

Man wählt  $b_{21}$  als freien Parameter, z.B.  $b_{21} = 0$  und erhält

$$\begin{bmatrix} 1 & 1 \\ 0 & 2/3 \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{31} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{array}{l} b_{11} = 1 \\ b_{31} = 0 \end{array}$$

Insgesamt ergibt sich

$$b_1(\theta) = \theta - \frac{3}{4}\theta^2, \quad b_2(\theta) \equiv 0, \quad b_3(\theta) = \frac{3}{4}\theta^2.$$

Die im Beispiel betrachtete dense output Formel ist nach Konstruktion stetig; ist sie auch stetig differenzierbar?

$$\mathbf{y}'_{\theta} = \frac{d}{d\theta} \mathbf{y}_{\theta} = h \sum_{i=1}^{s^*} b'_i(\theta) \mathbf{Y}'_i$$

$$\lim_{\theta \rightarrow 0^+} \mathbf{y}'_{\theta} = h \sum_{i=1}^2 b'_i(0) \mathbf{Y}'_i = h \mathbf{Y}'_1 = h \mathbf{f}(x_0, \mathbf{y}_0) \quad \text{ist ok}$$

$$\lim_{\theta \rightarrow 1^-} \mathbf{y}'(\theta) = h \sum_{i=1}^3 b'_i(1) \mathbf{Y}'_i = -\frac{h}{2} \mathbf{Y}'_1 + \frac{3h}{2} \mathbf{Y}'_3 \underset{\text{i.A.}}{\neq} h \mathbf{f}(x_1, \mathbf{y}_1)$$

Diese zusätzliche Forderung könnte in die Definition von  $b_j(\theta)$  eingebaut werden, beispielsweise über *Hermite Interpolation*.

Hermite Interpolation liefert für niedrigere Ordnung ( $p \leq 4$ ) dense output Formeln, die stetig differenzierbar sind. Man erhält die Formeln wie folgt: Man definiert  $\mathbf{p}(\theta)$  als Interpolationspolynom vom Grad 3 mit

$$\begin{array}{lll} \mathbf{p}(0) & = & \mathbf{y}_0, & \mathbf{p}'(0) & = & h \mathbf{f}(x_0, \mathbf{y}_0) & = & h \mathbf{f}_0 \\ \mathbf{p}(1) & = & \mathbf{y}_1, & \mathbf{p}'(1) & = & h \mathbf{f}(x_1, \mathbf{y}_1) & = & h \mathbf{f}_1 \end{array}$$

Das ergibt die dense output Formel

$$\mathbf{y}_{\theta} = \mathbf{p}(\theta) = (1 - \theta) \mathbf{y}_0 + \theta \mathbf{y}_1 + \theta(\theta - 1) \left( (1 - 2\theta)(\mathbf{y}_1 - \mathbf{y}_0) + (\theta - 1) h \mathbf{f}_0 + \theta h \mathbf{f}_1 \right).$$

*Beweis.* Numerische Mathematik 1. □

## 4.7 Steife Differentialgleichungen

Unter steifen Differentialgleichungen (engl. *stiff*: schwierig) versteht man Probleme, bei denen explizite Verfahren sehr schlecht funktionieren (sehr kleine Schritte bei automatischer Schrittweitenwahl bzw. *overflow* bei konstanter größerer Schrittweite). Dieses Verhalten wurde in den 50er Jahren bei gewissen Differentialgleichungen der Chemie (Reaktionskinetik), bei Problemen mit Diffusion (Wärmeleitung), bei Simulationen von elektrischen Schaltkreisen, u.a.m. entdeckt. Grund dafür ist mangelnde Stabilität des numerischen Verfahrens in speziellen Situationen (insbesondere in der Nähe exponentiell attraktiver Gleichgewichtspunkte).

Um dieses Verhalten besser zu verstehen, linearisieren wir die Lösung der autonomen Differentialgleichung

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{y}_0$$

um den stationären Gleichgewichtspunkt  $\mathbf{y}^*$  (in anderen Worten  $\mathbf{f}(\mathbf{y}^*) = \mathbf{0}$ ). Wegen

$$\mathbf{f}(\mathbf{y}(x)) = \underbrace{\mathbf{f}(\mathbf{y}^*)}_{=\mathbf{0}} + \underbrace{\frac{\partial \mathbf{f}}{\partial \mathbf{y}}(\mathbf{y}^*)}_{:=\mathbf{L}}(\mathbf{y}(x) - \mathbf{y}^*) + \dots$$

erhält man

$$(\mathbf{y}(x) - \mathbf{y}^*)' = \mathbf{y}'(x) = \mathbf{f}(\mathbf{y}(x)) = \mathbf{L}(\mathbf{y}(x) - \mathbf{y}^*) + \dots$$

Wir setzen nun  $\mathbf{z}(x) = \mathbf{y}(x) - \mathbf{y}^*$  und erhalten *die um  $\mathbf{y} = \mathbf{y}^*$  linearisierte Gleichung*

$$\mathbf{z}' = \mathbf{L}\mathbf{z} + \mathbf{g}(\mathbf{z}), \quad \mathbf{g}(\mathbf{z}) = \mathcal{O}(\|\mathbf{z}\|^2) \text{ für } \mathbf{z} \rightarrow \mathbf{0}.$$

Im Folgenden betrachten wir nur den linearen Teil der Gleichung, also  $\mathbf{u}' = \mathbf{L}\mathbf{u}$ .

Falls  $\mathbf{L}$  diagonalisierbar ist (was generisch der Fall ist), also

$$\mathbf{T}^{-1}\mathbf{L}\mathbf{T} = \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_m),$$

erhält man nach Transformation  $\mathbf{u}(x) = \mathbf{T}\mathbf{w}(x)$  das System

$$\mathbf{w}' = (\mathbf{T}^{-1}\mathbf{u})' = \mathbf{T}^{-1}\mathbf{L}\mathbf{u} = \mathbf{T}^{-1}\mathbf{L}\mathbf{T}\mathbf{T}^{-1}\mathbf{u} = \mathbf{\Lambda}\mathbf{w}$$

beziehungsweise

$$w_j = \lambda_j w_j, \quad j = 1, \dots, m.$$

Wegen der vorausgesetzten Stabilität muss man  $\text{Re } \lambda_j \leq 0$  annehmen. Es ist deshalb angezeigt, die folgende *Testgleichung* zu betrachten.

$$y' = \lambda y, \quad y(x_0) = y_0, \quad \text{Re } \lambda \leq 0$$

(Testgleichung von Dahlquist). Die exakte Lösung dieser Gleichung lautet

$$y(x) = e^{\lambda(x-x_0)} y_0$$

und hat folgende Eigenschaften:

- (i) Die Lösung ist durch den Anfangswert beschränkt:  $|y(x)| \leq |y_0|$ ,  $x \geq x_0$ .
- (ii) Für  $\text{Re } \lambda < 0$  gilt zusätzlich:  $|y(x)| \rightarrow 0$  für  $x \rightarrow \infty$ .

Das numerische Verfahren sollte zumindest ähnliche Eigenschaften besitzen.

Wir betrachten zunächst das explizite Eulerverfahren

$$\begin{aligned} y_1 &= y_0 + hf(y_0) = y_0 + h\lambda y_0 = (1 + h\lambda)y_0, \\ y_2 &= y_1 + hf(y_1) = (1 + h\lambda)y_1 = (1 + h\lambda)^2 y_0, \\ &\dots \\ y_n &= (1 + h\lambda)^n y_0. \end{aligned}$$

Damit die numerische Lösung durch den Anfangswert beschränkt bleibt, muss die Bedingung

$$|1 + h\lambda| \leq 1$$

gelten. Diese Bedingung sagt, dass  $z = h\lambda$  in der Kreisscheibe  $B_1(-1) \subset \mathbb{C}$  mit Mittelpunkt  $-1$  und Radius  $1$  liegen muss.

Für  $|\lambda| \gg 0$  muss die Schrittweite deshalb sehr klein gewählt werden, beispielsweise erfordert  $\lambda = -10^6$  die Wahl  $0 < h \leq 2 \cdot 10^{-6}$ , sonst ist das Verfahren instabil. Diese einschneidende Stabilitätsbedingung ist Schuld am schlechten Abschneiden expliziter Verfahren bei steifen Differentialgleichungen.

Im Gegensatz zum expliziten Eulerverfahren hat das *implizite Eulerverfahren*

$$y_1 = y_0 + hf(y_1)$$

viel bessere Stabilitätseigenschaften. Angewandt auf die Testgleichung  $y' = \lambda y$  ergibt sich

$$\begin{aligned} y_1 &= y_0 + hf(y_1), & y_1 &= \frac{1}{1 - h\lambda} y_0, \\ y_2 &= \left( \frac{1}{1 - h\lambda} \right)^2 y_0, \\ &\dots \\ y_n &= \left( \frac{1}{1 - h\lambda} \right)^n y_0. \end{aligned}$$

Das Verfahren ist somit immer stabil, da  $|1 - h\lambda| \geq 1$  für alle  $h > 0$  und  $\operatorname{Re} \lambda \leq 0$  gilt.

## 4.8 A-Stabilität

Wir betrachten die *Testgleichung*

$$y' = \lambda y, \quad y(x_0) = y_0, \quad \operatorname{Re} \lambda \leq 0,$$

deren exakte Lösung für  $x \geq x_0$  durch den Anfangswert beschränkt bleibt. Wir untersuchen, ob die numerische Lösung (berechnet mit konstanter Schrittweite  $h > 0$ ) dieselbe Eigenschaft besitzt.

Wendet man zur Lösung der obigen Gleichung ein explizites Runge-Kutta-Verfahren an, so erhält man

$$\begin{aligned} Y_i &= h_0 + h \sum_{j=1}^{i-1} a_{ij} Y_j', & Y_j' &= \lambda Y_j, & i &= 1, \dots, s, \\ y_1 &= y_0 + h \sum_{j=1}^s b_j Y_j'. \end{aligned}$$

Wir verwenden im folgenden die Notation:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ a_{21} & 0 & \cdots & \cdots & 0 \\ a_{31} & a_{32} & \ddots & & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{s,s-1} & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_s \end{bmatrix}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^s$$

und

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_s \end{bmatrix}, \quad \mathbf{Y}' = \begin{bmatrix} Y_1' \\ \vdots \\ Y_s' \end{bmatrix}, \quad z = h\lambda \in \mathbb{C}.$$

Damit lässt sich das Runge-Kutta-Verfahren kompakt schreiben als

$$\begin{aligned} \mathbf{Y}' &= \lambda \mathbf{Y}, \\ \mathbf{Y} &= \mathbf{1}y_0 + h\mathbf{A}\mathbf{Y}' = \mathbf{1}y_0 + z\mathbf{A}\mathbf{Y}, \\ y_1 &= y_0 + h\mathbf{b}^\top \mathbf{Y}' = y_0 + z\mathbf{b}^\top \mathbf{Y}. \end{aligned}$$

Daraus folgt

$$\mathbf{Y} = (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{1}y_0$$

und schließlich

$$y_1 = (1 + z\mathbf{b}^\top (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{1}) y_0.$$

Das führt zu folgender Definition.

**Definition 4.13** Gegeben Sei ein explizites Runge-Kutta-Verfahren mit Koeffizienten  $a_{ij}$  und  $b_j$ . Die Funktion

$$R(z) = 1 + z\mathbf{b}^\top (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{1}$$

heißt *Stabilitätsfunktion* des Runge-Kutta-Verfahrens.

**Satz 4.14** Für eine  $s$ -stufiges explizites Runge-Kutta-Verfahren ist die Stabilitätsfunktion ein Polynom vom Grad kleiner gleich  $s$ .

*Beweis:* Wegen  $\mathbf{A}^s = \mathbf{0}$  gilt für alle  $z \in \mathbb{C}$

$$(\mathbf{I} - z\mathbf{A})^{-1} = \sum_{j=0}^{\infty} z^j \mathbf{A}^j = \sum_{j=0}^{s-1} z^j \mathbf{A}^j.$$

Damit ist

$$R(z) = 1 + z\mathbf{b}^\top (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{1} = 1 + \mathbf{b}^\top \mathbf{1}z + \mathbf{b}^\top \mathbf{A}\mathbf{1}z^2 + \dots + \mathbf{b}^\top \mathbf{A}^{s-1} \mathbf{1}z^s$$

ein Polynom vom Grad kleiner gleich  $s$ . □

**Bemerkung 4.15** Für ein Runge-Kutta-Verfahren der Ordnung  $p$  gilt

$$\mathbf{b}^\top \mathbf{A}^\ell \mathbf{1} = \frac{1}{(\ell + 1)!}, \quad 0 \leq \ell \leq p - 1.$$

Das sind gerade die Ordnungsbedingungen der *langen* Bäume. Insbesondere gilt daher für  $p = s$  die Darstellung

$$R(z) = 1 + z + \frac{z^2}{2} + \dots + \frac{z^p}{p!},$$

unabhängig von den Koeffizienten des Verfahrens.

**Definition 4.16** Gegeben sei ein Runge-Kutta-Verfahren mit Stabilitätsfunktion  $R(z)$ . Die Menge

$$S = \{z \in \mathbb{C} \mid |R(z)| \leq 1\}$$

heißt *Stabilitätsbereich* des Runge-Kutta-Verfahrens (bzw. der Stabilitätsfunktion).

**Satz 4.17** Für explizite Runge-Kutta-Verfahren ist der Stabilitätsbereich beschränkt als Teilmenge von  $\mathbb{C}$ .

*Beweis:* Für explizite Runge-Kutta-Verfahren ist  $R(z)$  ein Polynom, folglich  $|R(z)| \rightarrow \infty$  für  $|z| \rightarrow \infty$ .  $\square$

**Beispiel 4.18** Der Stabilitätsbereich des expliziten Eulerverfahrens ist die Kreisscheibe  $|1 + z| \leq 1$  mit Mittelpunkt  $-1$  und Radius  $1$ .

## 4.9 Implizite Runge-Kutta-Verfahren

Im folgenden betrachten wir wieder die Anfangswertaufgabe

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{y}_0 \in \mathbb{R}^m.$$

**Definition 4.19** Ein implizites Runge-Kutta-Verfahren (kurz: IRK-Verfahren) ist durch folgende Formeln definiert:

$$\begin{aligned} \mathbf{Y}'_i &= \mathbf{f}(x_0 + c_i h, \mathbf{Y}_i) \\ \mathbf{Y}_i &= \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{Y}'_j, \quad i = 1, \dots, s, \\ \mathbf{y}_1 &= \mathbf{y}_0 + h \sum_{j=1}^s b_j \mathbf{Y}'_j. \end{aligned}$$

Die Größen  $\mathbf{Y}_i$  heißen *Stufen* des Verfahrens, die natürliche Zahl  $s$  heißt *Stufenzahl*. Die reellen Koeffizienten  $a_{ij}$ ,  $b_j$  und  $c_i$  gibt man i.A. im Butcher-Tableau an. Zu gegebener Schrittweite  $h$  ist  $\mathbf{y}_1$  die durch das IRK-Verfahren definierte Approximation an  $\mathbf{y}(x_0 + h)$ .

Das Verfahren heißt:

- (a) *diagonal* implizit (DIRK), falls  $a_{ij} = 0$  für alle  $j > i$ .
- (b) *einfach* (singly) *diagonal* implizit (SDIRK), falls zusätzlich  $a_{ii} = a_{11}$  für alle  $i$ .
- (c) *explizit* (ERK), falls  $a_{ij} = 0$  für alle  $j \geq i$ .

Offenbar gelten die Inklusionen

$$\text{ERK} \subset \text{SDIRK} \subset \text{DIRK} \subset \text{IRK}.$$

Um die Stufen  $(\mathbf{Y}_i)_{i=1}^s$  zu berechnen, muss in jedem Schritt ein (nichtlineares) Gleichungssystem der Größe  $ms$  gelöst werden. Für  $h$  klein genug ist das mit Fixpunktiteration möglich (da  $\mathbf{f}$  Lipschitz bezüglich der zweiten Komponente ist).

**Beispiel 4.20** (implizites Eulerverfahren)

$$\begin{array}{c|c} 0 & 1 \\ \hline & 1 \end{array}$$

Das Verfahren hat eine Stufe ( $s = 1$ ) und lautet

$$\mathbf{Y}'_1 = \mathbf{f}(x_0 + h, \mathbf{Y}_1), \quad \mathbf{Y}_1 = \mathbf{y}_0 + h\mathbf{Y}'_1 = \mathbf{y}_1 \quad \Leftrightarrow \quad \mathbf{y}_1 = \mathbf{y}_0 + h\mathbf{f}(x_0 + h, \mathbf{y}_1).$$

**Beispiel 4.21** (Trapezregel)

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Das Verfahren hat zwei Stufen und lautet

$$\mathbf{Y}'_1 = \mathbf{f}(x_0, \mathbf{Y}_1), \quad \mathbf{Y}_1 = \mathbf{y}_0, \quad \mathbf{Y}'_2 = \mathbf{f}(x_0 + h, \mathbf{Y}_2), \quad \mathbf{Y}_2 = \mathbf{y}_0 + \frac{h}{2}(\mathbf{Y}'_1 + \mathbf{Y}'_2)$$

und

$$\mathbf{y}_1 = \mathbf{y}_0 + \frac{h}{2}(\mathbf{Y}'_1 + \mathbf{Y}'_2) = \mathbf{Y}_2,$$

also

$$\mathbf{y}_1 = \mathbf{y}_0 + \frac{h}{2}(\mathbf{f}(x_0, \mathbf{y}_0) + \mathbf{f}(x_0 + h, \mathbf{y}_1)),$$

was auch den Namen des Verfahrens erklärt.

**Bemerkung 4.22** Wie schon bei den ERK-Verfahren setzen wir stets die folgenden vereinfachenden Annahmen voraus:

$$\sum_{j=1}^s b_j = 1, \quad \sum_{j=1}^s a_{ij} = c_i, \quad i = 1, \dots, s.$$

Damit sind die Koeffizienten  $c_i$  des Verfahrens definiert.

**Bemerkung 4.23** IRK sind Einschrittverfahren und erfüllen die Bedingung (ii) von Satz 4.7. Der Konvergenzbeweis von IRK -Verfahren verläuft daher wie für ERK-Verfahren. Auch die Bedingungsgleichungen sind dieselben, wobei jetzt aber alle Summen bis  $s$  laufen, also

$p$	Bedingungen
1	$\sum_{i=1}^s b_i = 1$
2	$\sum_{i=1}^s b_i c_i = \frac{1}{2}$
3	$\sum_{i=1}^s b_i c_i^2 = \frac{1}{3}$ $\sum_{i,j=1}^s b_i a_{ij} c_j = \frac{1}{6}$

Um die A-Stabilität von IRK-Verfahren zu untersuchen, wenden wir das Verfahren auf die skalare Testgleichung

$$y' = \lambda y, \quad \operatorname{Re} \lambda \leq 0$$

an. Wir verwenden wieder die Bezeichnungen

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1s} \\ \vdots & & \vdots \\ a_{s1} & \cdots & a_{ss} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_s \end{bmatrix}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^s$$

und

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_s \end{bmatrix} = \mathbf{1}y_0 + z\mathbf{A}\mathbf{Y}, \quad z = h\lambda.$$

Falls  $(\mathbf{I} - z\mathbf{A})$  invertierbar ist, d.h.  $\frac{1}{z}$  kein Eigenwert von  $\mathbf{A}$  ist, gilt

$$\mathbf{y}_1 = \mathbf{y}_0 + z\mathbf{b}^\top \mathbf{Y} = \underbrace{(1 + z\mathbf{b}^\top (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{1})}_{:= R(z)} \mathbf{y}_0.$$

Die Funktion  $R(z)$  heißt *Stabilitätsfunktion* des IRK-Verfahrens.

**Bemerkung 4.24** Für die Stabilitätsfunktion eines IRK-Verfahrens gilt

$$R(z) = \frac{P(z)}{Q(z)},$$

wobei  $P$  und  $Q$  Polynome vom Grad kleiner gleich  $s$  sind und  $Q(z) = \det(\mathbf{I} - z\mathbf{A})$ . Zum Beweis verwendet man die Cramersche Regel (Übung).

**Definition 4.25** Gegeben sei ein IRK-Verfahren mit Stabilitätsfunktion  $R(z)$ . Die Menge

$$S = \{z \in \mathbb{C} \mid (\mathbf{I} - z\mathbf{A}) \text{ invertierbar und } |R(z)| \leq 1\}$$

heißt *Stabilitätsbereich* des Verfahrens. Das Verfahren heißt *A-stabil*, falls  $\mathbb{C}^- \subset S$ .

**Beispiel 4.26** Das implizite Eulerverfahren besitzt die Stabilitätsfunktion

$$R(z) = \frac{1}{1 - z}.$$

Das Verfahren ist somit A-stabil.

**Beispiel 4.27** Die Trapezregel besitzt die Stabilitätsfunktion

$$R(z) = \frac{1 + z/2}{1 - z/2}.$$

Sie ist A-stabil (man verwende das Maximumprinzip auf  $\mathbb{C}^-$ ).

**Satz 4.28** Für die Stabilitätsfunktion  $R(z)$  eines Runge-Kutta-Verfahrens der Ordnung  $p$  gilt

$$R(z) - e^z = \mathcal{O}(z^{p+1}), \quad z \rightarrow 0.$$

*Beweis:* Man wendet ein IRK-Verfahren mit Schrittweite  $h > 0$  auf die Testgleichung  $y' = \lambda y$  mit  $y(0) = 1$  an und erhält

$$y_1 = R(z), \quad y(h) = e^z.$$

Die Aussage folgt mit der Definition der Ordnung des Verfahrens.

**Definition 4.29** Ein  $s$ -stufiges IRK-Verfahren heißt *stiffly accurate*, falls für alle  $i$  gilt:  $a_{si} = b_i$ .

**Satz 4.30** Gegeben sei ein IRK-Verfahren mit invertierbarer Koeffizientenmatrix  $\mathbf{A}$ .

Dann gilt:

(i)  $R(\infty) = 1 - \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{1}$ .

(ii) Falls das Verfahren *stiffly accurate* ist, folgt  $R(\infty) = 0$ .

Insbesondere gibt es im Fall (ii) also eine Umgebung  $\mathcal{U}$  von  $\infty$  mit  $\mathcal{U} \subset S$ .

*Beweis:* Da  $\mathbf{A}$  invertierbar ist, gilt

$$R(z) = 1 + z \mathbf{b}^\top (\mathbf{I} - z \mathbf{A})^{-1} \mathbf{1} = 1 + \mathbf{b}^\top \left( \frac{1}{z} \mathbf{I} - \mathbf{A} \right)^{-1} \mathbf{1} \rightarrow 1 - \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{1}$$

für  $z \rightarrow \infty$  und damit (i). Zum Beweis von (ii) überlegt man sich leicht, dass

$$\mathbf{b}^\top \mathbf{A}^{-1} = [0, \dots, 0, 1]$$

gilt. Das Resultat folgt somit aus (i). □

**Beispiel 4.31** Die Trapezregel ist *stiffly accurate*, ihre Koeffizientenmatrix ist aber nicht invertierbar und es gilt  $R(\infty) = -1$ .

Im Folgenden verwenden wir zusätzliche vereinfachende Bedingungen zur Konstruktion von Runge-Kutta-Verfahren. Die Bedingungen

$$B(p) : \sum_{i=1}^s b_i c_i^{\ell-1} = \frac{1}{\ell}, \quad 1 \leq \ell \leq p$$

sind notwendig für Verfahren der Ordnung  $p$  (die Bedingungsgleichungen der sg. Büsche). Wenn diese Bedingungsgleichungen erfüllt sind, dann hat die  $s$ -stufige Quadraturformel  $(b_i, c_i)_{i=1, \dots, s}$  Ordnung  $p$ .

Die Bedingungen

$$C(q) : \sum_{j=1}^s a_{ij} c_j^{\ell-1} = \frac{c_i^\ell}{\ell}, \quad 1 \leq i \leq s, 1 \leq \ell \leq q$$

heißen Stufenordnungsbedingungen der Ordnung  $q$ . Wenn diese Bedingungen erfüllt sind, dann hat die  $s$ -stufige Quadraturformel  $(a_{ij}, c_j)_{j=1, \dots, s}$  Ordnung  $q$  auf dem Intervall  $[0, c_i]$ .

**Definition 4.32** Ein Runge-Kutta-Verfahren hat *Stufenordnung*  $q$ , wenn  $C(q)$  gilt.



**Satz 4.33** Wenn ein Runge-Kutta-Verfahren Stufenordnung  $q$  hat, dann gilt für alle  $i$

$$\mathbf{Y}_i - \mathbf{y}(x_0 + c_i h) = \mathcal{O}(h^{q+1}),$$

wobei die Stufen  $\mathbf{Y}_i$  für  $\mathbf{y}_0 = \mathbf{y}(x_0)$  berechnet wurden.

*Beweis:* Die Defekte erfüllen

$$\mathbf{y}(x_0 + c_i h) = \mathbf{y}(x_0) + h \sum_{j=1}^s a_{ij} \mathbf{y}'(x_0 + c_j h) + \mathcal{O}(h^{q+1}),$$

denn Taylorentwicklung zeigt

$$\begin{aligned} \mathbf{y}(x_0 + c_i h) &= \sum_{\ell=0}^q \frac{c_i^\ell h^\ell}{\ell!} \mathbf{y}^{(\ell)}(x_0) + \mathcal{O}(h^{q+1}), \\ h \sum_{j=1}^s a_{ij} \mathbf{y}'(x_0 + c_j h) &= \sum_{\ell=1}^q \underbrace{\sum_{j=1}^s a_{ij} c_j^{\ell-1}}_{c_i^\ell / \ell} \frac{h^\ell}{(\ell-1)!} \mathbf{y}^{(\ell)}(x_0) + \mathcal{O}(h^{q+1}). \end{aligned}$$

Somit erhält man

$$\mathbf{y}(x_0 + c_i h) = \mathbf{y}(x_0) + h \sum_{j=1}^s a_{ij} \mathbf{f}(x_0 + c_j h, \mathbf{y}(x_0 + c_j h)) + \mathcal{O}(h^{q+1}).$$

Zieht man diesen Ausdruck von den Runge-Kutta-Gleichungen

$$\mathbf{Y}_i = \mathbf{y}(x_0) + h \sum_{j=1}^s a_{ij} \mathbf{f}(x_0 + c_j h, \mathbf{Y}_j)$$

ab und verwendet die Lipschitzbedingung an  $\mathbf{f}$ , so erhält man für  $h$  klein genug

$$\|\mathbf{Y}_i - \mathbf{y}(x_0 + c_i h)\| \leq ch^{q+1},$$

was das gewünschte Resultat ist. □

**Satz 4.34** Falls die Knoten  $c_1, \dots, c_s$  paarweise verschieden sind, dann ist die  $s \times s$  Matrix  $\mathbf{A}$  eindeutig durch C(s) bestimmt.

*Beweis:* Die Aussage folgt unmittelbar aus der Bedingung C(s)

$$\mathbf{A} \begin{bmatrix} 1 & c_1 & \dots & c_1^{s-1} \\ 1 & c_2 & \dots & c_2^{s-1} \\ \vdots & \vdots & & \vdots \\ 1 & c_s & \dots & c_s^{s-1} \end{bmatrix} = \begin{bmatrix} c_1 & \dots & \frac{1}{s} c_1^s \\ \vdots & & \vdots \\ c_s & \dots & \frac{1}{s} c_s^s \end{bmatrix}$$

und der Tatsache, dass die Vandermonde-Matrix (rechts von  $\mathbf{A}$ ) invertierbar ist. □

Für die folgenden Überlegungen erinnern wir kurz an Quadraturformeln (Numerische Mathematik 1, Kapitel 1.1, 1.4 und 1.5). Für (d) wird auf die Literatur verwiesen.

- (a) Eine Quadraturformel  $(b_i, c_i)_{i=1, \dots, s}$  hat Ordnung  $p$ , genau dann wenn die Bedingungen  $B(p)$  erfüllt sind.
- (b) Es existiert keine  $s$ -stufige Quadraturformel der Ordnung  $p \geq 2s + 1$ .
- (c) Es existiert eine eindeutige  $s$ -stufige Quadraturformel der Ordnung  $p = 2s$ , die Gauß'sche Quadraturformel. Sie hat positive Gewichte.
- (d) Die  $s$ -stufige Quadraturformel von Radau besitzt als Knoten die Nullstellen des Polynoms

$$\frac{d^{s-1}}{dx^{s-1}} (x^{s-1}(1-x)^s),$$

insbesondere ist  $c_s = 1$ . Die Knoten sind paarweise verschieden, die Gewichte berechnen sich daher eindeutig aus  $B(s)$ . Diese Quadraturformel hat Ordnung  $p = 2s - 1$ , ihre Gewichte sind positiv.

**Definition 4.35** Sei  $(b_i, c_i)_{i=1, \dots, s}$  die Gauß'sche Quadraturformel mit  $s$  Knoten. Das  $s$ -stufige IRK-Verfahren

$$\begin{array}{c|c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b}^\top \end{array},$$

wobei  $\mathbf{A}$  durch  $C(s)$  eindeutig bestimmt ist, heißt *Gauß-Verfahren*.

**Definition 4.36** Sei  $(b_i, c_i)_{i=1, \dots, s}$  die Radau'sche Quadraturformel mit  $s$  Knoten und  $c_s = 1$ . Das  $s$ -stufige IRK-Verfahren

$$\begin{array}{c|c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b}^\top \end{array},$$

wobei  $\mathbf{A}$  durch  $C(s)$  eindeutig bestimmt ist, heißt *Radau IIA-Verfahren*.

**Satz 4.37** Die eben definierten IRK-Verfahren haben die folgenden Eigenschaften:

	Stufen	Ordnung $p$	Stufenordnung $q$
Gauß	$s$	$2s$	$s$
Radau IIA	$s$	$2s - 1$	$s$

Beide Verfahrensklassen sind A-stabil. Die Radau IIA-Verfahren sind zudem stiffly accurate und erfüllen  $R(\infty) = 0$ . Die Gauß-Verfahren haben  $R(\infty) = (-1)^s$ .

*Beweis:* Aus  $C(s)$  folgt sofort  $q = s$ . Für den Beweis der klassischen Ordnung  $p$  und die Stabilität verweise ich auf die Literatur.

**Beispiel 4.38** (Gauß-Verfahren) Für  $s = 1$  sind die Koeffizienten  $b_1 = 1$  und  $c_1 = \frac{1}{2} = a_{11}$ . Das Verfahren lautet

$$\mathbf{Y}'_1 = \mathbf{f} \left( x_0 + \frac{h}{2}, \mathbf{y}_0 + \frac{h}{2} \mathbf{Y}'_1 \right), \quad \mathbf{y}_1 = \mathbf{y}_0 + \frac{h}{2} \mathbf{Y}'_1$$

und somit insgesamt

$$\mathbf{y}_1 = \mathbf{y}_0 + h \mathbf{f} \left( x_0 + \frac{h}{2}, \frac{\mathbf{y}_0 + \mathbf{y}_1}{2} \right).$$

Das Verfahren heißt *implizite Mittelpunktsregel*. Es hat Stufenordnung 1 und Ordnung 2. Seine Stabilitätsfunktion lautet

$$R(z) = 1 + z \left(1 - \frac{z}{2}\right)^{-1} = \frac{1 + \frac{z}{2}}{1 - \frac{z}{2}}.$$

Die Stabilitätsfunktion hat einen Pol bei  $z = 2$ , ist analytisch auf  $\overline{\mathbb{C}} \setminus \{2\}$  und auf der imaginären Achse durch 1 beschränkt. Nach dem Maximumprinzip ist das Verfahren daher A-stabil und es gilt  $R(\infty) = -1$ .

Für  $s = 2$  lauten die Koeffizienten  $b_1 = b_2 = \frac{1}{2}$ ,  $c_{1,2} = \frac{1}{2} \mp \frac{\sqrt{3}}{6}$  und

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & c_1 \\ 1 & c_2 \end{bmatrix} = \begin{bmatrix} c_1 & \frac{1}{2}c_1^2 \\ c_2 & \frac{1}{2}c_2^2 \end{bmatrix}.$$

Das ergibt das 2-stufige Gauß-Verfahren

$$\begin{array}{c|cc} \frac{1}{2} - \frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{1}{2} + \frac{\sqrt{3}}{6} & \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

der Ordnung 4. Seine Stabilitätsfunktion lautet:

$$R(z) = \frac{1 + \frac{z}{2} + \frac{z^2}{12}}{1 - \frac{z}{2} + \frac{z^2}{12}}.$$

Die Pole der Stabilitätsfunktion sind in der rechten Halbebene, die Stabilitätsfunktion ist auf der imaginären Achse durch 1 beschränkt. Damit ist das Verfahren nach dem Maximumprinzip A-stabil. Zudem ist  $R(\infty) = 1$ .

**Beispiel 4.39** (Radau IIA-Verfahren) Für  $s = 1$  lauten die Koeffizienten  $b_1 = 1$  und  $c_1 = 1 = a_{11}$ . Das 1-stufige Radau IIA-Verfahren ist somit das implizite Eulerverfahren. Dieses ist A-stabil und es gilt  $R(\infty) = 0$ .

Für  $s = 2$  lauten die Koeffizienten  $b_1 = \frac{3}{4}$ ,  $b_2 = \frac{1}{4}$ ,  $c_1 = \frac{1}{3}$ ,  $c_2 = 1$  und

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & c_1 \\ 1 & c_2 \end{bmatrix} = \begin{bmatrix} c_1 & \frac{1}{2}c_1^2 \\ c_2 & \frac{1}{2}c_2^2 \end{bmatrix}.$$

Das ergibt das 2-stufige Radau IIA-Verfahren

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}$$

der Ordnung 3. Seine Stabilitätsfunktion lautet:

$$R(z) = \frac{1 + \frac{z}{3}}{1 - \frac{2z}{3} + \frac{z^2}{6}}.$$

Die Pole der Stabilitätsfunktion sind in der rechten Halbebene, die Stabilitätsfunktion ist auf der imaginären Achse durch 1 beschränkt. Damit ist das Verfahren nach dem Maximumprinzip A-stabil. Zudem ist  $R(\infty) = 0$ .

## 4.10 Kollokationsverfahren

Wir betrachten wieder die Anfangswertaufgabe

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{y}_0 \in \mathbb{R}^m.$$

Für ihre Lösung wählt man eine Schrittweite  $h \neq 0$  und paarweise verschiedene Knoten (sg. *Kollokationsknoten*)

$$c_1, \dots, c_s \in [0, 1].$$

Man sucht nun ein Polynom  $\mathbf{p}$  vom Grad  $s$ , welches die folgenden Bedingungen erfüllt:

$$(\#) \quad \begin{cases} \mathbf{p}(x_0) = \mathbf{y}_0 \\ \mathbf{p}'(x_0 + c_i h) = \mathbf{f}(x_0 + c_i h, \mathbf{p}(x_0 + c_i h)), \quad i = 1, \dots, s \end{cases}$$

und wählt als numerische Approximation für  $\mathbf{y}(x_0 + h)$  die Größe

$$\mathbf{y}_1 := \mathbf{p}(x_0 + h).$$

Man verwendet dabei die folgende Sprechweise:

**Definition 4.40** Das eben definierte Polynom  $\mathbf{p}$  nennt man *Kollokationspolynom* mit Knoten  $c_1, \dots, c_s$ . Die Bedingungen (#) heißen *Kollokationsbedingungen*. Das zugehörige Verfahren heißt *Kollokationsverfahren*.

Kollokationsverfahren sind IRK-Verfahren, wie nachstehender Satz zeigt.

**Satz 4.41** Das Kollokationsverfahren mit Knoten  $c_1, \dots, c_s$  ist äquivalent zu dem  $s$ -stufigen IRK-Verfahren mit Koeffizienten

$$a_{ij} = \int_0^{c_i} \ell_j(t) dt, \quad b_j = \int_0^1 \ell_j(t) dt,$$

wobei

$$\ell_j(t) = \prod_{\substack{k=1 \\ k \neq j}}^s \frac{t - c_k}{c_j - c_k}$$

die zu  $c_1, \dots, c_j$  gehörigen Lagrange-Polynome bezeichnen.

*Beweis:* Wir definieren die Größen

$$\mathbf{Y}'_i = \mathbf{p}'(x_0 + c_i h)$$

und verwenden, dass  $\mathbf{p}'$  ein Polynom vom Grad kleiner gleich  $s - 1$  ist. Somit kann  $\mathbf{p}'$  eindeutig in der Basis der durch  $c_1, \dots, c_s$  definierten Lagrange-Polynome dargestellt werden

$$\mathbf{p}'(x_0 + th) = \sum_{j=1}^s \ell_j(t) \mathbf{Y}'_j.$$

Mittels Integration erhält man

$$\mathbf{p}(x_0 + c_i h) = \mathbf{p}(x_0) + h \int_0^{c_i} \mathbf{p}'(x_0 + th) dt = \mathbf{y}_0 + h \sum_{j=1}^s \underbrace{\int_0^{c_i} \ell_j(t) dt}_{a_{ij}} \mathbf{Y}'_j = \mathbf{Y}_i.$$

Wegen (#) gilt dann

$$\mathbf{Y}'_i = \mathbf{f}(x_0 + c_i h, \mathbf{Y}_i)$$

und somit

$$\mathbf{y}_1 = \mathbf{p}(x_0 + h) = \mathbf{p}(x_0) + h \int_0^1 \mathbf{p}'(x_0 + th) dt = \mathbf{y}_0 + h \sum_{j=1}^s \mathbf{b}_j \mathbf{Y}'_j,$$

was zu zeigen war. □

Der obige Satz zeigt auch, dass das Kollokationspolynom für genügend kleines  $h$  existiert und eindeutig ist (wegen der Eindeutigkeit der Runge-Kutta-Lösung).

**Satz 4.42** Gegeben sei ein Kollokationsverfahren mit Knoten  $c_1, \dots, c_s$ . Dann gilt: Die durch Satz 4.41 definierten Koeffizienten erfüllen die Bedingungen B(s) und C(s).

*Beweis:* Zum Nachweis von C(s) verwendet man

$$\sum_{j=1}^s a_{ij} c_j^{\ell-1} = \sum_{j=1}^s \left( \int_0^{c_i} \ell_j(t) dt \right) c_j^{\ell-1} = \int_0^{c_i} \underbrace{\sum_{j=1}^s \ell_j(t) c_j^{\ell-1}}_{t^{\ell-1}} dt = \frac{c_i^\ell}{\ell}, \quad 1 \leq \ell \leq s.$$

Ebenso zeigt man B(s). □

Der obige Satz zeigt: Bei gegebenen (paarweise verschiedenen) Knoten können die Koeffizienten  $a_{ij}$  und  $b_j$  aus den Bedingungen B(s) und C(s) berechnet werden (Gleichungen mit Vandermonde-Matrizen).

**Satz 4.43** Gegeben sei ein Kollokationsverfahren mit Knoten  $c_1, \dots, c_s$  und zugehörigen Gewichten  $b_1, \dots, b_s$ , berechnet aus B(s). Dann gilt: Das Kollokationsverfahren hat Ordnung  $p$ , genau dann wenn die Quadraturformel  $(b_i, c_i)_{i=1, \dots, s}$  Ordnung  $p$  hat.

Für den Beweis wird auf die Literatur verwiesen. Dieser Satz zeigt noch einmal, dass die Gauß-Verfahren Ordnung  $2s$  und die Radau IIA-Verfahren Ordnung  $2s - 1$  besitzen.

## 4.11 Kontraktivität und algebraische Stabilität von Runge-Kutta-Verfahren

In diesem Abschnitt lassen wir uns von folgender Idee leiten: Falls die Differentialgleichung  $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$  *kontraktiv* ist, d.h. für zwei beliebige Lösungen  $\mathbf{y}(x)$  und  $\mathbf{z}(x)$  gilt

$$\|\mathbf{y}(x_1) - \mathbf{z}(x_1)\| \leq \|\mathbf{y}(x_0) - \mathbf{z}(x_0)\|$$

für alle  $x_1 \geq x_0$ , so möchte man auch Kontraktivität für die numerische Lösung.

Gegeben sei die Anfangswertaufgabe  $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$  mit  $\mathbf{f} : \mathbb{R} \times \mathbb{C}^m \rightarrow \mathbb{C}^m$ . Weiters sei  $\langle \cdot, \cdot \rangle$  ein Skalarprodukt auf  $\mathbb{C}^m$  mit zugehöriger Norm  $\|\cdot\|$ . Für die Lösung der Differentialgleichung gilt:

$$\begin{aligned} \frac{d}{dx} \|\mathbf{y}(x) - \mathbf{z}(x)\|^2 &= \langle \mathbf{y}(x) - \mathbf{z}(x), \mathbf{y}'(x) - \mathbf{z}'(x) \rangle + \langle \mathbf{y}'(x) - \mathbf{z}'(x), \mathbf{y}(x) - \mathbf{z}(x) \rangle \\ &= 2 \operatorname{Re} \langle \mathbf{y}'(x) - \mathbf{z}'(x), \mathbf{y}(x) - \mathbf{z}(x) \rangle \\ &= 2 \operatorname{Re} \langle \mathbf{f}(x, \mathbf{y}(x)) - \mathbf{f}(x, \mathbf{z}(x)), \mathbf{y}(x) - \mathbf{z}(x) \rangle. \end{aligned}$$

Wir verwenden deshalb im Folgenden die Voraussetzung

$$(*) \quad \operatorname{Re} \langle \mathbf{f}(x, \mathbf{y}) - \mathbf{f}(x, \mathbf{z}), \mathbf{y} - \mathbf{z} \rangle \leq 0 \quad \text{für alle } x \in \mathbb{R} \text{ und alle } \mathbf{y}, \mathbf{z} \in \mathbb{C}^m.$$

Unter dieser Voraussetzung ist die Differentialgleichung  $\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$  kontraktiv.

**Definition 4.44** Ein Runge-Kutta-Verfahren heißt *kontraktiv* (B-stabil), wenn für jede Differentialgleichung, die (\*) erfüllt, gilt:

$$\|\mathbf{y}_1 - \mathbf{z}_1\| \leq \|\mathbf{y}_0 - \mathbf{z}_0\|,$$

wobei  $\mathbf{y}_1, \mathbf{z}_1$  die numerischen Lösungen nach einem Schritt mit beliebiger Schrittweite  $h > 0$  zu den Anfangswerten  $\mathbf{y}_0, \mathbf{z}_0$  bezeichnen.

Offensichtlich ist A-Stabilität eine Folge von B-Stabilität, man wähle einfach  $f(x, y) = \lambda y$  mit  $\operatorname{Re} \lambda \leq 0$ .

**Definition 4.45** Ein Runge-Kutta-Verfahren mit Koeffizienten  $c_i, a_{ij}, b_j$  heißt *algebraisch stabil*, wenn die folgenden beiden Bedingungen erfüllt sind:

- (i)  $b_i \geq 0, \quad i = 1, \dots, s$
- (ii)  $\mathbf{M} := \mathbf{B}\mathbf{A} + \mathbf{A}^\top \mathbf{B} - \mathbf{B}\mathbf{1} \cdot \mathbf{1}^\top \mathbf{B} \quad \text{mit } \mathbf{B} = \operatorname{diag}(b_1, \dots, b_s) \text{ ist positiv semidefinit.}$

Offenbar gilt  $\mathbf{M} = (m_{ij})$  mit  $m_{ij} = b_i a_{ij} + b_j a_{ji} - b_i b_j$ .

**Satz 4.46** Algebraische Stabilität impliziert B-Stabilität.

Für nicht konfluente (d.h. paarweise verschiedene) Knoten  $c_1, \dots, c_s$  sind beide Stabilitätsbegriffe äquivalent (für den Beweis, siehe Literatur).

*Beweis von Satz 4.46:* Es gilt

$$\mathbf{Y}_i = \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{Y}'_j, \quad \mathbf{Y}'_j = \mathbf{f}(x_0 + c_j h, \mathbf{Y}_j), \quad \mathbf{y}_1 = \mathbf{y}_0 + h \sum_{j=1}^s b_j \mathbf{Y}'_j$$

und

$$\mathbf{Z}_i = \mathbf{z}_0 + h \sum_{j=1}^s a_{ij} \mathbf{Z}'_j, \quad \mathbf{Z}'_j = \mathbf{f}(x_0 + c_j h, \mathbf{Z}_j), \quad \mathbf{z}_1 = \mathbf{z}_0 + h \sum_{j=1}^s b_j \mathbf{Z}'_j,$$

somit

$$\mathbf{y}_1 - \mathbf{z}_1 = \mathbf{y}_0 - \mathbf{z}_0 + h \sum_{j=1}^s b_j (\mathbf{Y}'_j - \mathbf{Z}'_j).$$

Damit folgt

$$\|\mathbf{y}_1 - \mathbf{z}_1\|^2 = \|\mathbf{y}_0 - \mathbf{z}_0\|^2 + 2h \sum_{j=1}^s b_j \operatorname{Re} \langle \mathbf{y}_0 - \mathbf{z}_0, \mathbf{Y}'_j - \mathbf{Z}'_j \rangle + h^2 \sum_{i,j=1}^s b_i b_j \langle \mathbf{Y}'_i - \mathbf{Z}'_i, \mathbf{Y}'_j - \mathbf{Z}'_j \rangle.$$

Als nächstes verwenden wir

$$\mathbf{y}_0 - \mathbf{z}_0 = \mathbf{Y}_j - \mathbf{Z}_j - h \sum_{i=1}^s a_{ji} (\mathbf{Y}'_i - \mathbf{Z}'_i),$$

in

$$\begin{aligned} 2 \sum_{j=1}^s b_j \operatorname{Re} \langle \mathbf{y}_0 - \mathbf{z}_0, \mathbf{Y}'_j - \mathbf{Z}'_j \rangle &= 2 \sum_{j=1}^s \underbrace{b_j}_{\geq 0} \underbrace{\operatorname{Re} \langle \mathbf{Y}_j - \mathbf{Z}_j, \mathbf{Y}'_j - \mathbf{Z}'_j \rangle}_{\leq 0 \text{ wegen } (*)} \\ &\quad - h \sum_{i,j=1}^s b_j a_{ji} (\langle \mathbf{Y}'_i - \mathbf{Z}'_i, \mathbf{Y}'_j - \mathbf{Z}'_j \rangle + \langle \mathbf{Y}'_j - \mathbf{Z}'_j, \mathbf{Y}'_i - \mathbf{Z}'_i \rangle). \end{aligned}$$

(Man hat hier  $2 \operatorname{Re} w = w + \bar{w}$  verwendet.) Damit hat man insgesamt

$$\|\mathbf{y}_1 - \mathbf{z}_1\|^2 \leq \|\mathbf{y}_0 - \mathbf{z}_0\|^2 - h^2 \sum_{i,j=1}^s m_{ij} \langle \mathbf{Y}'_i - \mathbf{Z}'_i, \mathbf{Y}'_j - \mathbf{Z}'_j \rangle,$$

und die Aussage des Satzes ergibt sich aus dem folgenden Lemma.  $\square$

**Lemma 4.47** Seien  $\mathbf{w}_1, \dots, \mathbf{w}_s \in \mathbb{C}^m$  und  $\mathbf{M} \in \mathbb{R}^{s \times s}$  symmetrisch und positiv semidefinit. Dann gilt

$$\sum_{i,j=1}^s m_{ij} \langle \mathbf{w}_i, \mathbf{w}_j \rangle \geq 0.$$

*Beweis:* Wegen der Voraussetzungen an  $\mathbf{M}$  gibt es eine orthogonale Matrix  $\mathbf{Q}$  mit

$$\mathbf{M} = \mathbf{Q}^T \operatorname{diag}(d_1, \dots, d_s) \mathbf{Q}, \quad d_i \geq 0,$$

also

$$m_{ij} = \sum_{k=1}^s q_{ki} d_k q_{kj}, \quad q_{\mu\nu} \in \mathbb{R}.$$

Damit ergibt sich

$$\begin{aligned} \sum_{i,j=1}^s m_{ij} \langle \mathbf{w}_i, \mathbf{w}_j \rangle &= \sum_{i,j=1}^s \sum_{k=1}^s q_{ki} d_k q_{kj} \langle \mathbf{w}_i, \mathbf{w}_j \rangle \\ &= \sum_{k=1}^s d_k \sum_{i,j=1}^s q_{ki} q_{kj} \langle \mathbf{w}_i, \mathbf{w}_j \rangle = \sum_{k=1}^s d_k \langle \mathbf{u}_k, \mathbf{u}_k \rangle \geq 0, \end{aligned}$$

wobei  $\mathbf{u}_k = \sum_{j=1}^s q_{kj} \mathbf{w}_j$ .

□

**Beispiel 4.48** (Gauß-Verfahren)

$$\begin{aligned} s = 1 : \quad & b_1 = 1, \quad a_{11} = \frac{1}{2}, \quad \mathbf{M} = \frac{1}{2} + \frac{1}{2} - 1 = 0 \\ s = 2 : \quad & b_1 = b_2 = \frac{1}{2}, \quad \mathbf{A} = \frac{1}{12} \begin{bmatrix} 3 & 3 - 2\sqrt{3} \\ 3 + 2\sqrt{3} & 3 \end{bmatrix}, \\ & \mathbf{M} = \frac{1}{2} (\mathbf{A} + \mathbf{A}^\top) - \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Allgemein gilt für die Gauß-Verfahren stets  $\mathbf{M} = \mathbf{0}$ . Außerdem haben sie positive Gewichte (vgl. Numerische Mathematik 1). Deshalb sind Gauß-Verfahren algebraisch stabil und folglich auch A-stabil.

**Beispiel 4.49** (Radau IIA-Verfahren)

$$\begin{aligned} s = 1 : \quad & b_1 = 1, \quad a_{11} = 1, \quad \mathbf{M} = 1 + 1 - 1 = 1 \\ s = 2 : \quad & b_1 = \frac{3}{4}, \quad b_2 = \frac{1}{4}, \quad \mathbf{A} = \frac{1}{12} \begin{bmatrix} 5 & -1 \\ 9 & 3 \end{bmatrix}, \\ & \mathbf{M} = \frac{1}{16} \begin{bmatrix} 10 & 2 \\ 2 & 2 \end{bmatrix} - \frac{1}{16} \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{aligned}$$

Radau IIA-Verfahren sind für beliebiges  $s$  algebraisch stabil und folglich auch A-stabil.

**Beispiel 4.50** (Trapezregel)

$$b_1 = b_2 = \frac{1}{2}, \quad \mathbf{A} = \frac{1}{2} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{M} = \frac{1}{4} \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Die Trapezregel ist nicht algebraisch stabil. Dennoch ist sie aber A-stabil.