数理解析研究所講究録
第 38 巻 1968 年 51-66



# The Newton Method and its Application to Boundary Value Problem with Nonlinear Boundary Conditions

Minoru URABE*

0. Introduction. The present note is concerned with the boundary value problem of the form

$$(0.1) \qquad \frac{dx}{dt} = X(x,t), \quad f(x) = 0,$$

where $x$ and $X(x,t)$ are n-dimensional vectors and $f(x)$ is an n-dimensional vector-valued functional. First the Newton method will be established for equations in linear normed spaces and then the results will be applied to the equation

$$(0.2) \qquad F(x) = [\frac{dx}{dt} - X(x,t),\ f(x)] = 0.$$

The theorem obtained for boundary value problem (0.1) is a generalization of the theorem obtained by the author in his previous paper [2]. Lastly the theorem newly obtained will be applied to the perturbation method and the author will give an explicit bound of the small parameter within which the perturbation method is really effective.

1. The Newton method for functional equations. We consider a linear normed space $E$, a Banach space $B$, and a function $F(x)$ mapping an open set $D$ of $E$ into $B$. For $F(x)$, we assume that for any $x \in D$, there is an additive homogeneous operator $J(x)$ mapping $E$ into $B$ such that

$$||F(x + h) - F(x) - J(x)h|| \,/\, ||h|| \to 0 \quad \text{as} \quad ||h|| \to 0.$$

* Research Institute for Mathematical Sciences, Kyoto Univ.

In the sequel, such $J(x)$ will be called the weak Fréchet derivative of $F(x)$.

Now we suppose the weak Fréchet derivative $J(x)$ of $F(x)$ is continuous on $D$. Then applying the Newton method to the equation

$$(1.1) \qquad F(x) = 0,$$

we get the following theorem.

Theorem 1. Suppose the equation (1.1) has an approximate solution $x = \bar{x} \in D$ for which there are an additive operator $L$ mapping $E$ into $B$, a positive number $\delta$, and a non-negative number $\varkappa < 1$ such that

(1.2) $L$ has a linear inverse operator $L^{-1}$,

$$(1.3) \qquad D_\delta = \{x \mid \|x - \bar{x}\| \leq \delta,\ x \in E\} \subset D,$$

$$(1.4) \qquad \|J(x) - L\| \leq \varkappa/M \quad \text{on } D_\delta,$$

$$(1.5) \qquad Mr/(1 - \varkappa) \leq \delta .$$

Here $r(\geq 0)$ and $M(> 0)$ are the numbers such that

$$(1.6) \qquad \|F(\bar{x})\| \leq r,$$

$$(1.7) \qquad \|L^{-1}\| \leq M.$$

Then the Newton iterative process

$$(1.8) \qquad x_{n+1} = x_n - L^{-1}F(x_n) \qquad (n = 0,1,2,\ldots)$$

with $x_0 = \bar{x}$ yields a fundamental sequence $\{x_n\}$ $(n = 0,1,2,\ldots)$ in $D_\delta$ and we have

(1.9) $||x_n - \bar{x}|| \leq Mr/(1 - \kappa)$ $(n = 0, 1, 2, \ldots)$.

If the above fundamental sequence $\{x_n\}$ $(n = 0, 1, 2, \ldots)$ is convergent in E, namely, there is an $\hat{x} \in E$ such that

(1.10) $||x_n - \hat{x}|| \to 0$ as $n \to \infty$,

then $\hat{x}$ is a unique solution of (1.1) in $D_\delta$ and we have

(1.11) $||\hat{x} - \bar{x}|| \leq Mr/(1 - \kappa)$.

In this theorem, evidently the latter conclusion holds if E is complete, namely, E is a Banach space.

The proof of the theorem can be carried out similarly to that of the theorem for the equation in Euclidean space (see Urabe [1]).

Remark 1. By the condition (1.4), L is an additive operator close to $J(\bar{x})$. Hence, in practical applications, it will be usually convenient to take $J(\bar{x})$ for L.

Remark 2. Conditions (1.3), (1.4) and (1.5) are related with the accuracy of the approximate solution $x = \bar{x}$. In fact, suppose $L = J(\bar{x})$ and $\bar{x}$ is accurate. Then, as seen from (1.6), we can take small r and hence we can take small $\delta$ satisfying (1.3) and (1.5) for any fixed $\kappa$. Since $J(x)$ is continuous in x, (1.4) will be then satisfied automatically for such small $\delta$. In such a sense, conditions (1.3), (1.4) and (1.5) are related with the accuracy of the approximate solution and it indeed provides the accuracy of the approximate solution to which Theorem 1 can be applied.

2. Lemma on mixed linear differential operators. In what follows we consider an n-dimensional vector space $R^n$ with any norm $||\ldots||$ and the space $C[I]$ consisting of n-dimensional vector-functions continuous on the interval $I = [a,b]$. For any $x(t) \in C[I]$, we define its norm by $\sup_{t \in I} ||x(t)||$ and denote it by $||x(t)||_C$. In what follows, we consider also the space $B = C[I] \times R^n$ and, for any $y = [x(t), v] \in B$, we define its norm $||y||_B$ by

$$(2.1) \qquad ||y||_B = ||x(t)||_C + ||v||.$$

The space B is evidently a Banach space with respect to this norm.

Now we consider the mixed linear differential operator L of the following form

$$(2.2) \qquad Lx = [\frac{dx}{dt} - A(t)x, \ \ell x],$$

where $A(t)$ is an $n \times n$ matrix continuous on I and $\ell$ is a linear operator mapping $C[I]$ into $R^n$. Evidently L mapps $E = C'[I] \subset C[I]$ into B where $C'[I]$ is the set of n-dimensional vector-functions continuously differentiable on I.

The operator L is evidently additive, but it is not bounded on E with respect to the norms $||\ldots||_C$ and $||\ldots||_B$.

Now corresponding to matrix $A(t)$, we consider an arbitrary fundamental matrix $\Phi(t)$ of the linear homogeneous differential system

$$(2.3) \qquad \frac{dz}{dt} = A(t)z$$

and, by $\ell[\Phi(t)]$, we denote the matrix whose column vectors are $\ell[\Phi_i(t)]$ $(i = 1,2,\ldots,n)$ where $\Phi_i(t)$ are the column vectors of the matrix $\Phi(t)$.

Then for the inverse operator of L, we can prove the following lemma.

Lemma. *If the matrix* $G = \ell[\Phi(t)]$ *is non-singular, namely,*

$$(2.4) \qquad \det G = \det \ell[\Phi(t)] \neq 0,$$

*then the operator* L *defined by* (2.2) *has a linear inverse operator* $L^{-1}$ *and, for* $||L^{-1}||_C$, *we have*

$$(2.5) \qquad ||L^{-1}||_C \leqq \max(\,||H_1||_C,\ ||H_2||_C\,).$$

*Here* $H_1$ *is a linear operator mapping* $C[I]$ *into* $E = C'[I] \subset C[I]$ *such that*

$$(2.6) \quad H_1\varphi = \Phi(t)\int_a^t \Phi^{-1}(s)\varphi(s)ds - \Phi(t)G^{-1}\ell[\Phi(t)\int_a^t \Phi^{-1}(s)\varphi(s)ds]$$

*and* $H_2$ *is a linear operator mapping* $R^n$ *into* E *such that*

$$(2.7) \quad H_2 v = \Phi(t)G^{-1}v.$$

Proof. Put

$$Lx = y = (\varphi, v),$$

then, by (2.2), we have

$$(2.8) \qquad \frac{dx}{dt} - A(t)x = \varphi(t),$$

$$(2.9) \qquad \ell x = v.$$

The general solution of (2.8) is

$$(2.10) \qquad x(t) = \Phi(t)c + \Phi(t)\int_a^t \Phi^{-1}(s)\,\varphi(s)\,ds,$$

where $c$ is an arbitrary constant vector. To determine $c$ so that (2.10) may satisfy (2.9), we substitute (2.10) into (2.9). Then we have

$$(2.11) \qquad \ell[\Phi(t)c] + \ell[\Phi(t)\int_a^t \Phi^{-1}(s)\varphi(s)ds] = v.$$

By (2.4), we then have

$$c = G^{-1}v - G^{-1}\ell[\Phi(t)\int_a^t \Phi^{-1}(s)\varphi(s)\,ds].$$

If we substitute this into (2.10), then we have

$$x(t) = \Phi(t)G^{-1}v - \Phi(t)G^{-1}\ell[\Phi(t)\int_a^t \Phi^{-1}(s)\varphi(s)ds] + \Phi(t)\int_a^t \Phi^{-1}(s)\varphi(s)d$$

By (2.6) and (2.7), this shows that

$$(2.12) \qquad x = L^{-1}y = H_1\varphi + H_2v,$$

which evidently implies the existence of the linear inverse operator $L^{-1}$.

Now, from (2.12), for any $y = (\varphi, v)$, we have

$$\begin{aligned} ||L^{-1}y||_C &\leqq ||H_1||_C \cdot ||\varphi||_C + ||H_2||_C \cdot ||v|| \\ &\leqq \max(||H_1||_C, ||H_2||_C) \cdot (||\varphi||_C + ||v||) \\ &= \max(||H_1||_C, ||H_2||_C) \cdot ||y||_B. \end{aligned}$$

This clearly implies (2.5) and hence this completes the proof.

Q. E. D.

Remark 1. As easily seen, condition (2.4), linear operators $H_1$ and $H_2$ are all independent of choice of the fundamental matrix $\Phi(t)$ of (2.3).

Remark 2. As seen from (2.8) and (2.9), the lemma gives solutions of general linear boundary value problems for linear differential equations.

3. An existence theorem of nonlinear boundary value problems. In the present section, we consider the nonlinear boundary value problem of the following form:

(3.1) $$\frac{dx}{dt} = X(x,t), \quad f(x) = 0,$$

where $x$ and $X(x,t)$ are n-dimensional vectors and $f$ is an operator mapping some set of $C[I[a,b]]$ into $R^n$.

Let $\Omega$ be the domain of the tx-space intercepted by two hyperplanes $t = a$ and $t = b$ (the boundary points of $\Omega$ on the hyperplanes $t = a$ and $t = b$ are supposed to be included in $\Omega$ and to make an open set on each hyperplane). Put

$$D = \left\{ x(t) \mid (t, x(t)) \in \Omega \quad \text{for} \quad t \in I, \; x(t) \in E = C'[I] \right\},$$

$$D' = \left\{ x(t) \mid (t, x(t)) \in \Omega \quad \text{for} \quad t \in I, \; x(t) \in C[I] \right\}.$$

Then it is evident that $D \subset D'$ and $D$ and $D'$ are open respectively in $E$ and $C[I]$.

In (3.1), we assume that $X(x,t)$ is defined and continuously differentiable with respect to $x$ on $\Omega$ and $f(x)$ is defined and continuously Fréchet differentiable on $D'$. By $X_x(x,t)$ and $f_x(x)$, let us denote respectively the Jacobian matrix of $X(x,t)$ with respect to $x$ and the Fréchet derivative of $f(x)$.

Now we consider an additive operator $L$ mapping $E$ into $B = C[I] \times R^n$ of the following form:

$$(3.2)\qquad Lh = \left[\frac{dh}{dt} - A(t)h,\ \ell h\right],$$

where $A(t)$ is an $n \times n$ matrix continuous on $I$ and $\ell$ is a linear operator mapping $C[I]$ into $R^n$. By $\Phi(t)$, let us denote an arbitrary fundamental matrix of the linear homogeneous system

$$(3.3)\qquad \frac{dz}{dt} = A(t)z.$$

Then, by the lemma of the preceding section, the operator $L$ has a linear inverse operator $L^{-1}$ if

$$(3.4)\qquad \det \ell[\Phi(t)] \neq 0.$$

Now let us apply Theorem 1 to our boundary value problem (3.1). Then we have the following existence theorem.

Theorem 2. Suppose the boundary value problem (3.1) has an approximate solution $x = \bar{x}(t) \in D$ for which there are an additive operator $L$ of the form (3.2) mapping $E$ into $B$, a positive number $\delta$, and a non-negative number $\kappa < 1$ such that

$$(3.5)\qquad \det \ell[\Phi(t)] \neq 0 \ \text{ for } L,$$

$$(3.6)\qquad D'_\delta = \left\{x \mid \|x - \bar{x}\|_C \leq \delta,\ x \in C[I]\right\} \subset D',$$

$$(3.7)\qquad \|X_x(x,t) - A(t)\|_C + \|f_x(x) - \ell\| \leq \kappa/M \ \text{ on } D'_\delta,$$

$$(3.8)\qquad Mr/(1-\kappa) \leq \delta.$$

Here $r(\geq 0)$ and $M(> 0)$ are the numbers such that

$$(3.9)\qquad \left\|\frac{d\bar{x}}{dt} - X(\bar{x},t)\right\|_C + \|f(\bar{x})\| \leq r,$$

(3.10) $\quad ||L^{-1}||_C \leq M.$

Then the boundary value problem (3.1) has one and only one solution $x = \hat{x}(t)$ in

(3.11) $\quad D_\delta = \{x \mid ||x - \bar{x}||_C \leq \delta,\ x \in E\},$

and, for this solution, we have

(3.12) $\quad ||\hat{x} - \bar{x}||_C \leq Mr/(1 - \kappa).$

By the lemma of section 2, condition (3.5) implies the existence of a linear inverse operator $L^{-1}$ and hence it implies the existence of a finite number $M$ satisfying (3.10).

Proof. From (3.6) and (3.11), we have

(3.13) $\quad D_\delta \subset D' \cap E = D.$

Now put

(3.14) $\quad F(x) = [\frac{dx}{dt} - X(x,t),\ f(x)],$

then evidently $F(x)$ mapps $D \subset E$ into $B$. It is easily seen that such $F(x)$ has a weak Fréchet derivative $J(x)$ of the following form:

(3.15) $\quad J(x)h = [\frac{dh}{dt} - X_x(x,t)h,\ f_x(x)h],$

where $h$ is an arbitrary element belonging to $E$. From (3.15), it is clear that $J(x)$ is continuous on $D$. Now let us compare (3.15) with (3.2). Then, by (3.7), we readily see that

(3.16) $\quad ||J(x) - L||_B \leq \varkappa / M$ on $D_\delta \subset D'$.

Now, for the approximate solution $x = \bar{x} \in D$, from (3.9), we have

(3.17) $\quad ||F(\bar{x})||_B = ||\frac{d\bar{x}}{dt} - X(\bar{x}, t)||_C + ||f(\bar{x})|| \leq r.$

Thus we see that, for the equation

(3.18) $\quad F(x) = 0$

and given $x = \bar{x}$, the conditions of Theorem 1 are all fulfilled.

Then by Theorem 1, we have a fundamental sequence $\{x_n\}$ (n = 0, 1,2,...) in $D_\delta$ produced by the Newton iterative process

(3.19) $\quad x_{n+1} = x_n - L^{-1}F(x_n) \qquad (n = 0,1,2,\ldots;\ x_o = \bar{x}).$

However, by (3.2) and (3.14), the above iterative process can be written as follows:

$$Lx_{n+1} = Lx_n - F(x_n)$$
$$= [X(x_n, t) - A(t)x_n,\ \ell x_n - f(x_n)].$$

Hence we have

(3.20) $\quad x_{n+1} = L^{-1}[X(x_n, t) - A(t)x_n,\ \ell x_n - f(x_n)] \quad (n = 0,1,2,\ldots).$

Now $\{x_n\}$ (n = 0,1,2,...) is a fundamental sequence in $E = C'[I] \subset C[I]$ with respect to the norm $||\ldots||_C$. Hence, by the completeness of the space $C[I]$, there exists a vector-function $\hat{x} \in C[I]$ such that

(3.21) $\quad ||x_n - \hat{x}||_C \to 0 \quad$ as $n \to \infty$.

Then, since $x_n \in D_\delta$, it is evident that

(3.22) $$||\hat{x} - \bar{x}||_C \leqq \delta .$$

Then by (3.6), $\hat{x} \in D'_\delta \subset D'$. Hence we have

$$X(\hat{x},t) - A(t)\hat{x} \in C[I] \quad \text{and} \quad \ell\hat{x} - f(\hat{x}) \in R^n.$$

Then letting $n \to \infty$ in (3.20), we have

(3.23) $$\hat{x} = L^{-1}[X(\hat{x},t) - A(t)\hat{x}, \ \ell\hat{x} - f(\hat{x})].$$

Since $L^{-1}$ is a linear operator mapping $B = C[I] \times R^n$ into $E = C'[I]$, the equality (3.23) implies

(3.24) $$\hat{x} \in E.$$

Then by Theorem 1, we see that $x = \hat{x}$ is a unique solution of (3.18) in $D_\delta$, namely, the unique solution of the given boundary value problem (3.1) in $D_\delta$, and further that

$$||\hat{x} - \bar{x}||_C \leqq Mr/(1 - \kappa).$$

This completes the proof. Q. E. D.

Remark 1. For the solution $x = \hat{x}(t)$ obtained in Theorem 2, we can easily prove that

(3.25) $$||\frac{d\hat{x}}{dt} - \frac{d\bar{x}}{dt}||_C \leqq (1 + \frac{KM}{1 - \kappa})r,$$

where $K$ is a non-negative number such that

(3.26) $$||X_x(x,t)||_C \leqq K \quad \text{on} \quad D'_\delta .$$

Remark 2. As in Theorem 1, in practical applications, it will be usually convenient to take $J(\tilde{x})$ for $L$.

Remark 3. Suppose the boundary condition is linear, namely, $f(x)$ is of the form

$$f(x) = \ell_o x - \gamma \quad ,$$

where $\ell_o$ is a linear operator mapping $C[I]$ into $R^n$ and $\gamma$ is a constant vector belonging to $R^n$. In such a case, evidently we can take $\ell_o$ for $\ell$ of $L$. Then Theorem 2 turns to the theorem that is a slight generalization of the one obtained by the author in his previous paper [2].

Remark 4. For the boundary value problem (3.1), we shall call the solution $x = \hat{x}(t) \in D$ the isolated solution if

(3.27) $\quad \det f_x(\hat{x})[\Phi(t)] \neq 0,$

where $\Phi(t)$ is an arbitrary fundamental matrix of the linear homogeneous system

$$\frac{dz}{dt} = X_x(\hat{x}, t)z.$$

The terminology comes from the fact that there is no other solution of the boundary value problem (3.1) in a sufficiently small neighborhood of the solution $x = \hat{x}(t)$ satisfying (3.27).

It can be easily proved that the solution $x = \hat{x}(t)$ obtained in Theorem 2 is an isolated solution.

4. Application to the perturbation method. Let the perturbed system of (3.1) be

(4.1) $\quad \dfrac{dx}{dt} = X(x,t) + \lambda R(x,t,\lambda), \quad f(x) + \lambda d(x,\lambda) = 0,$

where $\lambda$ is a small parameter such that

(4.2) $\quad \lambda \in \Lambda = \{\lambda \mid |\lambda| \leq \rho\} \qquad (\rho > 0).$

In (4.1), we suppose $R(x,t,\lambda)$ is continuously differentiable with respect to $x$ on $\Omega \times \Lambda$ and $d(x,\lambda)$ is continuously Frêchet differentiable with respect to $x$ on $D' \times \Lambda$. By $R_x(x,t,\lambda)$ and $d_x(x,\lambda)$, we denote respectively the Jacobian matrix of $R(x,t,\lambda)$ with respect to $x$ and the Frêchet derivative of $d(x,\lambda)$ with respect to $x$.

Let us assume (3.1) has an isolated solution $x = x_o(t)$. Then there is a positive number $\delta_o$ such that

$$\Omega_o = \{(t,x) \mid \|x - x_o(t)\| \leq \delta_o,\ t \in I\} \subset \Omega .$$

Putting

$$D'_o = \{x(t) \mid (t,x(t)) \in \Omega_o \ \text{ for } \ t \in I,\ x(t) \in C[I]\},$$

let us assume

(4.3) $\quad \|X_x(x',t) - X_x(x'',t)\|_C + \|f_x(x') - f_x(x'')\| \leq K_o\|x' - x''\|_C$

for any $x', x'' \in D'_o$;

(4.4) $$\begin{cases} \|R_x(x,t,\lambda)\|_C + \|d_x(x,\lambda)\| \leq K_1, \\ \|R(x,t,\lambda) - R(x,t,0)\|_C + \|d(x,\lambda) - d(x,0)\| \leq K_2|\lambda|, \end{cases}$$

on $D'_o \times \Lambda$.

If we put

$$F(x) = [\frac{dx}{dt} - X(x,t),\ f(x)]$$

and denote the weak Fréchet derivative of $F(x)$ by $J(x)$, then, by the isolatedness of $x = x_o(t)$, we have the linear inverse operator $J^{-1}(x_o)$ and consequently we have some positive number $M$ such that

$$(4.5) \qquad \|J^{-1}(x_o)\|_C \leqq M.$$

Now, if we apply the common perturbation method to the boundary value problem (4.1), then we get the first approximate solution $x = \bar{x}(t)$ of the form

$$(4.6) \qquad \bar{x} = x_o - \lambda u,$$

where

$$u = J^{-1}(x_o)U(x_o) \quad \text{and} \quad U(x_o) = [-R(x_o,t,0),\ d(x_o,0)].$$

Take $J(x_o)$ for $L$ of Theorem 2 and suppose

$$(4.7) \qquad \|u\|_C \leqq \sigma .$$

Then we easily see that the conditions of Theorem 2 are all satisfied by system (4.1) and the approximate solution $\bar{x}$ if

$$(4.8) \begin{cases} \delta + |\lambda|\sigma \leqq \delta_o, \\ K_o(\delta + |\lambda|\sigma) + |\lambda|K_1 \leqq \kappa/M, \\ \dfrac{M(\frac{1}{2}K_o\sigma^2 + K_1\sigma + K_2)}{1-\kappa}\lambda^2 \leqq \delta, \end{cases}$$

for some $\delta$ and $\kappa < 1$. If we put

$$\alpha = \delta/|\lambda|,$$

we can simplify the condition (4.8) and finally we see that the conditions of Theorem 2 are all satisfied if

(4.9) $\quad |\lambda| \leqq \lambda_0(\alpha, \varkappa),$

where

(4.10) $$\lambda_0(\alpha,\varkappa) = \min \left[\rho, \frac{\delta_0}{\alpha+\sigma}, \frac{\varkappa}{M[K_0(\varkappa+\sigma)+K_1]}, \frac{(1-\varkappa)\alpha}{M(\frac{1}{2}K_0\sigma^2+K_1\sigma+K_2)}\right].$$

By Theorem 2, this implies that <u>for any positive number</u> $\alpha$ <u>and any positive number</u> $\varkappa$ <u>less than</u> 1, <u>the quantity</u> $\lambda_0(\alpha,\varkappa)$ <u>given by</u> (4.10) <u>gives a bound of the parameter</u> $\lambda$ <u>within which the perturbation method is really effective</u>, in other words, <u>the perturbation method really produces an approximate solution</u>. By Theorem 2, it is easily seen that for $\lambda$ satisfying (4.9), the approximate solution (4.6) obtained by the perturbation method is within the error

$$O(\lambda^2) = \frac{M(\frac{1}{2}K_0\sigma^2 + K_1\sigma + K_2)}{1-\varkappa}\lambda^2.$$

<u>Remark</u>. Consider the special case where

$$\Omega_0 = \{(t,x) \mid t \in I, -\infty < \|x\| < \infty\} = \Omega$$

and the unperturbed system (3.1) is linear with respect to x. In such a case, we may suppose that

$$\delta_0 = \infty \text{ and } K_0 = 0.$$

Then from (4.10), we see that

$$\lambda_o(\alpha,\varkappa) = \min \left[\rho\, ,\ \frac{\varkappa}{MK_1}\, ,\ \frac{(1-\varkappa)\alpha}{M(K_1\sigma + K_2)}\right].$$

However we can take $\alpha$ as large as we desire. Hence we finally see that, in the special case under consideration, the bound of the parameter $\lambda$ can be given by

$$\min\,(\rho\, ,\ \varkappa/MK_1),$$

where $\varkappa$ is an arbitrary positive number less than 1.

## References


[1] Urabe, M., Galerkin's procedure for nonlinear periodic systems, Arch. Rational Mech. Anal., 20(1965), 120-152.

[2] Urabe, M., An existence theorem for multi-point boundary value problems, Funkcial. Ekvac., 9(1966), 43-60.