Publ. RIMS, Kyoto Univ.
Vol. 5 (1969), pp. 1-9

A Geometric Method of Numerical Solution of Nonlinear Equations and Error Estimation by Urabe's Proposition

By

Yoshitane SHINOHARA*

1. Introduction

In his paper [1], Rybashov proposed a method of approximate solution by an analogue computer of a system of nonlinear equations

$$(E) \qquad F(x) = \{f_k(x_1, x_2, \cdots, x_n)\} = 0 \qquad (k=1, 2, \cdots, n).$$

However his method can be practised easily on a digital computer and by doing so, we can get a more accurate solution of nonlinear equations. In the present note, we shall describe a method to practise Rybashov's method on a digital computer and illustrate our method with a system of transcendental equations which appears in making a map of Japan by a simple conic projection.

In numerical solution of nonlinear equations, after finding an approximate solution in any way, it is also important to verify the existence of an exact solution and to know the error bound of the approximate solution obtained. In the present note, we shall show that we can indeed verify the existence of an exact solution and know the error bound of the approximate solution obtained from the approximate solution itself by the use of Urabe's proposition [2].

The author expresses his hearty gratitude to Professor Urabe for his kind guidance and constant advice.

Received March 5, 1969.
Communicated by M. Urabe.
* Technical College of Tokushima University.

2. The Method of Computation

From equations of the system (E), we choose $n-1$ equations, say,

$$f_\alpha(x_1, x_2, \cdots, x_n) = 0 \qquad (\alpha = 1, 2, \cdots, n-1)\,. \tag{2.1}$$

The condition necessary for a system of equations (2.1) is only that the rank of the matrix $(\partial f_\alpha/\partial x_i)$ $(\alpha = 1, 2, \cdots, n-1\,;\ i = 1, 2, \cdots, n)$ is equal to $n-1$. The system of equations (2.1) then determines a curve

$$C:\ \boldsymbol{x} = \boldsymbol{x}(s)\,,$$

for which from (2.1) we have

$$\sum_{i=1}^{n} \frac{\partial f_\alpha}{\partial x_i} \cdot \frac{dx_i}{ds} = 0 \qquad (\alpha = 1, 2, \cdots, n-1)\,. \tag{2.2}$$

Put

$$D_i = (-1)^i \cdot \frac{\partial(f_1, f_2, \cdots\cdots, f_{n-1})}{\partial(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)} \qquad (i = 1, 2, \cdots, n)\,, \tag{2.3}$$

then from (2.2) we have

$$\frac{dx_i}{ds} = \lambda D_i \qquad (i = 1, 2, \cdots, n)\,, \tag{2.4}$$

where λ is an arbitrary parameter. Let us choose a parameter s so that s may be an arc length of the curve C. Then we readily see that λ is given by

$$\lambda = \pm\Big[\sum_{i=1}^{n} D_i^2\Big]^{-1/2}\,. \tag{2.5}$$

Hence from (2.4), for the curve C, we have a system of differential equations of the form

$$\frac{d\boldsymbol{x}}{ds} = \boldsymbol{X}(\boldsymbol{x})\,. \tag{2.6}$$

Now we take a point $\boldsymbol{x} = \boldsymbol{x}^{(0)}$ on the curve C and suppose $\boldsymbol{x}^{(0)} = \boldsymbol{x}(0)$. Then we can compute the curve C integrating numerically equation (2.6) by a step-by-step method, say, the Runge-Kutta method. Let $\boldsymbol{x}^{(l)}$ $(l = 1, 2, \cdots)$ be the approximate value of $\boldsymbol{x}(s)$

obtained at the l-th step by the numerical integration. Then we may have $f_n[\boldsymbol{x}^{(0)}]\cdot f_n[\boldsymbol{x}^{(1)}]\leqq 0$. Otherwise we continue the numerical integration of (2.6) until we have

$$f_n[\boldsymbol{x}^{(l-1)}]\cdot f_n[\boldsymbol{x}^{(l)}]\leqq 0\,. \tag{2.7}$$

Once we have had (2.7) for some l, we check if $|f_n[\boldsymbol{x}^{(l-1)}]|$ or $|f_n[\boldsymbol{x}^{(l)}]|$ is smaller than a specified positive number ε. If this is not satisfied, we multiply the step-size of the numerical integration by 2^{-p} ($p\geqq 1$) and repeat the numerical integration starting from the value $\boldsymbol{x}^{(l-1)}$. If we repeat this process, then after a finite number of repetitions we shall have

$$\begin{cases} f_n[\boldsymbol{x}^{(l-1)}]\cdot f_n[\boldsymbol{x}^{(l)}]\leqq 0\,, \\ |f_n[\boldsymbol{x}^{(l-1)}]| \quad \text{or} \quad |f_n[\boldsymbol{x}^{(l)}]|<\varepsilon\,, \end{cases} \tag{2.8}$$

provided on the curve C there is a simple solution of (E), that is, a solution of (E) for which the Jacobian of $\boldsymbol{F}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ does not vanish. The value $\boldsymbol{x}^{(l-1)}$ or $\boldsymbol{x}^{(l)}$ satisfying (2.8) gives an approximate solution of the given system of equations (E). Starting from $\boldsymbol{x}^{(l-1)}$ or $\boldsymbol{x}^{(l)}$, we then can compute a solution of (E) by, say, Newton's method. However, if ε is very small, $\boldsymbol{x}^{(l-1)}$ or $\boldsymbol{x}^{(l)}$ itself will give an accurate approximate solution of (E).

Remark 1. In the course of the numerical integration of (2.6), it may happen that

$$|f_\alpha[\boldsymbol{x}^{(l)}]|\geqq\zeta \tag{2.9}$$

at some l-th step for some positive integer $\alpha\leqq n-1$, where ζ is a prescribed positive number. When (2.9) happens, one however can correct $\boldsymbol{x}^{(l)}$ so that the corrected value $\tilde{\boldsymbol{x}}^{(l)}$ may satisfy inequalities

$$|f_\alpha[\tilde{\boldsymbol{x}}^{(l)}]|<\zeta$$

for all $\alpha=1, 2, \cdots, n-1$. To do so, it suffices to apply Newton's method to (2.1) starting from $\boldsymbol{x}=\boldsymbol{x}^{(l)}$ leaving one of $x_i^{(l)}$'s ($i=1, 2, \cdots, n$) fixed.

Remark 2. To find a point $\boldsymbol{x}=\boldsymbol{x}^{(0)}$ on the curve C, it suffices

to find a solution of the system (2.1) consisting of $n-1$ equations after asigning a suitable value to some one of x_i's ($i=1, 2, \cdots, n$). Our method is clearly applicable to systems consisting of $n-1$ equations. Repeating such a process, the problem thus is reduced to finding a solution of a single equation.

Remark 3. If we continue the numerical integration of (2.6) through the approximate solution obtained or begin the numerical integration of (2.6) in the reverse direction, then we shall have (2.8) again provided there is another simple solution of (E) on the curve C. Continuing our process, we thus can get numerically all finite simple solutions of (E) lying on the curve C.

3. Urabe's Proposition

Urabe's proposition [2]. *Let*

$$\boldsymbol{F}(\boldsymbol{x}) = 0 \tag{3.1}$$

be a given real system of equations and suppose that $\boldsymbol{F}(\boldsymbol{x})$ *is continuously differentiable with respect to* $\boldsymbol{x}$ *in a region* Ω *of the* $\boldsymbol{x}$*-space. Assume that equation* (3.1) *possesses an approximate solution* $\boldsymbol{x}=\hat{\boldsymbol{x}}$ *such that the Jacobian matrix* $J(\boldsymbol{x})$ *of* $\boldsymbol{F}(\boldsymbol{x})$ *with respect to* $\boldsymbol{x}$ *is non-singular for* $\boldsymbol{x}=\hat{\boldsymbol{x}}$ *and there are a positive constant* δ *and a non-negative number* $\kappa<1$ *satisfying the following conditions for any norm of vectors and matrices*:

$$\begin{cases} \text{(i)} & \Omega_\delta = \{\boldsymbol{x} \mid \|\boldsymbol{x}-\hat{\boldsymbol{x}}\| \leqq \delta\} \subset \Omega\,, \\ \text{(ii)} & \|J(\boldsymbol{x})-J(\hat{\boldsymbol{x}})\| \leqq \dfrac{\kappa}{M} \quad \textit{for any} \quad \boldsymbol{x}\in\Omega_\delta\,, \\ \text{(iii)} & \dfrac{Mr}{1-\kappa} \leqq \delta\,, \end{cases} \tag{3.2}$$

where r *and* M ($\geqq 0$) *are numbers such that*

$$\|F(\hat{\boldsymbol{x}})\| \leqq r \quad \textit{and} \quad \|J^{-1}(\hat{\boldsymbol{x}})\| \leqq M\,. \tag{3.3}$$

Equation (3.1) *then possesses one and only one solution* $\boldsymbol{x}=\bar{\boldsymbol{x}}$ *in* Ω_δ *and we have*

$$(3.4) \qquad \|\hat{x}-\bar{x}\| \leqq \frac{Mr}{1-\kappa}.$$

For the proof, see [2], pp. 124–125.

The proposition gives conditions under which the existence of an approximate solution implies the existence of an exact solution, and the error estimate of an approximate solution can be got from itself without knowing an exact solution. Hence by verifying the conditions of the above proposition for an approximate solution obtained by computation, we can know the existence of an exact solution and at the same time get the error bound of the approximate solution obtained by computation.

4. An Example

In order to make a map of Japan by a simple conic projection, it is necessary to find a function $y(x)$ which satisfies the first order differential equation

$$(4.1) \qquad \frac{dy}{dx}+\frac{\eta}{\sin x}\cdot y = 2$$

and the boundary condition

$$(4.2) \qquad y'(\alpha) = y'(\beta) = \frac{\eta}{\sin \theta}\cdot y(\theta)$$

for an appropriate positive constant η. In (4.2), θ is a solution of the equation

$$(4.3) \qquad y''(\theta) = 0,$$

and

$$\alpha = 44^\circ, \quad \beta = 66^\circ.$$

As readily seen, the general solution of (4.1) can be given by

$$(4.4) \qquad y = y_\eta(x) = 2\tan^{-\eta}(x/2)\left[\int_{x_1}^{x} \tan^{\eta}(x/2)\,dx + C_\eta\right],$$

where x_1 may be supposed without loss of generality to be equal to $55^\circ=(44^\circ+66^\circ)/2$. From the boundary condition $y'(\alpha)=y'(\beta)$, it

readily follows that

$$C_\eta = [\sin\alpha\cdot\tan^\eta(\alpha/2)\cdot\int_{x_1}^{\beta}\tan^\eta(x/2)\,dx - \sin\beta\cdot\tan^\eta(\beta/2)\cdot\int_{x_1}^{\alpha}\tan^\eta(x/2)\,dx]\Big/ [\sin\beta\cdot\tan^\eta(\beta/2) - \sin\alpha\cdot\tan^\eta(\alpha/2)]\,. \tag{4.5}$$

Now, differentiating equation (4.1) with respect to x, we see that equation (4.3) is equivalent to the equation

$$y_\eta(\theta) - \frac{2\sin\theta}{\eta+\cos\theta} = 0\,. \tag{4.6}$$

From the boundary condition $y'(\beta)=\eta\cdot y(\theta)/\sin\theta$, we further have

$$2 - \frac{\eta}{\sin\beta}\cdot y_\eta(\beta) = \frac{\eta}{\sin\theta}y_\eta(\theta)\,. \tag{4.7}$$

Thus we see that for solving the initial problem, it suffices to solve numerically the system of equations

$$\begin{cases} F(\eta,\theta) \triangleq y_\eta(\theta) - \dfrac{2\sin\theta}{\eta+\cos\theta} = 0\,, \\ G(\eta,\theta) \triangleq y_\eta(\theta) - \dfrac{2\sin\theta}{\eta} + \dfrac{\sin\theta}{\sin\beta}\cdot y_\eta(\beta) = 0 \end{cases} \tag{4.8}$$

for the function $y_\eta(x)$ given by (4.4) and (4.5).

We shall now apply our method to the above system of equations. According to (2.4)~(2.5), for the curve C determined by $F(\eta,\theta)=0$, we have

$$\frac{d\eta}{ds} = -\frac{F_\theta}{\sqrt{F_\eta^2+F_\theta^2}}\,, \quad \frac{d\theta}{ds} = \frac{F_\eta}{\sqrt{F_\eta^2+F_\theta^2}}\,, \tag{4.9}$$

and solving the equation $F(0,\theta)=0$ numerically, we get

$$\eta = 0\,, \quad \theta = 0.95114\ 50249\,, \tag{4.10}$$

which can be assumed to be an initial value for the solution of (4.9) corresponding to $s=0$.

We specify the values of ε and ξ so that

$$\varepsilon = 10^{-9}\,, \quad \xi = 10^{-8}\,. \tag{4.11}$$

In the numerical integration of (4.9), we start with step-size 2^{-5} and multiply the step-size by 2^{-3} when the second inequality of (2.8) is detected to be unfulfilled.

On the computer TOSBAC 3400, we have got ;

$$\begin{aligned}(4.12)\qquad &\eta^{(54)} = 0.57355\ 02977\,, \quad \theta^{(54)} = 0.94679\ 70998\,;\\ &\eta^{(53)} = 0.57355\ 02976\,, \quad \theta^{(53)} = 0.94679\ 70998\,;\\ &G(\eta^{(53)}, \theta^{(53)})\cdot G(\eta^{(54)}, \theta^{(54)}) < 0\,;\end{aligned}$$

$$(4.13)\qquad \begin{cases} F(\eta^{(54)}, \theta^{(54)}) = -0.43655\ 74569\times 10^{-10}\,, \\ G(\eta^{(54)}, \theta^{(54)}) = \ \ 0.43655\ 74569\times 10^{-10}\,. \end{cases}$$

For the purpose of making a map, it is unnecessary to find all solutions, but it suffices to get any one of the solutions. Hence we have stopped the computation after having found the above $(\eta^{(54)}, \theta^{(54)})$.

In order to get an error bound for the approximate solution

$$(4.14)\qquad \hat{\boldsymbol{x}} = \{\hat{\eta}, \hat{\theta}\} = \{\eta^{(54)}, \theta^{(54)}\}\,,$$

we apply Urabe's proposition to the equation (4.8) using the Euclidean norms. Put

$$(4.15)\qquad \boldsymbol{x} = \{\eta, \theta\}, \quad \boldsymbol{F}(\boldsymbol{x}) = \{F(\eta, \theta), G(\eta, \theta)\}\,,$$

then from (4.13) readily follows

$$(4.16)\qquad ||\boldsymbol{F}(\hat{\boldsymbol{x}})|| < 0.437\times\sqrt{2}\times 10^{-10} < r = 0.437\times 1.415\times 10^{-10}\,.$$

Put

$$J(x) = \begin{bmatrix} F_\eta(\eta, \theta) & F_\theta(\eta, \theta) \\ G_\eta(\eta, \theta) & G_\theta(\eta, \theta) \end{bmatrix},$$

then for $\boldsymbol{x}=\hat{\boldsymbol{x}}$, we have

$$J(\hat{\boldsymbol{x}}) = \begin{bmatrix} -0.74233\ 20101\times 10^{-2} & -0.98257\ 30255\times 10^{0} \\ \ \ 0.24747\ 58356\times 10 & \ \ 0.18917\ 48980\times 10^{-9} \end{bmatrix},$$

$$J^{-1}(\hat{\boldsymbol{x}}) = \begin{bmatrix} \ \ 0.77797\ 54129\times 10^{-10} & \ \ 0.40407\ 98559\times 10^{0} \\ -0.10177\ 36060\times 10 & -0.30528\ 15454\times 10^{-2} \end{bmatrix},$$

and hence

(4.17) $$||J^{-1}(\hat{\boldsymbol{x}})|| < M = 1.096\,.$$

Let Ω be the region such that

(4.18) $$\Omega = \{\boldsymbol{x} = (\eta, \theta) : |\eta - \hat{\eta}| \leqq H,\ \hat{\theta} - 11H \leqq \theta \leqq \hat{\theta} + 10H\}$$

where $H = 0.01745\ 32925$ ($0.01745\ 32925$ radian $= 1°$). Then, computing the values of $||J(\boldsymbol{x}) - J(\hat{\boldsymbol{x}})||$ for grid points

$$x = x_{ij} = (\eta_i, \theta_j) = \left(\hat{\eta} + \frac{H}{8}i,\ \hat{\theta} + \frac{H}{8}j\right)$$
$$(i = 0, \pm 1, \cdots, \pm 8\,;\ j = 0, \pm 1, \cdots, \pm 80, -81, -82, \cdots, -88)\,,$$

we see that

(4.19) $$||J(\boldsymbol{x}) - J(\hat{\boldsymbol{x}})|| \leqq 0.890$$

for any $\boldsymbol{x} \in \Omega$. Put

(4.20) $$\delta = 0.0174\,,$$

then evidently

(4.21) $$\Omega_\delta \subset \Omega\,,$$

and we have (4.19) for any $\boldsymbol{x} \in \Omega_\delta$. Hence by (4.16), (4.17), (4.19) and (4.20), we see that the conditions (ii) and (iii) of (3.2) in Urabe's proposition are fulfilled if there is a positive number $\kappa < 1$ satisfying the following inequalities:

(4.22) $$\begin{cases} 0.890 \leqq \dfrac{\kappa}{1.096}\,, \\ \dfrac{1.096 \times 0.437 \times 1.415 \times 10^{-10}}{1 - \kappa} \leqq 0.0174\,. \end{cases}$$

These inequalities are equivalent to the inequalities

$$0.890 \times 1.096 \leqq \kappa \leqq 1 - \frac{1.096 \times 0.437 \times 1.415}{0.0174} \times 10^{-10}\,,$$

that is,

(4.23) $$0.97544 \leqq \kappa \leqq 0.99999\ 99961\,0 \cdots\,.$$

Hence, indeed, there is a positive number $\kappa < 1$ satisfying (4.22). This proves that all the conditions of Urabe's proposition are ful-

filled by the approximate solution $\boldsymbol{x}=\hat{\boldsymbol{x}}$. Thus we see that equation (4.8) possesses one and only one exact solution $\boldsymbol{x}=\bar{\boldsymbol{x}}$ in Ω_8 and that

$$(4.24)\qquad \|\hat{\boldsymbol{x}}-\bar{\boldsymbol{x}}\| \leqq \frac{1.096\times 0.437\times 1.415}{1-\kappa}\times 10^{-10},$$

where κ is an arbitrary number satisfying (4.23). From (4.23) and (4.24), we then see that

$$(4.25)\qquad \|\hat{\boldsymbol{x}}-\bar{\boldsymbol{x}}\| \leqq \frac{1.096\times 0.437\times 1.415}{1-0.97544}\times 10^{-10} < 2.76\times 10^{-9},$$

which gives an error bound for the approximate solution $\boldsymbol{x}=\hat{\boldsymbol{x}}=\{\eta^{(54)}, \theta^{(54)}\}$ given by (4.12).

For the function $\hat{y}(x)=y_{\hat{\eta}}(x)$, we have

$$\hat{y}'(\alpha) = 0.99072\ 83110,$$
$$\hat{y}'(\beta) = 0.99072\ 83110,$$

and

$$\hat{\eta}\hat{y}(\hat{\theta})/\sin\hat{\theta} = 0.99072\ 83109.$$

These show that $y=\hat{y}(x)$ satisfies the given boundary conditions accurately.

Remark. In the above computation, we have used the following formulas:

$$\int_{x_1}^{x_2}\tan^{\eta}(x/2)\,dx = 2\sum_{i=0}^{\infty}\frac{(-1)^i}{\eta+2i+1}(t_2^{\eta+2i+1}-t_1^{\eta+2i+1}),$$

$$\int_{x_1}^{x_2}\tan^{\eta}(x/2)\cdot\log\tan(x/2)\cdot dx$$
$$= -2\sum_{i=0}^{\infty}\frac{(-1)^i}{(\eta+2i+1)^2}[t_2^{\eta+2i+1}\{1-(\eta+2i+1)\log t_2\}$$
$$-t_1^{\eta+2i+1}\{1-(\eta+2i+1)\log t_1\}]$$

where $t_i=\tan(x_i/2)$ $(i=1,2)$.

References

[1] Рыбашов, М. В., Об одном методе решения глобальной задачи отыскания корней конечных уравнений при помощи злектронной модели, Автомат. и Телемех. **23** (1962), 1396–1398.

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