

Series A

I. MATHEMATICA

341

ÜBER DIE INTEGRATION VON
RANDWERTAUFGABEN GEWÖHNLICHER
DIFFERENTIALGLEICHUNGEN

VON

PENTTI LAASONEN

HELSINKI 1964
SUOMALAINEN TIEDEAKATEMIA

Vorgelegt am 13. März 1964

KESKUSKIRJAPAINO
HELSINKI 1964

Über die Integration von Randwertaufgaben gewöhnlicher Differentialgleichungen

Die von gewöhnlichen Differentialgleichungen oder von Systemen solcher Gleichungen hergeleiteten Probleme pflegt man in zwei Kategorien zu klassifizieren, je nachdem die zusätzlichen Bedingungen sich auf einen einzigen Punkt der unabhängigen Veränderlichen oder auf mehrere Werte derselben beziehen. Diese Einteilung ist ja durch die vielen Verschiedenheiten begründet, welche die Probleme der beiden Kategorien hinsichtlich wichtiger Eigenschaften wie z.B. der Existenz, der Eindeutigkeit, der Stabilität Störungen gegenüber u.dgl. aufweisen.

Die gleiche Zweiteilung erstreckt sich indessen bis in die Lösungsmethodik dieser Probleme, und zwar besonders wenn es sich um angenäherte Lösungen für diskretisierte Variablenwerte handelt, wenngleich aus den genannten verschiedenen Eigenschaften keineswegs Schlüsse auf notwendige Verschiedenartigkeit der zugehörigen Lösungsmethoden möglich sind.

Eine Differentialgleichung mit Anfangswerten, d.h. ein sogenanntes Marschproblem mit diskretisierten Variablenwerten wird üblicherweise nach irgendeiner Methode gelöst, die im Grunde der Integration der zugehörigen Normalgleichung gleichkommt:

$$\begin{cases} \frac{d\bar{y}}{dx} = f(x, \bar{y}), \\ \bar{y}(x_0) = \bar{y}_0; \end{cases} \quad \bar{y}(x) = \bar{y}_0 + \int_{x_0}^x f(x, \bar{y}) dx.$$

Dies trifft sowohl für die Einzelschrittverfahren z.B. vom Runge-Kutta-Typ als auch für die charakteristischen Extrapolations- und Interpolationsmethoden mit endlichen Differenzen zu. Liegt dagegen ein sog. Juryproblem vor, welches Zusatzbedingungen in wenigstens zwei Punkten besitzt, dann ist es gebräuchlich, die Differentialgleichung einfach durch die entsprechende Differenzengleichung zu ersetzen.

Es gibt jedoch eine Möglichkeit, diese Probleme mit diskretisierten Variablenwerten im wesentlichen auf Integration zurückzuführen, selbst wenn wir die Reduktion des Randwertproblems auf wiederholte Anfangswertprobleme ausschliessen. Wir wollen ein einfaches Beispiel nehmen, das geeignet ist, das Wesen dieses Verfahrens zu erläutern.

Die Randwerte seien in zwei Punkten, x_1 und x_2 , vorgeschrieben und das Differentialgleichungssystem sei derart in Normalform ausdrückbar,

$$\begin{cases} \frac{d\bar{y}}{dx} = f(x, \bar{y}, \bar{z}), \\ \frac{d\bar{z}}{dx} = g(x, \bar{y}, \bar{z}), \end{cases} \quad \begin{cases} \bar{y}(x_1) = \bar{y}_1, \\ \bar{z}(x_2) = \bar{z}_2, \end{cases}$$

dass die Vektoren \bar{y}_1 und \bar{z}_2 aus denjenigen Unbekannten bestehen, deren Randwerte in x_1 , bzw. x_2 , vorgeschrieben sind. In integrierter Form ist das obige System daher gleichwertig mit dem System

$$\begin{cases} \bar{y}(x) = \bar{y}_1 + \int_{x_1}^x f(x, \bar{y}, \bar{z}) dx, \\ \bar{z}(x) = \bar{z}_2 + \int_{x_2}^x g(x, \bar{y}, \bar{z}) dx. \end{cases}$$

Wenn man jetzt die Integration in dem diskretisierten Variablenraum mittels Multiplikation mit einer zweckdienlichen Matrix, dem Integrator, durchführt, geht dieses System auf ein System mit $k \cdot (m + n)$ skalaren Unbekannten zurück, wenn k die gesamte Anzahl der Punkte von x ist sowie m und n die Dimensionszahlen der Vektoren \bar{y} bzw. \bar{z} sind. Ist das ursprüngliche System linear, so gilt das gleiche auch für das neue System, weshalb dieses nach den üblichen direkten Methoden gelöst werden kann.

Ein besonderer Vorteil des Verfahrens aber ist, dass das System auch eine iterative Lösung zulässt. Dies ist besonders bedeutungsvoll, wenn das ursprüngliche Differentialgleichungssystem nichtlinear ist. Auch wenn die fortgesetzte iterative Folge divergiert, ist es vorteilhaft, wie stets bei linearer Divergenz und Konvergenz, den sogenannten Delta-Prozess von Aitken anzuwenden und derart das Divergenz- oder Konvergenzzentrum schnell zu bestimmen.

Vor dem Übergang zu ein paar Beispielen seien einige spezielle Faktoren erwähnt, die beim Vergleich dieses I -Verfahrens mit der gebräuchlichen Differenzenmethode, dem D -Verfahren, von Bedeutung sind. Die Genauigkeit des D -Verfahrens kann bekannterweise dadurch verbessert werden, dass Differenzenquotienten höherer Approximation zur Ersetzung der Ableitungen angewandt werden. Eine ähnliche Verschärfung ist auch bei dem I -Verfahren möglich — sie wird ganz einfach durch die Wahl einer der gewünschten Genauigkeit angepassten Integrator-Matrix bewirkt. Ist die Differentialgleichung von höherer Ordnung, so werden auch einige Rand-

bedingungen Ableitungen höherer Ordnung enthalten, und gerade in den Randpunkten sind beim D -Verfahren die Gegenstücke höherer Approximation zu solchen Ableitungen recht kompliziert und ihre Aufstellung gestaltet sich uneinheitlich. Für das I -Verfahren hingegen stellt das Beibehalten der Genauigkeit bis zum Randpunkt überhaupt kein Problem dar.

Ein zweiter Unterschied zugunsten des I -Verfahrens ist die bereits erwähnte Möglichkeit einer zwanglosen Iteration bei der Lösung des diskretisierten Systems. Was die erreichbare Genauigkeit anbelangt, haben ausgeführte Versuche gezeigt, dass wenn das Differenzieren bzw. die Integration theoretisch gleichgradig fortgeführt wird und die Anzahl der Schritte die gleiche ist, das I -Verfahren immer kleinere Fehler als das D -Verfahren ergibt.

Als erstes sehr einfaches Beispiel sei die nichtlineare Gleichung

$$y'' = \sinh y - 2$$

mit den Randbedingungen $y(0) = y'(\frac{1}{2}) = 0$ betrachtet, die in dem Buch von Henrici »Discrete Variable Methods in Ordinary Differential Equations« nach dem D -Verfahren behandelt wird. Zerlegung in die Form

$$\begin{cases} y' = z, \\ z' = \sinh y - 2; \end{cases} \quad \begin{cases} y(0) = 0, \\ z(\frac{1}{2}) = 0, \end{cases}$$

gibt die einfachen Rekursionsformeln

$$\begin{cases} z_v(x) = \int_{\frac{1}{2}}^x (\sinh y_{v-1}(x) - 2) dx, \\ y_v(x) = \int_0^x z_v(x) dx. \end{cases}$$

Wird die Schrittlänge zu 0,1 gewählt (wie auch bei Henrici) und wird für den Multiplikator die Matrix

$$\frac{1}{240} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 10 & 16 & -2 & 0 & 0 & 0 \\ 8 & 32 & 8 & 0 & 0 & 0 \\ 9 & 27 & 27 & 9 & 0 & 0 \\ 8 & 32 & 16 & 32 & 8 & 0 \\ 8 & 32 & 17 & 27 & 27 & 9 \end{pmatrix}$$

genommen, die eine Genauigkeit zweiter Ordnung besitzt, so erhält man die sukzessiven Resultate in Tafel 1. Die Konvergenz ist aber schneller, wenn nach je zwei oder drei Rekursionen die ersten Werte von $y(x)$ für die fol-

Tafel 1

0,0	0,1	0,2	0,3	0,4	0,5
0,00000000	0,00000000	0,00000000	0,00000000	0,00000000	0,00000000
0,00000000	0,09000000	0,16000000	0,21000000	0,24000000	0,25000000
0,00000000	0,08177361	0,14441577	0,18865643	0,21499596	0,22374148
0,00000000	0,08261594	0,14601912	0,19086426	0,21759190	0,22647130
0,00000000	0,08252897	0,14585352	0,19063611	0,21732354	0,22618907
0,00000000	0,08253796	0,14587064	0,19065969	0,21735128	0,22621825
0,00000000	0,08253703	0,14586887	0,19065725	0,21734842	0,22621523
0,00000000	0,08253713	0,14586905	0,19065750	0,21734871	0,22621554
0,00000000	0,08253712	0,14586903	0,19065748	0,21734868	0,22621551
0,00000000	0,08253712	0,14586903	0,19065748	0,21734869	0,22621551

gende Wertgruppe nach dem Deltaprozess extrapoliert werden; die derart erhaltenen sukzessiven Näherungen von $y(\frac{1}{2})$ sind in den zwei letzten Spalten von Tafel 2 angeführt. Tafel 3 zeigt nebeneinander die (in sieben Stellen) korrekten Werte von y , sowie die Resultate des D -Verfahrens bzw. des I -Verfahrens. Es ist zu bemerken, dass die letztgenannten Approximationen auf Differentiation bzw. Integration von gleicher Genauigkeit basieren.

Tafel 2

0,00000000	0,00000000	0,00000000
0,25000000	0,25000000	0,25000000
0,22374148	0,22374148	0,22374148
0,02264713		0,22647130
0,22618907	0,22623737	
0,22621825	0,22621783	0,22621424
0,22621523	0,22621521	0,22621557
0,22621554		0,22621551
0,22621551	0,22621481	0,22621551
0,22621551	0,22621583	
	0,22621548	
		0,22621557
		0,22621551
		0,22621551

Tafel 3

0,0	0,0000000	0,0000000	0,0000000
0,1	0,0825292	0,0824662	0,0825371
0,2	0,1458689	0,1457580	0,1458690
0,3	0,1906572	0,1905125	0,1906575
0,4	0,2173486	0,2171837	0,2173487
0,5	0,2262154	0,2260438	0,2262155

Ändern wir die Gleichung um in

$$y'' = 20 \sinh y - 2$$

unter Beibehaltung der Randbedingungen, so wird die gewöhnliche iterative Folge divergent: erste Spalte in Tafel 4. Unter Verwendung des Delta-Prozesses aber erhält man leicht das Ergebnis: die zwei letzten Spalten der gleichen Tafel.

Tafel 4

0,00000000	0,00000000	0,00000000
0,25000000	0,25000000	0,25000000
- 0,27517034	- 0,27517034	- 0,27517034
0,80215111		0,80215111
- 1,52775864	0,08062744	
4,34741749	0,07900646	0,07789061
- 55,4148718	0,07744220	0,07964272
		0,07740242
	0,03430427	0,08164193
	0,09190969	
	0,05651644	0,07886817
		0,07881211
	0,06998633	0,07881506
	0,07634289	0,07883528
	0,08475817	
		0,07881161
		0,07882512
		0,07881179
		0,07883664

Unser letztes Beispiel ist alles andere als künstlich, indem es seinen Ursprung in der Festigkeitsuntersuchung eines gewölbten, zylindersymmetrischen Gusseisendeckels hat. Die Randbedingungen sind auch komplizierter und bedingen daher die Benützung zweckdienlicher Umwege.

Das Problem enthält insgesamt neun Unbekannte, die linearen Verschiebungskomponenten u und v , drei dimensionslose Hilfsgrößen ε , β und ω , sowie drei Kraftkomponenten N_r , N_t , Q_r und ein Biegemoment M_r . Alle diese Unbekannten sowie gewisse bekannte Größen ϱ , α , A , D und p sind Funktionen der unabhängigen Veränderlichen r . Das System der neun Gleichungen zwischen diesen Funktionen ist linear und als Differentialsystem der sechsten Ordnung:

$$(1) \quad \frac{du}{dr} = \varepsilon + \frac{v}{\varrho}$$

$$(2) \quad \frac{dv}{dr} = \beta - \frac{u}{\varrho}$$

$$(3) \quad \frac{d\beta}{dr} = \frac{\omega}{\varrho}$$

$$(4) \quad \frac{d}{dr} (rN_r - \alpha rQ_r) = N_t$$

$$(5) \quad \frac{d}{dr} (rQ_r + \alpha rN_r) = -pr$$

$$(6) \quad \frac{d}{dr} (rM_r) = \nu M_r + rQ_r$$

$$(7) \quad N_r = A \left[\frac{\nu}{r} (u - \alpha v) + \varepsilon \right] - D\omega$$

$$(8) \quad N_t = A \left[\frac{1}{r} (u - \alpha v) + \nu\varepsilon \right] - \nu D\omega$$

$$(9) \quad M_r = -D \frac{\omega}{\varrho}$$

(ν ist hierbei eine numerische Konstante). Die Randbedingungen beziehen sich auf zwei Randwerte r_1 und r_2 der Variablen, und zwar in folgender Weise:

$$\begin{cases} u = 0, \\ \beta = 0, \\ Q_r + \alpha N_r = \frac{1}{2} rp; \end{cases} \quad \text{für } r = r_1$$

$$\begin{cases} \nu = 0, \\ N_r - \alpha Q_r = k_1 u - k_2 \beta = 0, \\ M_r + k_3 u - k_4 \beta = 0; \end{cases} \quad \text{für } r = r_2$$

(k_1, k_2, k_3, k_4 sind gewisse Konstanten). Tafel 5 enthält die vorgeschriebenen Funktionen sowie die Werte der Konstanten.

Tafel 5

r	$1/\varrho$	α	A	D
cm	cm ⁻¹		kp/cm	kpcm
	10 ⁻³	10 ⁻³	10 ⁴	10 ⁵
20	239	-218	1320	1584
30	130	-2	858	435
40	97	226	649	188
50	-31	356	473	73
60	-260	325	550	115
$\nu = 0,3$ $k_1 = 1891000 \text{ kp/cm}^2$ $k_2 = 397000 \text{ kp/cm}$ $k_3 = 18020000 \text{ kp/cm}$ $k_4 = 1890000 \text{ kp}$				

Umordnung der integrierten Gleichungen (1)–(6) liefert unter Beachtung der Randbedingungen das Rekursionssystem

$$(10) \quad \beta = \int_{r_1} \frac{1}{\varrho} \omega dr,$$

$$(11) \quad v = \int_{r_2} \left(\beta - \frac{1}{\varrho} u \right) dr,$$

$$(12) \quad u = \int_{r_1} \left(\varepsilon + \frac{1}{\varrho} v \right) dr,$$

$$(13) \quad r[Q_r + \alpha N_r] = - \int_{r_1} pr dr - \frac{1}{2} (pr)_{r_1} = r q,$$

$$(14) \quad r[N_r - \alpha Q_r] = \int_r N_i dr + r_2(-k_1 u + k_2 \beta)_{r_2},$$

$$(15) \quad rM_r = \int_{r_2} (\nu M_r + r Q_r) dr + r_2(-k_3 u + k_4 \beta)_{r_2},$$

wo q eine Hilfsgrösse mit dem Ausdruck

$$q = -\frac{1}{r} \left[\int_{r_1} pr \, dr + \frac{1}{2} (pr)_{r_1} \right]$$

und mithin eine bekannte Funktion ist.

Obwohl wir in (7)–(9), (10)–(15) genau die zur direkten Bestimmung der Unbekannten benötigten neun Gleichungen besitzen, bestimmt man ihre Werte in den Teilungspunkten der Randpunktstrecke vorteilhafter folgendermassen. Mit Hilfe der Gleichungen (10), (11), (7), (8), (9) und (13) kann man sukzessiv alle anderen Grössen in ω , u und ε ausdrücken. Die Zahl der Teilstrecken zwischen r_1 und r_2 sei $n - 1$, also ist $n - 2$ die Zahl der inneren Teilungspunkte und n die Ordnung eines jeden der unbekanntenen Vektoren. Als Hauptunbekannte wird jetzt der aus den Vektoren ω , u und ε zusammengesetzte Vektor der Ordnung $3n$ gewählt. Die Vektoren aller anderen Unbekannten ausser Q_r können lediglich durch Multiplikationen mittels dieses Grundvektors bestimmt werden, und zwar sind die Faktormatrizen zur Herbeiführung hiervon, die von der Ordnung $n \times 3n$ sind, sogleich durch die genannten Gleichungen (10), (11), (7), (8), (9) definiert.

Um dies nachzuweisen, seien die verschiedenen in dem Vorgang enthaltenen Operationen betrachtet. Zuerst sei bemerkt, dass natürlich auch den Grössen ω , u und ε gewisse Faktormatrizen dieser Art zugehören, nämlich

$$(I \ 0 \ 0), \quad (0 \ I \ 0), \quad (0 \ 0 \ I),$$

wo I und 0 die Einheitsmatrix bzw. Nullmatrix der Ordnung $n \times n$ ist.

Der Multiplikation mit einer von r abhängigen Grösse, z.B. $\frac{1}{\varrho}$, entspricht ganz einfach die Multiplikation mit der zugehörigen Diagonalmatrix; der Integration entspricht gleicherweise die Multiplikation mit der zugehörigen Integratormatrix usf. Wenn Q_r aus den noch übrig bleibenden Gleichungen eliminiert wird, erhält man die Gleichungen

$$\begin{cases} u - \int_{r_1} \left(\varepsilon + \frac{1}{\varrho} v \right) dr = 0, \\ rN_r + \alpha^2 rN_r - \int_{r_2} N_t \, dr + r(k_1 u - k_2 \beta)_{r_2} = \alpha q r, \\ rM_r - \int_{r_2} (vM_r - \alpha rN_r) \, dr + r_2(k_3 u - k_4 \beta)_{r_2} = \int_{r_2} q r \, dr. \end{cases}$$

In allen diesen Gleichungen sind die rechten Seiten bekannte Spaltenvektoren der Ordnung n , weil alle Glieder auf den linken Seiten von eben

beschriebener Art sind, d.h. Produkte einer $(n \times 3n)$ -Matrix und des Grundvektors. Dies gilt auch für die Glieder von der Form

$$r_2(k_1u - k_2\beta)_{r_2},$$

denn diese können aus dem gewöhnlichen unbestimmten Glied

$$r(k_1u - k_2\beta)$$

durch Multiplikation mit einer solchen quadratischen n -Matrix erhalten werden, deren einzige nichtverschwindende Elemente in der letzten Spalte stehen, die allerdings aus blossen Einer-Elementen besteht.

Die drei Gleichungen zusammen bilden nun eine Matrixgleichung der Ordnung $3n$, die zur Lösung des Grundvektors angewandt werden kann, wonach sich alle einzelnen Unbekannten durch Multiplikation mit den zugehörigen Faktormatrizen ergeben.

Die iterative Lösungsmethode gestaltet sich völlig verschieden. Das Gleichungssystem wird zu einem solchen Zyklus umgearbeitet, in dem die linken Seiten aus den unbekanntem Grössen bestehen und zwar eine jede derselben genau einmal vorkommt. Die rechten Seiten können sowohl von den bekannten als auch von den unbekanntem Grössen abhängig sein. Z.B. ist die folgende Reihe möglich:

$$\beta = \int_{r_1}^* \frac{1}{\varrho} \omega dr.$$

$$v = \int_{r_2}^* \left(\beta - \frac{1}{\varrho} u \right) dr.$$

$$u = \int_{r_1}^* \left(\varepsilon + \frac{1}{\varrho} v \right) dr,$$

$$\varepsilon = \frac{D}{A} \omega + \frac{1}{A} N_r - \frac{r}{r} (u - \alpha v),$$

$$N_t = \nu N_r + \frac{1 - \nu^2}{r} A(u - \alpha v),$$

$$Q_r = q - \alpha N_r,$$

$$N_r = \alpha Q_r + \frac{1}{r} \left[\int_{r_2}^* N_t dr + r_2(-k_1u + k_2\beta)_{r_2} \right],$$

$$M_r = \frac{1}{r} \left[\int_{r_2}^* (\nu M_r + r Q_r) dr + r_2(-k_3u + k_4\beta)_{r_2} \right],$$

$$\omega = -\frac{\varrho}{D} M_r.$$

Die Iteration des Systems kann von beliebigen Vektoren ausgehen, denn der Delta-Prozess liefert auf alle Fälle den Fixpunkt, d.h. die richtige Lösung. Dabei ist jedoch zu bemerken, dass das erste Glied in jeder Tripelgruppe erst nach wenigstens einmaligem Durchlaufen des Zyklus ermittelt werden darf. Die Rechnungen sind jetzt weniger umfangreich als im Fall der direkten Lösung, da man es nunmehr nur mit Vektoren von der Ordnung n anstatt mit Matrizen der Ordnung $n \times 3n$ zu tun hat. In den Versuchen hat sich die iterative Methode als wesentlich schneller als das direkte Verfahren erwiesen. Schon eine ganz grobe Abschätzung der benötigten Rechenzeit macht diesen Unterschied handgreiflich.

Die folgende Tafel enthält sowohl die bekannten als auch die auf beide Weisen bestimmten unbekannt Funktionen, wenn $n = 5$ ist. Verdoppelung von n bewirkte eine Änderung der Resultate höchstens um wenige Prozente.

Tafel 6

r	u	v	ε	β	ω	N_r	N_t	M_r	Q_r
cm	cm	cm				kp/cm	kp/cm	kp	kp/cm
	10^{-5}	10^{-5}	10^{-7}	10^{-6}	10^{-5}				10^{-1}
20	0	2612	-2227	0	-182	-1731	2901	-4831	3275
30	394	2415	-1221	-299	-124	-673	836	-1319	-736
40	784	1831	-849	-614	-169	-359	438	-609	-1812
50	879	939	-805	-924	-482	-220	418	-204	-2034
60	816	0	-639	-937	-338	-120	645	539	-1889

Infolge des einfachen Aufbaus des Programms dürfte das I -Verfahren besonders vorteilhaft sein, wenn es auf eine schnelle, obwohl vielleicht grobe Approximation bei einem komplizierten System von hoher Ordnung ankommt. Wenn die Lösung durch Iterationen gewonnen wird, kann eine dichtere Einteilung der Variablen, dabei aber Verwendung eines möglichst einfachen Integrators, etwa der Trapezformel entsprechend, lohnend sein indem dann die Multiplikationen meistens durch Additionen ersetzt werden können. Die von Stiefel empfohlene Technik, die bei sukzessiven Integrationen mit verdoppelten Punktmengen fortschreitet, kann zur Verbesserung der Genauigkeit herangezogen werden.