44

SINGLE STEP METHODS FOR THE SOLUTION OF SINGULAR INITIAL VALUE PROBLEMS

by

David L. Russell

Computer Sciences Technical Report #51

December 1968

SINGLE STEP METHODS FOR THE SOLUTION
OF SINGULAR INITIAL VALUE PROBLEMS*

by

David L. Russell†

1. The Differential Equation

We consider a system of nonlinear ordinary differential equations

(1.1) $$x^r \frac{dy}{dx} = A(x)y + f(x,y).$$

Here r is a positive real number, y is an n-dimensional vector with real or complex components, x is a real variable, and $A(x)$ is an $n \times n$ matrix with real or complex entries which is continuous for $0 \leq x \leq b$, $y \in C^n$, with the properties

(1.2) $$f(0,0) = 0,$$

(1.3) $$\|f(x,y_1) - f(x,y_2)\| \leq L(\eta)\|y_1 - y_2\|$$

where $\eta = \max\{\|y_1\|, \|y_2\|\}$. The Lipschitz factor $L(\eta)$ satisfies

(1.4) $$\lim_{\eta \to 0+} L(\eta) = 0.$$

---

*Supported in part by the National Science Foundation under Grant GP-8927

†Departments of Computer Sciences and Mathematics, University of Wisconsin, Madison

Our interest lies with continuous vector valued functions $y(x)$ defined for $0 \leq x \leq a \leq b$ with

(1.5) $$y(0) = 0$$

and satisfying (1.1) for $x > 0$. In general no such functions exist and if they do there is no guarantee that there is only one of them. The scalar equation

$$x \frac{dy}{dx} = g(x), \quad g(x) = \begin{cases} \frac{1}{\log x}, & 0 < x < 1, \\ 0, & x = 0 \end{cases}$$

has no such solution because for each $x > 0$ the integral

$$\int_0^x \frac{d\xi}{\xi \log \xi} = \int_0^x \frac{1}{\log \xi} \frac{d}{d\xi} (\log \xi)\, d\xi = \int_{-\infty}^{\log x} \frac{1}{u}\, du = -\infty.$$

On the other hand the equation

$$x \frac{dy}{dx} = y$$

is satisfied by every linear function

$$y(x) = \alpha x, \quad \alpha \text{ real},$$

and thus has infinitely many solutions of the specified type.

The following theorem provides sufficient conditions for (1.1) to have exactly one solution satisfying (1.5). In addition it provides us with a tool which will prove useful later when we consider a certain numerical technique for solving (1.1). No particular originality is claimed for the proof. (See, e.g., [1], p. 55ff.)

<u>Theorem 1</u>. <u>Assume that all of the eigenvalues of</u> $A = A(0)$ <u>have negative real parts and that</u>

$$T : C[0,a] \to L^{\infty}[0,a]$$

<u>is a mapping which satisfies</u>

$$T(0) = 0,$$

$$\|T(f_1) - T(f_2)\|_{\infty} \leq \|f_1 - f_2\|_C ,$$

<u>where</u> $\| \ \|_C$ <u>denotes the "sup" norm on</u> $C[0,a]$, <u>the space of continuous</u> n-<u>vector functions on</u> $[0,a]$, <u>and</u> $\| \ \|_{\infty}$ <u>denotes the "essential sup" norm on</u> $L^{\infty}[0,a]$. <u>Then, if</u> a <u>is a sufficiently small positive number, there is exactly one function</u> $y \in C[0,a]$ <u>which satisfies</u> (1.5) and

$$(1.6) \qquad x^r \frac{dy}{dx} = Ay + T(\tilde{A}(\cdot)y(\cdot) + f(\cdot, y(\cdot)))(x)$$

<u>for almost all</u> x <u>in</u> $(0,a]$, <u>where</u>

$$A(x) = A + \tilde{A}(x).$$

<u>Remark</u>. Choosing T to be the identity operator injecting $C[0,a]$ into $L^{\infty}[0,a]$ we obtain the desired existence and uniqueness theorem for the equation (1.1).

<u>Proof of Theorem 1.</u> We put

$$(1.7) \qquad K(x,\xi) = \begin{cases} \exp\left(\dfrac{A}{1-r}\left(\dfrac{1}{x^{r-1}} - \dfrac{1}{\xi^{r-1}}\right)\right), & r \neq 1 \\[2ex] \exp(A(\log x - \log \xi)) \equiv \left(\dfrac{x}{\xi}\right)^{A}, & r = 1, \end{cases}$$

and consider the mapping $V : C[0,a] \to C[0,a]$ defined by

(1.8) $$V(y)(x) = \int_0^x K(x,\xi)\frac{1}{\xi^r} T(\tilde{A}(\cdot)y(\cdot) + f(\cdot, y(\cdot)))(\xi)\, d\xi,$$

which is, of course, inspired by the variation of parameters formula. A fixed point of V provides us with a solution of (1.5), (1.6). We let Y denote the complete metric space

$$Y = \{y \in C[0,a] \mid \|y\|_C \le B\}.$$

The theorem will have been proved if we can show that V represents a contraction mapping from Y into itself for a, B sufficiently small.

To show $V : Y \to Y$ we must compute

$$\|V(y)(x)\| \le \int_0^x \|K(x,\xi)\|\frac{1}{\xi^r}\|T(\tilde{A}(\cdot)y(\cdot) + f(\cdot, y(\cdot)))(\xi)\|\, d\xi$$

$$\le \|T(\tilde{A}(\cdot)y(\cdot) + f(\cdot, y(\cdot)))\|_\infty \int_0^x \|K(x,\xi)\|\frac{1}{\xi^r}\, d\xi$$

$$\le \|\tilde{A}(\cdot)y(\cdot) + f(\cdot, y(\cdot)))\|_C \int_0^x \|K(x,\xi)\|\frac{1}{\xi^r}\, d\xi$$

Let us put

$$\|\tilde{A}(\cdot)\|_C = \delta \ge 0$$

and observe that we can make δ as small as we wish by choosing a small, since $\tilde{A}(x)$ is continuous and $\tilde{A}(0) = 0$. The notation $\|\tilde{A}(\cdot)\|_C$ refers, of course, to the norm of $\tilde{A}(\cdot)$ considered as a linear operator on $C[0,a]$. On the other hand

$$\|f(\cdot,y(\cdot))\|_C \le \|f(\cdot,y(\cdot)) - f(\cdot,0)\|_C + \|f(\cdot,0)\|_C$$

$$\le L(B)B + \|f(\cdot,0)\|_C \le L(B)B + \varepsilon$$

provided $y \in Y$ and a is chosen sufficiently small. Here we have used (1.2), (1.3) and the continuity of $f(x,0)$.

Assuming for the moment that

(1.9) $$\int_0^x \|K(x,\xi)\| \frac{1}{\xi^r} d\xi \le K, \quad x \in [0,a]$$

where K is some positive constant, we have

$$\|V(y)\|_C \le K(\delta B + L(B)B + \varepsilon)$$

for $y \in Y$. Thus $V: Y \to Y$ if

(1.10) $$K(\delta B + L(B)B + \varepsilon) \le B.$$

Since δ and ε can be made as small as we wish by choosing a small and, by (1.4), $L(B) \to 0$ as $B \to 0$, we can ensure (1.10) by making a and B sufficiently small.

To establish the contracting property of V in Y we compute

(1.11) $$\|V(y_1)(x) - V(y_2)(x)\| = \Big\|\int_0^x K(x,\xi)\frac{1}{\xi^r}(T(\tilde{A}(\cdot)$$

$$(y_1(\cdot) - y_2(\cdot)) + f(\cdot,y_1(\cdot)) - f(\cdot,y_2(\cdot)))(\xi)\,d\xi\Big\|$$

$$\le \left(\int_0^x \|K(x,\xi)\|\frac{1}{\xi^r}d\xi\right)\left(\|\tilde{A}(\cdot)\|_C\|y_1(\cdot) - y_2(\cdot)\|_C\right.$$

$$\left.+ \|f(\cdot,y_1(\cdot)) - f(\cdot,y_2(\cdot))\|_C\right)$$

$$\le K(\delta + L(B))\|y_1(\cdot) - y_2(\cdot)\|_C$$

If a and B are sufficiently small we have

$$K(\delta + L(B)) = \alpha, \quad 0 \leq \alpha < 1.$$

Then, since (1.11) holds for all $x \in [0,a]$ we have

$$\|V(y_1)(\cdot) - V(y_2)(\cdot)\|_C \leq \alpha \|y_1(\cdot) - y_2(\cdot)\|$$

and $V: Y \to Y$ is a contraction.

Thus, to complete the proof of the theorem, all that remains is to prove (1.9).

Because the eigenvalues of A all have negative real parts there are positive numbers M and λ such that

$$\|\exp(At)\| \leq M e^{-\lambda t}, \quad t \geq 0.$$

Thus, in order to estimate $\int_0^x \|K(x,\xi)\| \frac{1}{\xi^r} d\xi$ we need only evaluate an integral of the form

$$\int_0^x \exp(-\lambda(f(x) - f(\xi)))\, f'(\xi)\, d\xi$$

where $f(x)$ is monotone increasing for $0 < x \leq a$. (cf. (1.7)).
Letting

$$s = f(x), \quad \sigma = f(\xi)$$

we have

$$(1.12) \qquad \int_0^x \|K(x,\xi)\| \frac{1}{\xi^r} d\xi \leq M \int_{\sigma = f(0+)}^{\sigma = s = f(x)} \exp(-\lambda(s - \sigma))\, d\sigma$$

$$\leq M \int_{-\infty}^{s} \exp(-\lambda(s - \sigma))\, d\sigma = \frac{M}{\lambda} = K.$$

With this the proof of the theorem is complete.

2. An Explicit First Order Method

We note first of all that the usual explicit first order (i.e. Euler's) method

$$(2.1)\qquad y_{k+1} = y_k + \frac{h}{x_k^r}[A(x_k)y_k + f(x_k, y_k)],$$

where $h = x_{k+1} - x_k$ is the step length, is, in general, of no use. We might attempt to avoid the singularity at $x = 0$ by putting $y_1 = 0$ and carrying out (2.1) for $k \geq 1$. The example of the scalar equation

$$(2.2)\qquad x^r \frac{dy}{dx} = -y + x$$

shows that this is futile. For this equation (2.1) becomes

$$y_{k+1} = y_k + \frac{h}{(kh)^r}[-y_k + kh] = \left(1 - \frac{h}{(kh)^r}\right)y_k + \frac{h}{(kh)^{r-1}} .$$

Thus

$$y_{k+1} = \sum_{\ell=1}^{k-1}\left[\prod_{j=\ell+1}^{k}\left(1 - \frac{h}{(jh)^r}\right)\right]\frac{h}{(\ell h)^{r-1}} + \frac{h}{(kh)^{r-1}}$$

The highest power of h^{-1} which occurs here occurs in just one term and that term is

$$\left[\prod_{j=2}^{k}\left(-\left(\frac{1}{j}\right)^r h^{-r+1}\right)\right] h^{-r+2} = \alpha\, h^{-kr+(k+1)}$$

which approaches $\pm\,\infty$ for $r > 1 + \frac{1}{k}$, $h \to 0$, k fixed. It follows that if $r > 1$ the sequences $\{y_k\}$ obtained by the use of this method do not ever remain bounded as $h \to 0$ and thus cannot converge to the true solution of (2.2).

We will now develop an explicit first order method based on the integral equation (1.8) and we will show that the method is convergent if the hypotheses of Theorem 1 are satisfied.

We note from (1.7) that $K(x,\xi)$ satisfies

$$\text{(2.3)} \qquad \frac{\partial K}{\partial x} = \frac{A}{x^r} K, \quad \frac{\partial K}{\partial \xi} = \frac{-A}{\xi r} K.$$

Moreover, $y(x)$ satisfies

$$y(x_{k+1}) = K(x_{k+1}, x_k)\, y(x_k)$$

$$+ \int_{x_k}^{x_{k+1}} K(x_{k+1},\xi) \frac{1}{\xi r} (\tilde{A}(\xi) y(\xi) + f(\xi, y(\xi)))\, d\xi$$

$$\approx K(x_{k+1},x_k) y(x_k) + \left[\int_{x_k}^{x_{k+1}} K(x_{k+1},\xi) \frac{1}{\xi r}\, d\xi \right] \Big(\tilde{A}(x_k) y(x_k)$$

$$+ f(x_k, y(x_k)) \Big).$$

Now (2.3) shows that

$$\text{(2.4)} \qquad \int_{x_k}^{x_{k+1}} K(x_{k+1},\xi) \frac{1}{\xi r} d\xi = -A^{-1} \left(\int_{x_k}^{x_{k+1}} \frac{\partial K(_{k+1},\xi)}{\partial \xi}\, d\xi \right)$$

$$= A^{-1} (K(x_{k+1},x_k) - I).$$

It therefore becomes plausible to approximate $y(x_k)$ by y_k, where the sequence $\{y_k\}$ is determined by

(2.5) $$y_0 = 0,$$

$$y_{k+1} = K(x_{k+1},x_k)y_k + A^{-1}[K(x_{k+1},x_k) - I]$$

$$(\tilde{A}(x_k)y_k + f(x_k,y_k)), \; k \leq 0.$$

The proof of the convergence of this method proceeds as follows. We define an operator

$$D_h : C[0,a] \to L^{\infty}[0,a],$$

called the discretization operator for step-length h, by

(2.6) $$(D_h g)(x) = g(x_k), \; x \in [x_k, x_{k+1}).$$

We note that $\|D_h g\|_{\infty} \leq \|g\|_C$ for all $g \in C[0,a]$ and $\|D_h\|$ (as an operator from $C[0,a]$ into $L^{\infty}[0,a]) = 1$. Thus D_h satisfies the hypotheses on T in Theorem 1. Replacing T by D_h in (1.8) we see that the nonlinear operator defined by

(2.7) $$V_h(y)(x) = \int_0^x K(x,\xi)\frac{1}{\xi^r} D_h(\tilde{A}(\cdot)y(\cdot) + f(\cdot,y(\cdot)))(\xi)\,d\xi$$

has a unique fixed point $\tilde{y} \in Y$. Now we compute

$$\tilde{y}(x_{k+1}) = \int_0^{x_{k+1}} K(x_{k+1},\xi)\frac{1}{\xi^r} D_h(\tilde{A}(\cdot)\tilde{y}(\cdot) + f(\cdot,\tilde{y}(\cdot)))(\xi)\,d\xi$$

$$= \int_{x_k}^{x_{k+1}} K(x_{k+1},\xi)\frac{1}{\xi^r} D_h(\tilde{A}(\cdot)\tilde{y}(\cdot) + f(\cdot,\tilde{y}(\cdot)))(\xi)\,d\xi$$

$$+ K(x_{k+1},x_k)\int_0^{x_k} K(x_k,\xi)\frac{1}{\xi^r} D_h(\tilde{A}(\cdot)\tilde{y}(\cdot) + f(\cdot,\tilde{y}(\cdot)))(\xi)\,d\xi$$

$$= \left(\int_{x_k}^{x_{k+1}} K(x_{k+1},\xi)\frac{1}{\xi^r}\,d\xi\right)(\tilde{A}(x_k)\tilde{y}(x_k) + f(x_k,\tilde{y}(x_k))) + K(x_{k+1},x_k)\tilde{y}(x_k).$$

Then from (2.4) we see that

$$(2.8)\qquad \tilde{y}(x_{k+1}) = K(x_{k+1},x_k)\tilde{y}(x_k) + A^{-1}[K(x_{k+1},x_k) - I]$$
$$\times (\tilde{A}(x_k)\tilde{y}(x_k) + f(x_k,\tilde{y}(x_k)))$$

Comparing (2.8) with the second equation in (2.5) and noting that $\tilde{y}(0)$ is obviously zero, we see that y_k and $\tilde{y}(x_k)$ satisfy the same recursion equation with the same initial condition and therefore

$$y_k = \tilde{y}(x_k),\ k = 0,1,2,\cdots.$$

Thus to estimate $\|y(x_k) - y_k\|$, $k = 0,1,2,\cdots$ it will be sufficient to estimate $\|y-\tilde{y}\|_c$.

If we put $\eta_0 = y$ and define

$$\eta_{k+1} = V_h(\eta_k),\ k \geq 0,$$

with V_h as given by (2.7) we know from the contraction property of V_h in Y that

$$\lim_{k\to\infty} \|\eta_k - \tilde{y}\|_c = 0.$$

Therefore

$$(2.9)\qquad \|y - \tilde{y}\|_c = \lim_{k\to\infty} \|y - \eta_k\|_c \leq$$

$$\lim_{k\to\infty} (\|\eta_0 - \eta_1\|_c + \|\eta_1 - \eta_2\|_c + \cdots + \|\eta_{k-1} - \eta_k\|_c)$$

$$= \lim_{k\to\infty} (\|\eta_0 - \eta_2\|_c + \|V_h(\eta_0) - V_h(\eta_1)\|_c + \cdots \|V_h^{k-1}(\eta_0) - V_h^{k-1}(\eta_1)\|_c)$$

$$\leq \lim_{k\to\infty} \|\eta_0-\eta_1\|_c \left(\sum_{\ell=0}^{h-1} \gamma\,\ell\right) = \|\eta_0-\eta_1\|_c \sum_{\ell=0}^{\infty} \gamma\,\ell$$

$$= \frac{\|\eta_0-\eta_1\|_c}{1-\gamma} .$$

Now

$$\|\eta_0-\eta_1\|_c = \|y-\eta_1\|_c = \|V(y) - V_h(y)\|_c$$

V being given by (1.8) with $T =$ the identity and V_h as defined by (2.7). Now

$$(2.10) \quad ((V-V_h)y(\cdot))(x) = \int_0^x K(x,\xi)\frac{1}{\xi^r} \{\tilde{A}(\xi)y(\xi) + f(\xi, y(\xi))$$

$$- D_h(\tilde{A}(\cdot)y(\cdot) + f(\cdot, y(\cdot)))(\xi)\}\, d\xi.$$

Let $\varepsilon(\delta)$, $\delta > 0$, be the modulus of continuity of the continuous vector valued function $\tilde{A}(x)y(x) + f(x,y(x))$ for $x \in [0,a]$, i.e.,

$$\varepsilon(\delta) = \sup_{\substack{x,x'\in[0,a] \\ |x-x'|\leq\delta}} \|\tilde{A}(x_k)y(x_k) + f(x_k, y(x_k)) - \tilde{A}(x_k')y(x_k') - f(x_k', y(x_k'))\|$$

From the uniform continuity of the function on $[0,a]$ we know that

$$\lim_{\delta\to 0+} \varepsilon(\delta) = 0.$$

Then from (1.12), (2.6) and (2.10) we see that

$$\|((V-V_h)y(\cdot))(x)\| \leq \frac{M}{\lambda}\varepsilon(h) = K\,\varepsilon(h)$$

which gives

$$\|\eta_0 - \eta_1\|_c \leq K\,\varepsilon(h)$$

and (2.9) then shows that

$$\lim_{h\to 0} \|y - \tilde{y}\|_c \leq \lim_{h\to 0} \frac{\|\eta_0 - \eta_1\|_c}{1 - \gamma} = 0$$

and the convergence has been proved.

The fact that the above arguments only prove the convergence of the numerical method in a "sufficiently small" interval [0,a] is not too serious. For when $x \geq a$ the equation (1.1) is no longer singular and the usual techniques (such as in [2], e.g.,) can be used to study the usefulness of this integration technique.

The most serious draw-back of this explicit numerical integration method is the fact that one must continually evaluate $K(x_{k+1}, x_k)$ as given by (1.7). This could be time consuming. If A can be reduced to Jordan form the nilpotent part of A should be combined with $\tilde{A}(x)$ so that we are left with a diagonal matrix Λ instead of A in (1.7). This makes things much simpler but there is still the question of the time required for the evaluation of the exponential function for large arguments, some of which might well be complex.

For this and other reasons we shall now investigate an implicit method for the numerical solution of (1.1) which requires no evaluation of exponentials.

3. An Implicit First Order Method

The method we have in mind consists in approximating $y(x_k) = y(kh)$, $h > 0$ being the step length, by y_k, where $\{y_k\}$ is generated by the recursion equation

$$(3.1)\quad y_{k+1} = y_k + \frac{h}{x_{k+1}^r}[A(x_{k+1})y_{k+1} + f(x_{k+1}, y_{k+1})],$$

which we might call the implicit Euler's method. The plausibility of this method rests on the fact that if $y(x)$ is twice differentiable, then

$$(3.2)\quad y(x_{k+1}) = y(x_k) + \frac{h}{x_{k+1}^r}[A(x_{k+1})y(x_{k+1}) + f(x_{k+1}, y(x_{k+1}))] - \frac{h^2}{2}\hat{y}_k,$$

where $\hat{y}_k$ is a vector whose components are second derivatives of the corresponding components of $y(x)$ evaluated at certain points in the interval (x_k, x_{k+1}).

We assume that the hypotheses of Theorem 1 are valid so that

$$A(x) = A + \tilde{A}(x)$$

where all of the eigenvalues of A have negative real parts and $\lim_{x \to 0+} \tilde{A}(x) = 0$. A similarity transformation

$$P^{-1}AP = \Lambda + \alpha N$$

can be applied to A in such a way that Λ is a diagonal matrix whose diagonal entries are the eigenvalues of A, N is a matrix whose only non-zero entries are possible 1's in the first super-diagonal and $\alpha > 0$ is as small as we wish. If in (3.1) and (3.2) we set

$$y_k = Pz_k, \quad y(x_k) = Pz(x_k)$$

we obtain new equations

(3.3) $$z_{k+1} = z_k + \frac{h}{x_{k+1}^r}[(\Lambda + C(x_{k+1}))z_{k+1} + g(x_{k+1}, z_{k+1})],$$

(3.4) $$z(x_{k+1}) = z(x_k) + \frac{h}{x_{k+1}^r}[(\Lambda + C(x_{k+1}))z(x_{k+1})$$

$$+ g(x_{k+1}, z(x_{k+1})) + \frac{h^2}{2}\hat{z}_k$$

where $C(x)$ is continuous in $[0,a]$ with

(3.5) $$C(0) = \alpha N,$$

$$\hat{z}_k = P^{-1}\hat{y}_k, \quad \text{and}$$

$$g(x,z) = P^{-1} f(x, P_z)$$

satisfies

$$\|g(x,z) - g(x,\tilde{z})\| = \|P^{-1}(f(x,Pz) - f(x,P\tilde{z})\|$$

$$\leq \|P^{-1}\|L(\zeta)\|Pz - P\tilde{z}\| \leq \|P^{-1}\| \ \|P\|L(\rho)\|z - \tilde{z}\|$$

where $\rho = \max\{\|Pz\|, \|P\tilde{z}\|\}$. Thus

(3.6) $$\|g(x,z) - g(x,\tilde{z})\| \leq \hat{L}(\zeta)\|z - \tilde{z}\|$$

where $\zeta = \max\{\|z\|, \|\tilde{z}\|\}$ and, by (1.4)

(3.7) $$\lim_{x \to 0+} \hat{L}(\zeta) = 0$$

Equations (3.3) and (3.4) may be rewritten

$$(3.8)\quad (x_{k+1}I - h\Lambda)\, z_{k+1} = hC(x_{k+1})z_{k+1} + hg(x_{k+1}, z_{k+1}) + x_{k+1}^{r}\, z_k,$$

$$(3.9)\quad (x_{k+1}I - h\Lambda)z(x_{k+1}) = hC(x_{k+1})z(x_{k+1}) + hg(x_{k+1}, z(x_{k+1})) + x_{k+1}^{r}\, z(x_k) + \frac{x_{k+1}^{r} h^2}{2}\, \hat{z}_k$$

If we put

$$R(h\Lambda, x) = (xI - h\Lambda)^{-1}$$

then, taking

$$-\lambda = \max\{\mathrm{Re}(\nu)\,|\,\nu \text{ an eigenvalue of } A\},$$

and remembering that Λ is a diagonal matrix whose entries are the eigenvalues of A, we have

$$\|R(h\Lambda, x)\| \le \frac{1}{x^r + h\lambda}$$

and so we may write

$$(3.10)\quad R(h\Lambda, x) = \frac{T(h\Lambda, x)}{x^r + h\lambda}, \quad \|T(h\Lambda, x)\| \le 1.$$

Then (3.8) and (3.9) become

$$(3.11)\quad z_{k+1} = \frac{h}{x_{k+1}^{r} + h\lambda}\, T(h\Lambda, x_{k+1})\, [C(x_{k+1})\, z_{k+1} + g(x_{k+1}, z_{k+1})] + \frac{x_{k+1}^{r}}{x_{k+1}^{r} + h\lambda}\, T(h\Lambda, x_{k+1}) z_k,$$

$$(3.12)\quad z(x_{k+1}) = \frac{h}{x_{k+1}^r + h\lambda} T(h\Lambda, x_{k+1})[C(x_{k+1})z(x_{k+1}) + g(x_{k+1}, z(x_{k+1}))]$$

$$+ \frac{x_{k+1}^r}{x_{k+1}^r + h\lambda} T(h\Lambda, x_{k+1})z(x_k) + \frac{x_{k+1}^r h^2}{2(x_{k+1}^r + h\lambda)} \hat{z}_k.$$

The equation (3.11) defines z_{k+1} implicitly in terms of z_k. Because of (3.6), (3.7) together with the fact that α may be made as small as we wish, (3.5) and (3.10) and the fact that $\frac{h}{x_{k+1}^r + h\lambda} < \frac{1}{\lambda}$, $\frac{x_{k+1}^r}{x_{k+1}^r + h\lambda} < 1$, whenever $h > 0$, $x_{k+1} > 0$, it is easily verified that for sufficiently small x_{k+1} and sufficiently small z_k (3.11) may be solved by repeated iterations

$$(3.13)\quad z_{k+1,\ell+1} = \frac{h}{x_{k+1}^r + h\lambda} T(h\Lambda, x_{k+1})[C(x_{k+1})z_{k+1,\ell} + g(x_{k+1}, z_{k+1,\ell})]$$

$$+ \frac{x_{k+1}^r}{x_{k+1}^r + h\lambda} T(h\Lambda, x_{k+1}) z_k$$

beginning with $z_{k+1,0} = z_k$. In actual practice the iterations would be performed for the equation (3.1). It is not immediately clear from (3.1) itself that the iteration procedure would work but the fact that (3.13) works to solve (3.11) can be used to prove it. For (3.1) we would set

$$(x_{k+1}^r I - hA)y_{k+1,\ell+1} = h\tilde{A}(x_{k+1})y_{k+1,\ell} + hf(x_{k+1}, y_{k+1,\ell})$$

$$+ x_{k+1}^r y_k.$$

The precise result concerning (3.13) has the following formulation

Lemma 1. Given any $\rho > 0$ *there is an* η, $\rho \geq \eta > 0$, *such that if* $\|z_k\| \leq \eta$ *and* α *and* a *are sufficiently small, then the iteration process* (3.13) *converges to a solution* z_{k+1} *of* (3.11) *which satisfies*

$$\|z_{k+1}\| \leq \rho,$$

as long as $x_{k+1} \leq a$.

We will now examine the convergence of the y_k to the $y(x_k)$ via the equations (3.11) and (3.12) satisfied by the corresponding vectors z_k, $z(x_k)$. Assuming that $\|\hat{y}_k\| \leq K$, i.e., that we have a uniform bound on the second derivatives of the components of $y(x)$ for $0 \leq x \leq a$, we have

$$\|\hat{z}_k\| \leq \|P^{-1}\|K.$$

If we assume a chosen sufficiently small we will have

$$\|z(x)\| \leq \eta/2, \quad x \in [0,a]$$

We will follow the approximations z_k only so long as

$$(3.14) \quad \|z_k - z(x_k)\| \leq \eta/2, \quad x_{k+1} \leq a.$$

This means that we shall always have

$$\|z_{k+1}\| \leq \rho,$$

by Lemma 1.

Subtracting (3.11) from (3.12) we have

$$\|z(x_{k+1}) - z_{k+1}\| \leq \frac{h}{x_{k+1}^r + h\lambda} (\|C(x_{k+1})\| + \hat{L}(\rho))$$

$$\times \|z(x_{k+1}) - z_{k+1}\| + \frac{x_{k+1}^r}{x_{k+1}^r + h\lambda} \|z(x_k) - z_k\|$$

$$+ \frac{x_{k+1}^r \, h^2}{2(x_{k+1}^r + h\lambda} \|P^{-1}\| K$$

If α, ρ and a are sufficiently small we can assume that

$$\|C(x_{k+1})\| + \hat{L}(\rho) \leq \frac{\lambda}{2}.$$

Then letting

$$e_k = \|z(x_k) - z_k\|$$

we have

$$e_{k+1} \leq \frac{\frac{h\lambda}{2}}{x_{k+1}^r + h\lambda} e_{k+1} + \frac{x_{k+1}^r}{x_{k+1}^r + h\lambda} e_k$$

$$+ \frac{x_{k+1}^r \, h^2}{2(x_{k+1}^r + h\lambda)} \|P^{-1}\| K$$

so that

$$\text{(3.15)} \quad e_{k+1} \leq \frac{x_{k+1}^r}{x_{k+1}^r + \frac{h\lambda}{2}} (e_k + \frac{h^2}{2} \|P^{-1}\| K)$$

$$\leq e_k + \frac{h^2}{2} \|P^{-1}\| K.$$

Since $e_0 = 0$ (3.15) implies

(3.16) $$e_{k+1} \leq \frac{a}{h}\frac{h^2}{2}\|P^{-1}\|K = (\frac{a}{2}\|P^{-1}\|K)h.$$

If we require that

$$h < \frac{\eta}{a\|P^{-1}\|K}$$

then (3.14) will remain true as long as $0 \leq x_{k+1} \leq a$ and thus the estimate (3.15) is also valid there. We see, therefore, that we have proved the following theorem.

Theorem 2. *For sufficiently small* a, α *the error in approximating* $y(x_k)$ *by* y_k, *where* y_k *satisfies* (3.1) *and* $x_k \leq a$, *is bounded by*

$$\|y(x_k) - y_k\| \leq (\|P\| \ \|P^{-1}\| \ \frac{aK}{2})\, h$$

and the method (3.1) *is therefore convergent.*

4. An Application

Let us consider a nonhomogeneous linear system

(4.1) $$x^r \frac{dz}{dx} = A(x)z + h(x)$$

where $r > 0$, z is an n-vector with complex components, $A(x)$ is a continuous complex $n \times n$ matrix valued function for $0 \leq x \leq x_0$ and $h(x)$ is a complex n-vector valued function continuous for $0 \leq x \leq x_0$ with $h(0) = 0$. We will assume that $A = A(0)$ can be written in the form

$$A = \begin{pmatrix} A^{11} & 0 \\ 0 & A^{22} \end{pmatrix}$$

where A^{11} is an $n_1 \times n_1$ matrix whose eigenvalues all have positive real parts and A^{22} is an $n_2 \times n_2$ matrix whose eigenvalues all have negative real parts. If A is not already in this form it may be possible to reduce it to this form by a similarity transformation. Of course, $n_1 + n_2 = n$.

Given a with $0 < a \leq x_0$ it can be shown that there is an affine subspace $S_a \subseteq R^n$ of dimension n_1 such that if $z_0 \in S_a$ then the solution $z(x)$ of (4.1) which satisfies $z(a) = z_0$ also satisfies

$$\lim_{x \to 0+} z(x) = 0.$$

In general it is not easy to identify S_a by analytical procedures. We will show, however, that the numerical techniques developed in Sections 2 and 3 enable us to determine equations for S_a whose coefficients are correct within any desired degree of accuracy.

We seek a nonsingular linear transformation

$$(4.2) \quad z = \hat{P}(x)w = \begin{pmatrix} I_{n_1} & 0 \\ P(x) & I_{n_2} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}, \quad P(0) = 0,$$

which transforms (4.1) into a system

$$(4.3)\quad x^r \frac{dw}{dx} = B(x)w + g(x)$$

where $B(x)$ has the special form

$$B(x) = \begin{pmatrix} B^{11}(x) & B^{12}(x) \\ 0 & B^{22}(x) \end{pmatrix}$$

and

$$B^{11}(0) = A^{11}, \quad B^{22}(0) = A^{22}.$$

Substituting (4.2) into (4.1) and comparing with (4.3) we see that we need

$$\hat{P}(x)\, B(x) = A(x)\, \hat{P}(x) - x^r \frac{d\hat{P}}{dx}, \quad g(x) = \hat{P}^{-1}(x)\, h(x),$$

which leads to the following form equations for the blocks of the matrices involved:

$$(4.4)\quad B^{11}(x) = A^{11}(x) + A^{12}(x)\, P(x),$$

$$B^{12}(x) = A^{12}(x),$$

$$P(x)\, B^{11}(x) = A^{21}(x) + A^{22}(x)\, P(x) - x^r \frac{dP}{dx},$$

$$P(x)\, B^{12}(x) + B^{22}(x) = A^{22}(x).$$

If we substitute the right hand side of the first of these equations for $B^{11}(x)$ in the third equation we obtain an equation

$$(4.5)\quad x^r \frac{dP}{dx} = A^{22}(x)\, P(x) - P(x)\, A^{11}(x) - P(x)\, A^{12}(x)\, P(x) + A^{21}(x).$$

The idea now is to solve (4.5) for the initial condition $P(0) = 0$ and then obtain $B^{11}(x)$, $B^{12}(x)$ and $B^{22}(x)$ from the first, second and fourth equations, respectively, in (4.4).

The matrix P may be considered a vector of dimension $n_1 n_2$ and

$$T(P) = A^{22} P - P A^{11}$$

a linear operator on $C^{n_1 n_2}$. It is known (see, e.g. [3], Chap. VIII) that, under our assumptions on A^{11} and A^{22}, all eigenvalues of T have negative real parts. Thus (4.5) is an equation of the type (1.1) and, according to Theorem 1, has a unique solution $P(x)$ defined on $[0,a]$ with $P(0) = 0$, provided a is sufficiently small. Moreover, $P(x)$ can be approximated by the numerical techniques of Sections 2 and 3.

Since

$$\hat{P}(x)^{-1} = \begin{pmatrix} I_{n_1} & 0 \\ P(x) & I_{n_2} \end{pmatrix}^{-1} = \begin{pmatrix} I_{n_1} & 0 \\ -P(x) & I_{n_2} \end{pmatrix}$$

we may rewrite (4.3) in the form

$$x^r \frac{d}{dx} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} B^{11}(x) & B^{12}(x) \\ 0 & B^{22}(x) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + \begin{pmatrix} h_1(x) \\ -P(x)h_1(x) + h_2(x) \end{pmatrix}.$$

The equation for w_2 is

$$(4.6) \qquad x^r \frac{dw_2}{dx} = B^{22}(x) w_2 - P(x) h_1(x) + h_2(x)$$

and may also be treated numerically by the methods of Sections 2 and 3 since $B^{22}(0) = A^{22}$ has eigenvalues with negative real parts only.

Note that we need the computed values of $P(x)$ to perform this computation. Preferably a common step length h should be used in treating both (4.5) and (4.6). Once $w_2(x)$ has been computed we have

$$(4.7)\quad x^r \frac{dw_1}{dx} = B^{11}(x)w_1 + B^{12}(x)w_2(x) + h_1(x).$$

Since $B^{11}(0) = A^{11}$, all of whose eigenvalues have positive real parts, it can be shown that all solutions of (4.7) satisfy

$$\lim_{x\to 0+} w_1(x) = 0.$$

This equation can be treated by a number of methods, e.g., the explicit Euler's method, starting at $x = a$ and working in the negative direction. Note that previously computed values of $w_2(x)$ are required.

Thus we see that if

$$w(a) = \begin{pmatrix} w_1(a) \\ w_2(a) \end{pmatrix},$$

where $w_1(a) \in C^{n_1}$ is arbitrary, the solution $w(x)$ of (4.3) assuming such a value at a will satisfy

$$\lim_{x\to 0+} w(x) = 0.$$

Since

$$z(a) = \hat{P}(a)w(a) = \begin{pmatrix} I_{n_1} & 0 \\ P(a) & I_{n_2} \end{pmatrix} \begin{pmatrix} w_1(a) \\ w_2(a) \end{pmatrix}$$

we see that if

$$(4.8)\quad z(a) = \begin{pmatrix} z_1(a) \\ z_2(a) \end{pmatrix}, \quad z_2(a) = \hat{P}(a)z_1(a) + w_2(a)$$

then

$$\lim_{x \to 0+} z(x) = 0.$$

Thus (4.8) provides us with the defining equation for S_a and, as we have seen, the coefficients $\hat{P}(a)$, $w_2(a)$ may be computed with whatever accuracy is required by applying the methods of Sections 2 and 3 to the differential equations (4.5) and (4.6).

5. An Example

We consider the scalar second order equation

$$(5.1)\quad \frac{d^2y}{dt^2} = t^2 y + \frac{1}{1+t}$$

and ask: what relationship should hold between $y(1)$ and $y'(1)$ in order that

$$(5.2)\quad \lim_{t \to +\infty} y(t) = 0?$$

The equivalent first order system

$$\frac{d}{dx}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ t^2 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{1+t} \end{pmatrix}$$

can be changed by the transformations

$$(5.3)\quad \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -t & t \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

$$t = \frac{1}{x}$$

into another system

$$x^3 \frac{d}{dx} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1 + \frac{1}{2} x^2 & -\frac{1}{2} x^2 \\ -\frac{1}{2} x^2 & -1 + \frac{1}{2} x^2 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} \frac{x^3}{2(x+1)} \\ \frac{-x^3}{2(x+1)} \end{pmatrix}$$

which is of the type ( ). It is clear that ( ) holds if $\dot{z}(x) \to 0$ as $x \to 0+$.

The equation for $p(x)$ is

$$x^3 \frac{dp}{dx} = \left(-1 + \frac{1}{2} x^2\right) p - p\left(1 + \frac{1}{2} x^2\right) - p\left(-\frac{1}{2} x^2\right) p - \frac{1}{2} x^2$$

or

$$(5.4)\quad x^3 \frac{dp}{dx} = -2p + \frac{1}{2} x^2 p^2 - \frac{1}{2} x^2$$

and the equation for w_2 is

$$x^3 \frac{dw_2}{dx} = \left(-1 + \frac{1}{2} x^2 - p(x)\left(-\frac{1}{2} x^2\right)\right) w_2 - p(x)\left(\frac{x^3}{2(x+1)}\right) + \left(\frac{-x^3}{2(x+1)}\right)$$

or

$$(5.5)\quad x^3 \frac{dw_2}{dx} = \left(-1 + \frac{1}{2} x^2 \left(1 + p(x)\right)\right) w_2 - \frac{x^3}{2(x+1)} \left(1 + p(x)\right).$$

We shall employ the numerical techniques of Section 3, i.e. the implicit Euler method to solve these equations. Thus we have, for (5.4),

$$p_{k+1} = \frac{x_{k+1}^3}{x_{k+1}^3 + 2h} p_k + \frac{h}{x_{k+1}^3 + 2h}\left[\frac{x_{k+1}^2}{2} p_{k+1}^2 - \frac{x_{k+1}^2}{2}\right].$$

In general one must use iteration to obtain p_{k+1} here. For (5.5), a linear equation, the implicit Euler method yields an equation which can be solved explicitly for $w_{2,k+1}$. Thus

$$w_{2,k+1} = \frac{x_{k+1}^3 w_{2,k} - \dfrac{x_{k+1}^3}{2(x_{k+1}+1)}(1 + p_{k+1})}{x_{k+1}^3 + h(1 - \frac{1}{2} x_{k+1}^2 (1 + p_{k+1}))}.$$

These equations were used, with $h = .05$, to obtain the following table.

k	x_k	p_k	$w_{2,k}$
0	.00	.000	-.000
1	.05	-.001	-.003
2	.10	-.002	-.009
3	.15	-.006	-.028
4	.20	-.010	-.061
5	.25	-.015	-.111
6	.30	-.021	-.173
7	.35	-.028	-.254
8	.40	-.035	-.349
9	.45	-.043	-.454
10	.50	-.051	-.569
11	.55	-.060	-.693
12	.60	-.070	-.826
13	.65	-.079	-.964
14	.70	-.089	-1.108
15	.75	-.098	-1.252
16	.80	-.108	-1.400
17	.85	-.118	-1.554
18	.90	-.128	-1.710
19	.95	-.138	-1.868
20	1.00	-.148	-2.025

The equation (4.8) thus becomes

$$z_2(1) = (-.148)\,z_1(1) + (-2.025)$$

whence, from (5.3)

$$\frac{y_1(1) + y_2(1)}{2} = -.148\left(\frac{y_1(1) - y_2(1)}{2}\right) - 2.025$$

so that

$$y_2(1) = -1.347\, y_1(1) - 4.754,$$

that is, for (5.1)

$$y'(1) = -1.347\, y(1) - 4.754$$

is approximately the relationship which should hold if we are to have (5.2).

References

1. Wasow, W.: "Asymptotic Expansions for Ordinary Differential Equations," Interscience Publishers, New York, 1965.

2. Henrici, P.: "Discrete Variable Methods in Ordinary Differential Equations," John Wiley & Sons, Inc., New York, 1962.

3. Gantmacher, F.R.: "The Theory of Matrices," Chelsea Publishing Company, New York, 1960.