{Y}} \leqq \frac{\alpha c}{1-\bar{\beta}}$$

holds, where $c = \sup_{0 \leqq i < \infty} \|y_i\|_Y$.

Proof. As both sequences $\{y_n\}$, $\{\bar{y}_n\}$ are convergent according to our assumptions and the operator φ is continuous, the limit in (9) exists; it remains only to prove the inequality.

1) First, we shall prove that for any positive integer n the inequality

(10) $$\|\varphi y_{n+1} - \bar{y}_{n+1}\|_{\bar{Y}} \leqq \frac{\alpha c_n}{1-\bar{\beta}}$$

holds, where $c_n = \max \|y_i\|_Y$.

Evidently, for any $u \in S$

$$\|\varphi u - \varphi y_1\| \leqq \|\varphi\| \, \|u - y_1\| \leqq \|u - y_1\| \leqq \varkappa$$

holds, i.e. $\varphi u \in \bar{S} \subset \bar{S}_1$.

Then it follows from our assumptions that

$$\|\varphi y_1 - \bar{y}_1\|_{\bar{Y}} = \|\varphi K y_0 - \bar{K} \varphi y_0\|_{\bar{Y}} \leqq \alpha \|y_0\|_Y$$

and

$$\|\varphi y_{n+1} - \bar{y}_{n+1}\|_{\bar{Y}} = \|\varphi K y_n - \bar{K} \bar{y}_n\|_{\bar{Y}} \leqq$$

$$\leqq \|\varphi K y_n - \bar{K} \varphi y_n\|_{\bar{Y}} + \|\bar{K} \varphi y_n - \bar{K} \bar{y}_n\|_{\bar{Y}} .$$

As $y_n \in S \subset F$, (8) can be applied on the last by one term further, $\varphi y_n \in \bar{S}$, $\bar{y}_n \in \bar{S}_1$, consequently φy_n and $\bar{y}_n$ belong to $\bar{F}$ and (6) can be applied on the last term; consequently

$$\|\varphi y_{n+1} - \bar{y}_{n+1}\|_{\bar{Y}} \leqq \alpha \|y_n\|_Y + \bar{\beta} \|\varphi y_n - \bar{y}_n\|_{\bar{Y}} \leqq$$

$$\leqq \alpha \|y_n\|_Y + \bar{\beta}\, [\alpha \|y_{n-1}\|_Y + \bar{\beta} \|\varphi y_{n-1} - \bar{y}_{n-1}\|_{\bar{Y}}] \leqq \dots$$

$$\leqq \alpha \sum_{i=0}^{n} \bar{\beta}^i \|y_{n-i}\|_Y < \frac{\alpha\, c_n}{1-\bar{\beta}} \quad .$$

2) It is evident that

$$\|\varphi y^* - \bar{y}_n\|_{\bar{Y}} \leqq \|\varphi y^* - \varphi y_n\|_{\bar{Y}} + \|\varphi y_n - \bar{y}_n\|_{\bar{Y}} \leqq$$

$$\leqq \|y^* - y_n\|_Y + \frac{\alpha\, c_n}{1-\bar{\beta}}$$

holds. As $\|y^* - y_n\| \to 0$ and $c_n \leqq c$, the inequality (9) follows immediately.

<u>Note</u>. If the operator K and its approximation $\bar{K}$ are linear bounded operators mapping complete spaces Y resp. $\bar{Y}$ into themselves, then we can put $F = Y$, $\bar{F} = \bar{Y}$, $\beta = \|K\|$, $\bar{\beta} = \|\bar{K}\|$ and the assumptions (3b),(3c),(7b),(7c) are to be dropped.

III

In practice however, as the actual computation is realised by digital numbers, rounding-off errors in the process (4) arise. Consequently, the computation procedure is not defined by the iteration formula (4) but in general by the following one:

$$\tilde{y}_0 = \varphi y_0 + \bar{\eta}$$
(11)
$$\tilde{y}_{n+1} = \tilde{K} \tilde{y}_n \;, \qquad n = 0, 1, 2, \dots$$

where $\bar{\eta} \in \bar{Y}$ and the operator $\tilde{K}$ is defined on the same subspace $\bar{F}$ as $\bar{K}$ and approximates $\bar{K}$ in terms of definition

(12) $\qquad \| \bar{K}\bar{u} - \tilde{K}\bar{u} \|_{\bar{Y}} \leqq \xi\,, \quad \bar{u} \in \bar{Y}\,,$

where ξ is a nonnegative number (the upper bound of error caused by accumulation of rounding-off errors when computing the value $\bar{K}\bar{u}$).

It is the process (11) only which allways can be realised.

Theorem 2. Let the following conditions be fulfilled: x)

1) The inequality (12) holds for any $\bar{u} \in \bar{Y}$,

2) (7a),(7b),(7c) hold,

3) $\tilde{y}_0 \in \bar{F}$,

4) $\bar{U}\,(\tilde{y}_1\,, \tilde{\kappa} + \gamma) \subset \bar{F}\,,$

where $\bar{U}\,(\tilde{y}_1, \tilde{\kappa} + \gamma)$ is the closed sphere, $\tilde{y}$ its centre, $\tilde{\kappa} + \gamma$ its radius,

$$\gamma = \frac{\bar{\beta}\,\|\bar{\eta}\| + \xi}{1-\bar{\beta}}\,, \quad \tilde{\kappa} = \frac{\bar{\beta}}{1-\bar{\beta}}\left[\|\tilde{y}_1 - \tilde{y}_0\| + \|\bar{\eta}\|\right] + \gamma\,.$$

Then 1) $\tilde{y}_i \in \bar{U}$ for all $i = 1, 2, \ldots$

2) The estimation

(13) $$\|\bar{y}_n - \tilde{y}_n\| \leqq \bar{\beta}^n\,\|\bar{\eta}\| + \frac{\xi}{1-\bar{\beta}}$$

holds.

Proof: 1) We shall prove that the sphere $\bar{S}_1$ defined by the formula (7c) is contained in the sphere

$$\bar{\Sigma}\,(\tilde{y}_1\,, \tilde{\kappa})\,,$$

---

x) In the following we omit to designate spaces when writing norms.

i.e. that $h \in \bar{S}_1$ implies $\| h - \tilde{y}_1 \| \leqq \tilde{\varkappa}$.

Really, for any $\tilde{y}_0 \in \bar{\bar{F}}$

$$\| h - \tilde{y}_1 \| \leqq \| h - \bar{y}_1 \| + \| \bar{y}_1 - y_1 \|$$

$$\| h - \bar{y}_1 \| \leqq \frac{\bar{\beta}}{1-\bar{\beta}} \| \bar{y}_1 - \bar{y}_0 \| \leqq \frac{\bar{\beta}}{1-\bar{\beta}} (\| \bar{y}_1 - \tilde{y}_1 \| + $$

$$+ \| \tilde{y}_1 - \bar{y}_0 \|)$$

$$\| \bar{y}_1 - \tilde{y}_1 \| \leqq \| \bar{K} \bar{y}_0 - \bar{K} \tilde{y}_0 \| + \| \bar{K} \tilde{y}_0 - \tilde{K} \tilde{y}_0 \| \leqq \bar{\beta} \| \bar{y}_0 - \tilde{y}_0 \| + \xi \, .$$

From these inequalities we get simply

$$\| h - \tilde{y}_1 \| \leqq \frac{\bar{\beta} \| \tilde{\eta} \| + \xi}{1 - \bar{\beta}} + \frac{\bar{\beta}}{1-\bar{\beta}} (\| \tilde{y}_1 - \tilde{y}_0 \| + \| \tilde{\eta} \|) = \tilde{\varkappa} \, .$$

2) Evidently γ -neighborhood of $\bar{S}_1$ is contained in γ -neighborhood of $\bar{\bar{\Sigma}}$ and consequently in $\bar{\mathcal{U}}$. We shall prove by induction that if $\tilde{y}_i \in \bar{\mathcal{U}}$, then also $\tilde{y}_{i+1} \in \bar{\mathcal{U}}$. For $i = 1$ it is proved; for $i \geqq 1$ we have

$$\| \bar{y}_i - \tilde{y}_i \| = \| \bar{K} \bar{y}_{i-1} - \tilde{K} \tilde{y}_{i-1} \| \leqq$$

$$\| \bar{K} \bar{y}_{i-1} - \bar{K} \tilde{y}_{i-1} \| + \| \bar{K} \tilde{y}_{i-1} - \tilde{K} \tilde{y}_{i-1} \| \leqq \bar{\beta} \| \bar{y}_{i-1} - \tilde{y}_{i-1} \| + \xi \leqq \cdots$$

$$\leqq \bar{\beta} \{ \bar{\beta} [\cdots (\bar{\beta} \| \bar{y}_1 - \tilde{y}_1 \| + \xi) + \cdots + \xi \} + \xi =$$

$$= \bar{\beta}^i \| \bar{\eta} \| + \xi \cdot (1 + \bar{\beta} \cdots + \bar{\beta}^{i-1}) < \bar{\beta}^i \| \bar{\eta} \| + \frac{\xi}{1-\bar{\beta}} \, ,$$

i.e. the estimation (13) holds. From it follows that

$$\| \bar{y}_i - \tilde{y}_i \| \leqq \frac{\bar{\beta}^i \| \bar{\eta} \| + \xi}{1 - \bar{\beta}} < \gamma \, .$$

As $\bar{y}_i \in \bar{S}_1$, the first part of the assertion is also proved.

<u>Note</u>. The influence of truncation and of rounding errors was studied by M. Urabe. Theorem 2 is slightly generalized result of his paper [3], whose formula (2.5) p. 481 is

a special case of our formula (13) when $\bar{y}_0 = \tilde{y}_0$.

IV

Conclusion. Summing up the results of items 2 and 3, we get the following theorem:

Theorem 3. Let the conditions of the 1st and 2nd theorem be satisfied. Then the following assertions hold:

1) All elements of the sequence (11) belong to $\bar{F}$.

2) For the distance of the n-th approximation $\tilde{y}_n$ from the element φy^* the estimation

$$(14)\quad \|\varphi y^* - \tilde{y}_n\|_{\bar{Y}} < \|y^* - y_n\|_Y + \bar{\beta}^n \|\bar{\eta}\|_{\bar{Y}} + \frac{\alpha c_n + \delta}{1-\bar{\beta}}$$

holds, where c_n is defined in the theorem 1 .

Proof: Evidently

$$\|\varphi y^* - \tilde{y}_n\| \leqq \|\varphi y^* - \varphi y_n\| + \|\varphi y_n - \bar{y}_n\| + \|\bar{y}_n - \tilde{y}_n\| .$$

The first term of the right is at most equal to the error of the n-th approximation in the process (2). For the second term we use the estimation (10) and for the third the estimation (13).

Note. The influence of errors $\|y^* - y_n\|$ and $\|\bar{\eta}\|$ diminishes with $n \to \infty$ to zero. But as δ is a fixed positive number, it does not follow from (14), that the process (11) should converge in current sense. We can assert only, that a number V being given,

$$V > \frac{\alpha c + \delta}{1-\bar{\beta}}, \quad c = \sup_n c_n ,$$

such an integer n_0 exists that $\tilde{y}_n$ belongs to the sphere $\bar{S}_V(\varphi y^*, V)$ for all $n \geq n_0$. In practice, as a rule, $\bar{Y}$ will be an Euclidean m-dimensional space R^m. In every finite part of R^m there is a finite number of vectors whose components are digital numbers with given number of figures; let us assume that the sphere $\bar{S}_V$ contains just $\mathcal{N}$ elements. Then evidently, if the sequence $\{\tilde{y}_n\}$ does not converge, it will be periodic beginning from a certain $n_1 \geq n_0$, with the period $\mathcal{N}$ at most. In other words, the sequence $\{\tilde{y}_n\}$ will reach the state of numerical convergence in the sense of M. Urabe [3]. An arbitrary element $\tilde{y}_{n_1+i}$, $i = 0, 1, \ldots$ can be accepted as an approximation of y^* the error of which does not exceed the number V.

The state of numerical convergence need not take place when especially from the conditions of theorem 2 it is the 2nd condition only which is not fulfilled. However, in the case mentioned in the Note of item II the 2nd condition is to be dropped.

Literature:

[1] L.V. KANTOROVIČ: Funkcional'nyj analiz i prikladnaja matěmatika. UMN, 3, vyp. 6(1948), 89 - 185.

[2] L. COLLATZ: Einige Anwendungen funktionalanalytischer Methoden in der praktischen Analysis, ZAMP, 4 (1953), 327 - 357.

[3] M. URABE: Convergence of numerical iteration in solution of equations. J.of Sci.Hiroshima Univ., Ser. A, 19 No. 3 (1956), 479 - 489.