Proc. Estonian Acad. Sci. Phys. Math., 1991, 40, N 2, 99—104
https://doi.org/10.3176/phys.math.1991.2.05

УДК 517.988; 519.615

*Otu VAARMANN**

A THEOREM FOR METHODS WITH HIGH-ORDER CONVERGENCE

(Presented by N. Alumäe)

A local convergence theorem is proved for some approximate methods having a large order of convergence. From this theorem, as special cases, the convergence for approximate Newton's methods, approximate methods of tangent parabolas and tangent hyperbolas, etc. follow. The approach adopted in this paper is the use of iterative methods to obtain approximations for the exact inverse operators or, for the approximate solution of the corresponding linear equations. Certain conditions are established for which the approximate methods preserve the convergence order intrinsic to these standard iterative methods.

For solving the equation

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

where F is a differentiable (sufficiently often with respect to x) operator from one Banach space E_1, into another E_2, we consider the iterative methods of the type

$$x_{k+1}=x_k-Q(x_k,\ A^i_k),\quad i\in J=[1,\ \ldots,\ r],\ k=0,\ 1,\ \ldots, \tag{2}$$

where $Q(x,\ A^i_k)$ is an operator from E_1 into itself and A^i_k $(i\in J)$ are some approximations to the inverses occurring in the exact method $x_{k+1}=x_k-Q(x_k,\ \cdot)$ having the convergence order $p\geqslant 2$. Let the quantities γ_{ik} characterize the accuracy of approximation and let c_j, h_i and $\omega_{i\upsilon}$ be certain nonnegative constants except for $c_p>0$. In particular, A_k might be an approximation to the inverse $[F'(x)]^{-1}$ with $\|I-F'(x_k)A_k\|\leqslant\gamma_k$, where I denotes the identity mapping. Taking $A_k=[F(y_k;\ x_k)]^{-1}$, where $F(y_k;x_k)$ is a first-order divided difference, and its basic elements x_k and y_k satisfying the condition $\|y_k-x_k\|\leqslant C_1\|F(x_k)\|$ and the symmetric second-order divided difference $F(y_k;\ x_k;\ x_k)$ being bounded, one can take $\gamma_k=C_2\|F(x_k)\|$ $(C_1,\ C_2<\infty)$.

Now assume that the sequence (2) satisfies the condition

$$\|F(x_{k+1})\|\leqslant\sum_{i=1}^{r}\sum_{\upsilon=1}^{n_i}\omega_{i\upsilon}\gamma_{ik}^{a_{i\upsilon}}\|F(x_k)\|^{b_{i\upsilon}}+\sum_{j=p}^{n_0}c_j\|F(x_k)\|^j,$$

$$a_{i\upsilon},\ b_{i\upsilon}\geqslant 0, \tag{A}$$

and the quantities γ_{ik} can be expressed in the form

$$\gamma_{i\,k+1}=\varphi(\gamma_{ik},\ \|F(x_k)\|), \tag{B}$$

where φ is a positive monotonically decreasing scalar function of its arguments, or they vanish in the limit as $k\to\infty$ according to some other

* Eesti Teaduste Akadeemia Küberneetika Instituut (Institute of Cybernetics, Estonian Academy of Sciences). 200 108 Tallinn, Akadeemia tee, 21. Estonia.

rule, or $\gamma_{ik} \leqslant \gamma_i < 1$ $(k=0, 1, \ldots)$, where γ_i can as well amount to zero. If the quantities γ_{ik} are defined by

$$\gamma_{i\,k+1} = \varphi(\gamma_{ik}, \|F(x_k)\|) = (\gamma_{ik} + h_i \|F(x_k)\|)^{q_i} \quad \text{with}$$

$$\gamma_{i0} + h_0 \|F(x_0)\| < 1, \quad h_i > 0 \tag{C}$$

we shall introduce the sequences $\{b_k\}$ and $\{d_k\}$.

Let $s_i = \min\limits_v b_{iv}$ and $l_i = \min\limits_v a_{iv}$ be for all terms in (A) with $\omega_{iv} \neq 0$. We define for $a_{iv} + b_{iv} \geqslant p$ with $\omega = 0$ and $\min\limits_i q_i = q \geqslant p$

$$b_k \geqslant \max \{\max_i \gamma_{ik}' \|F(x_k)\|\} \tag{3}$$

$$d_k = \max \left\{ \max (1+h_i)^q b_k^{q-p}, \sum_{i=1}^{r} \sum_{v=1}^{n_i} \omega_{iv} b_k^{a_{iv}+b_{iv}-p} + \sum_{j=p}^{n_0}{}' c_j b_k^{j-p} \right\} \tag{4}$$

and for $l_i < p - s_i$ and for q_i sufficiently large such that $q_i \geqslant \dfrac{p(p-s_i)}{l_i}$,

$$b_k = \max \{\max \gamma_k^{l_i/(p-s_i)}, \|F(x_k)\|\}, \quad \gamma_k = \max_i \gamma_{ik}, \tag{5}$$

$$d_k = \max \left\{ \max (1+h_i)^{q l_i/(p-s_i)} b_k^{q l_i/(p-s_i)-p}, \right.$$

$$\left. \sum_{i=1}^{r} \sum_{v=1}^{n_i} \omega_{iv} b_k^{(p-s_i)a_{iv} + l_i(b_{iv}-p)/l_i} + \sum^{n} c_j b_k^{i-p} \right\}. \tag{6}$$

Generally speaking, the quantities ω_{iv} and c_j may depend on index k but for the sake of simplicity it is dropped here.

In the case of $\gamma_{ik} = 0$ we denote

$$d_k = c_p + \sum_{j=p+1}^{n} c_i b_k^{j-p}.$$

If $p \geqslant 2$, $b \geqslant 1$ and $0 \leqslant \gamma_{ik} < 1$, we take

$$\delta_m = \sum_{i=1}^{r} \sum_{v=1}^{n_i} \omega_{iv} \gamma_k^{a_{iv}} \|F(x_k)\|^{b_{iv}-1} + \sum_{j=p}^{n_0} c_j \|F(x_k)\|^{j-1}. \tag{7}$$

T h e o r e e m. *Let* $x_0 \in E_1$, $S = \{x \in E_1 : \|x - x_0\| \leqslant \varrho\}$ *and let* $F(x)$ *be a differentiable operator fulfilling the condition* (A) *and let the inequality*

$\|Q(x_k, A^i_k)\| \leqslant \lambda \|F(x_k)\|$, $x_k \in S$, $k = 1, 2, \ldots$, $i \in J$, $0 < \lambda < \infty$, be *be valid in S.*

Then 1) *If* $p \geqslant 2$, $\min\limits_{i,v} b_{iv} = b = 1$, $\gamma_k \leqslant \gamma \neq 0$, $r_1 = \lambda \|F(x_0)\| / (1-\delta) \leqslant \varrho$ $\delta = \delta_0^{(1)} < 1$ $(\gamma_0 = \gamma)$ *then the equation* $F(x) = 0$ *has a solution* x^* *in S,* $\|x^* - x_0\| \leqslant r_1$, *to which the sequence* (2) *converges with*

$$\|x_k - x^*\| \leqslant r_1 \delta^k;$$

if $p \geqslant 2$ *provided* $\gamma_{i0} \geqslant \gamma_{i1} \geqslant \ldots \geqslant \gamma_{ik} \geqslant \ldots \geqslant 0$ *and* $\gamma_{ik} \to 0$ *as* $k \to \infty$, *then* $\delta_k^{(1)} \searrow 0$ *and the sequence* (2) *converges at least superlinearly, i. e.*

$$\|x_k - x^*\| \leqslant r_1 \prod_{m=0}^{k-1} \delta_m^{(1)}\ *;$$

---

* If $k=0$ then $\prod\limits_{i=0}^{k-1} \delta_i$ is referred to as 1.

2) *if* $p \geqslant 2$ *and the quantities* γ_{ik} *are expressed in the form* (*B*) *or* $\gamma_{ik}=0$, $i \in J$, $k=0, 1, \ldots$, $\delta = \sqrt[p-1]{d_0} b_0 < 1$, $r_p = \lambda H_0(\delta) / \sqrt[p-1]{d} \leqslant \varrho$, *where* $\lim_{k\to\infty} d = d_k > 0$,

$$H_k(\delta) = \sum_{m=k}^{\infty} \delta^{p^m},$$

then the equation $F(x)=0$ *has a solution* x^* *in* S, $\|x^* - x_0\| \leqslant r_p$, *to which the sequence* (2) *converges and*

$$\|x_k - x^*\| \leqslant (\lambda / \sqrt[p-1]{d})\, H_k(\delta).$$

Proof. 1) Taking into account relations $\|x_{k+1} - x_k\| \leqslant \lambda \|F(x_k)\|$, $\delta = \delta_0^{(1)} < 1$ and (A), we obtain

$$\|F(x_{k+1}) \leqslant \delta_{k-1}^{(1)} \|F(x_{k-1})\| \leqslant | P_0 F(x_0) \| \prod_{i=1}^{k-1} \delta_i^{(1)} \leqslant \| P_0 F(x_0)\| \delta^k,$$

$$\|x_{k+1} - x_k\| \leqslant \lambda \|F(x_0)\| \prod_{i=1}^{k-1} \delta_i^{(1)} \leqslant \lambda \|F(x_0)\| \delta^k,$$

$$\|x_m - x_k\| \leqslant \sum_{i=k}^{m-1} \|x_{i+1} - x_i\| \leqslant r_1(\delta^k - \delta^m) \qquad (m \geqslant k).$$

This means that the sequence (2) is fundamental and, consequently,

$$x^* = \lim x_k, \;\; F(x^*) = 0, \;\; \|x_k - x^*\| \leqslant r_1 \prod_{i=0}^{k-1} \delta_i^{(1)} \leqslant r_i \delta^k,$$

$$\|x_k - x_0\| \leqslant r_1 (1 - \delta^k) \leqslant r_1 \leqslant \varrho, \quad \text{i.e. all} \quad x_k \in S.$$

In the case $\gamma_k \to 0$ as $k \to \infty$, obviously, $\delta_k^{(1)} \to 0$ as well and therefore the sequence (2) converges at least ($\min_{i,\upsilon} b_{i\upsilon} = b = 1$) superlinearly to a solution of (1).

2) If relations (C) hold, and $\gamma_{ik} = 0(\|F(x_k)\|)$, or $\gamma_{ik} = 0$, then, on the basis of (3)—(7), one concludes that in the capacity of b_{k+1} one can take $d_k b_k^p$, where

$$b_{k+1} = d_k b_k^p \leqslant d^{-\frac{1}{p-1}} \left(d_k^{\frac{1}{p-1}} b_k \right)^p \leqslant \ldots \leqslant d^{-\frac{1}{p-1}} \delta^{k+1}. \tag{8}$$

We shall show the correctness of this assertion for $l_i < p - s_i$. Owing to $\gamma_{ik} \leqslant b_k^{(p-s_i)/l_i}$ it follows from (A) that

$$\|F(x_{k+1})\| \leqslant \left(\sum_{i=1}^{r} \sum_{\upsilon=1}^{n_t} \omega_{i\upsilon} b_k^{((p-s_i)a_{i\upsilon} + l_i(b_{i\upsilon} - p))/l_i} + \sum_{j=p}^{n_0} c_j b_k^{j-p} \right) b_k^p \leqslant d_k b_k^p \tag{9}$$

while $(p - s_i) \dfrac{a_{i\upsilon}}{l_i} + (b_{i\upsilon} - p) \geqslant (p - s_i) + (s_i - p) = 0$ by reason of $a_{i\upsilon} \geqslant l_i$ and $b_{i\upsilon} \geqslant s_i$.

Analogously we have

$$\gamma_{ik+1}^{l_i/(p-s_i)} = (\gamma_{ik} + h_i \|F(x_k)\|)^{ql_i/(p-s_i)} \leqslant (b_k^{(p-s_i)/l_i} + h_i b_k)^{ql/(p-s_i)} =$$

$$= (1 + h_i)^{ql_i/(p-s_i)} b_k^{ql_i/(p-s_i) - p} b_k^p \leqslant d_k b_k^p. \tag{10}$$

From $\delta=\delta_0^{(p)}<1$ it follows that $b_1=d_0b_0^p=d_0b_0^{p-1}b_0=\delta^{p-1}b_0\leqslant\delta b_0\leqslant b_0$ for $p\geqslant 2$ and then also $d_1\leqslant d_0$. Further we have by induction $b_0\geqslant b_1\geqslant\ldots\ldots\geqslant b_k\geqslant\ldots\geqslant 0$ and $d_0\geqslant d_1\geqslant\ldots\geqslant d_k$. On the other hand the sequence $\{d_k\}$ is bounded from below.

Indeed, because of the assumption $c_p>0$ we have $\sum_{j=p}^{n_0} c_j b_k^{j-p}\geqslant c_p>0$.

Therefore there exists a limit $d=\lim_{k\to\infty} d_k$. Combining (9) with (10) yields $b_{k+1}=d_k b_k^2\leqslant d^{-\frac{1}{p-1}}\left(d_k^{\frac{1}{p-1}}b_k\right)^p\leqslant\ldots\leqslant d^{-\frac{1}{p-1}}\delta^{p^{k+1}}$. In a similar way we may be convinced that (8) holds for $l_i\geqslant p-s_i$.

Thus

$$\|x_{k+1}-x_k\|\leqslant\lambda d^{-\frac{1}{p-1}}\delta^{p^k},\quad \|x_n-x_k\|\leqslant\lambda d^{-\frac{1}{p-1}}[H_k(\delta)-H_n(\delta)]\,(n\geqslant k),$$

i. e. the sequence $\{x_k\}$ is fundamental and therefore it has a limit point $x^*=\lim_{k\to\infty} x_k$. Now it should not be hard to prove that

$$F(x^*)=0,\quad \|x_k-x^*\|\leqslant\lambda d^{-\frac{1}{p-1}}H_k(\delta),\ \|x_0-x^*\|\leqslant\lambda d^{-\frac{1}{p-1}}H_0(\delta).$$

Corollary. *Let* τ *be a real number,* $2\leqslant\tau\leqslant p-1$. *If* a) $q_1\geqslant\tau$ *and* $a_{i\upsilon}+b_{i\upsilon}\geqslant\tau$, $i\in I$, $\upsilon=1,\ \ldots,\ n_i$ or b) $l_i<\tau-s_i$ *and* $q_i\geqslant\dfrac{\tau(\tau-s_i)}{l_i}$, *then the order of the convergence the of iteration process* (2) *is at least equal to* τ.

Remark. Let the approximations to $[F'(x_k)]^{-1}$ be defined by

$$A_{k+1}=\sum_{i=1}^{q-1}(I-F'(x_{k+1})A_k)^i. \tag{11}$$

Then

$$\|I-F'(x_{k+1})A_k\|\leqslant\|I-F'(x_k)A_k\|+\|F'(x_{k+1})-F'(x_k)\|\leqslant$$
$$\leqslant\gamma_k+h\|F(x_k)\|,$$

where $h=L_1\lambda$ and $\|F''(x)\|\leqslant L_1$. Accordingly,

$$\|I-F'(x_{k+1})A_{k+1}\|\leqslant\|I-F'(x_{k+1})A_k\|^q\leqslant(\gamma_k+h\|F(x_k)\|)^q=\gamma_{k+1}.$$

Let us demonstrate how this Theorem can be applied to some modifications of tangent parabolas and hyperbolas.

Consider a modification of tangent parabolas

$$x_{k+1}=x_k-A_kF(x_k)-\frac{1}{2}A_kF''(x_k)(A_kF(x_k))^2, \tag{12}$$

for which we have

$$F(x_{k+1})=F(x_k)+F'(x_k)(x_{k+1}-x_k)+R=$$
$$=(I-F'(x_k)A_k)F(x_k)-\frac{1}{2}(I-F'(x_k)A_k)F''(x_k)(A_kF(x_k))^2+R,$$

where $R=\int_0^1[F''(x_k+t(x_{k+1}-x_k))-F''(x_k)](x_{k+1}-x_k)^2(1-t)\,dt$.

Further,

$$\|F(x_{k+1})\|\leqslant\gamma\|F(x_k)\|+\frac{1}{2}L_1\lambda^2\|F(x_k)\|^2+\frac{1}{3}L_2\lambda^3\|F(x_k)\|^3,$$

where $\|F''(x_k)\|\leqslant L_2$.

Here $p=3$, $s=\min_{\upsilon} b\upsilon=1$, and $l=\min_{\upsilon} a\upsilon=1$. In order to quarantee the cubic convergence rate for (12) with A_k defined by (11) the inequality $q \geqslant \frac{p(p-s)}{l}=6$ must be valid.

Using, in (12), the following approximation (discretization)

$$F''(x_k)(A_kF(x_k))^2=F''(x_k)[x_k-A_kF(x_k))-x_k]^2\approx$$
$$\approx 2[F(x_k-A_kF(x_k))-F(x_k)-A_k^{-1}((x_k-A_kF(x_k))-x_k)],$$

one gets

$$x_{k+1}=x_k-A_kF(x_k)-A_kF(y_k), \quad y_k=x_k-A_kF(x_k). \tag{13}$$

In a similar way as before one obtains (cf. [1])

$$\|F(x_{k+1})\|\leqslant\gamma_k^2\|F(x_k)\|+\omega_2\gamma_k\|F(x_k)\|^2+c_3\|F(x_k)\|^3.$$

In this case $a_\upsilon+b_\upsilon=p=3$ ($\upsilon=1$, 2) and therefore $q\geqslant 3$ provides cubic convergences for (11)—(12).

Let us consider now an inexact method of tangent hyperbolas

$$x_{k+1}=x_k-V_kA_kF(x_k), \tag{14}$$

where $V_k\approx\left[A_kF(x_k)-\frac{1}{2}A_kF''(x_k)A_kF(x_k)\right]^{-1}$ and let V_k and A_k satisfy the following conditions $\|I-U_kV_k\|\leqslant\gamma_{1k}$ and $\|I-F'(x_k)A_k\leqslant\leqslant\gamma_{2k}$. Then (cf. [2])

$$\|F(x_{k+1})\|\leqslant\omega_{11}\gamma_{1k}\|F(x_k)\|+\omega_{12}\gamma_{1k}\|F(x_k)\|^2+\omega_{21}\gamma_{2k}\|F(x_k)\|^2+$$
$$+c_3\|F(x_k)\|^3, \tag{15}$$

and, to assure the cubic convergence rate for (13), it is necessary that $q_1\geqslant 6$ and $q_2\geqslant 3$ provided V_k is defined by

$$V_{k+1}=\sum_{i=1}^{q_1-1}(I-U_{k+1}V_k)^i$$

and A_k by (11) with $1=0, 1, \ldots, q_2$, respectively.

Remark 2. In the case of $0<\gamma_{ik}\leqslant\gamma_i<1$ or $\gamma_{ik}=0(\|F(x_k)\|$ the quantities γ_{ik} or their expressions may be included into the expression of $\omega_{i\upsilon}$ and the convergence rate is then determined by the quantities $b=\min_{i,\upsilon} b_{i\upsilon}$ and p (see also Corollary). If $A_k=[F(2x_k-x_{k-1};\ x_{k-1})]^{-1}$ or $A_k=[F(2y_{k-1}-x_{k-1};\ x_{k-1})]^{-1}$ then $\gamma_k=0(\|F(x_{k-1})\|^2$ and $\gamma_k==0(\|F(x_{k-1})\|\,\|F(x_{k-2})\|\ldots\|F(x_0)\|)$, respectively [1]. Thus the Theorem is not directly applicable. If expressions of this type are valid for γ_k then they may also be taken into account by determining the quantities $\omega_{i\upsilon}$ but in this case the asymptotic convergence rate according to [3] is attainable for the iterative process (2).

Remark 3. Making use of approximations instead of the exact inverses is, theoretically, equivalent to solving the corresponding linear equations at each step within a certain tolerance, say $\eta\|F(x_k)\|$ ($\eta_k>0$). Therefore the Theorem under consideration is applicable also in the case when some iterative methods are exploited for obtaining solutions of the subproblems. For example, on the basis of the Theorem it is evident that in order to preserve the cubic convergence rate for the inexact method of tangent parabolas one must have $\eta_k=0(F\|(x_k)\|^2)$. The sequence $\{\eta_k\}$ is known as the forcing sequence.

REFERENCES

1. *Ваарманн О., Полль В.* 1977. Изв. АН ЭССР. Физ. Матем., **26**, № 2, 123—127.
2. *Ваарманн О.* 1987. Изв. АН ЭССР. Физ. Матем., **36**, № 3, 269—276.
3. *Schmidt, J. W.* 1963. Z. angew. Math. and Mech., **43**, H 3, 5, 97—110.

Received
Nov. 30, 1990

Otu VAARMANN

TEOREEM KÕRGE KOONDUVUSJÄRGUGA ITERATSIOONIMEETODITE KOHTA

Mittelineaarse operaatorvõrrandi $F(x)=0$ lahendamiseks Banachi ruumis on vaadeldud kiirelt koonduvaid iteratsioonimeetodeid, mille koonduvusjärk $p\geqslant 2$. On tõestatud koonduvusteoreem mitmel eri viisil tuletatavate iteratsioonimeetodite kohta ning antud tingimused, mille täidetuse korral vaadeldavad modifikatsioonid säilitavad neile põhimeetoditele omase koonduvuskiirguse järgu. Teoreemis toodud tulemuste illustreerimiseks on käsitletud lähemalt puutuvate paraboolide ja hüperboolide modifikatsioone, kus operaatorite $F''(x)$, $Q^{-1}(x)$ ja $[F'(x)]^{-1}$ asemel on kasutatud nende aproksimatsioone. Vastavad lineaarvõrrandid on igal iteratsioonisammul lahendatud mõne iteratsioonimeetodi abil ligikaudu.

Оту ВААРМАНН

ОДНА ТЕОРЕМА ДЛЯ МЕТОДОВ ВЫСОКОГО ПОРЯДКА СКОРОСТИ СХОДИМОСТИ

Цель настоящей работы — исследование общих вычислительных схем итерационных методов высокого порядка скорости сходимости $p\geqslant 2$, являющихся, в частности, аппроксимационными аналогами методов касательных парабол и гипербол, в которых вместо производных и обратных операторов могут быть использованы их приближения или выражения, содержащие приближенные решения соответствующих уравнений. Для получения их с заданной наперед точностью применяются некоторые итерационные процедуры. Определены условия, при выполнении которых эти модификации сохраняют присущие основным методам порядки скорости сходимости.