Kapitel 6

Zur Wahl der Gitterpunkte

6.1 Überblick

Die Wahl der inneren Gitterpunkte $(t_i)_{i=2}^{N-1}$ bzw. $(\tau_i)_{i=2}^{N-1}$ kann den Bedarf an Rechenzeit und Speicherplatz sowie die Qualität der diskretisierten Lösung stark beeinflussen.
Grundlage einer Anpassung der Gitterpunkte ist immer eine geeignete Fehlerschätzung. Zwei Vorgehensweisen sind prinzipiell möglich:

(i) Bei der *statischen* Anpassung der Gitterpunkte wird, ausgehend von der Lösung des diskretisierten Problems, für eine vorgegebene erste Gitterpunktverteilung eine neue zweite Gitterpunktverteilung berechnet, die den geschätzten Fehler minimieren bzw. reduzieren soll. Das Optimierungsproblem wird dann ein zweites Mal mit den neuen Gitterpunkten gelöst. Dabei kann auch die Anzahl der Gitterpunkte verändert werden. Solche statischen Gitteranpassungen werden üblicherweise bei Kollokationsverfahren für Randwertprobleme angewandt.

(ii) Bei der *beweglichen* oder *dynamischen* Anpassung der Gitterpunkte werden die inneren Gitterpunkte als variabel angesehen. Für jeden beweglichen inneren Gitterpunkt ist es notwendig, eine zusätzliche Gleichung als Nebenbedingung im Optimierungsproblem zu formulieren, die diesen z. B. im Sinne der Minimierung einer Fehlerschätzung festlegt.
Diese beweglichen Gitterpunktanpassungen sind bei Kollokationsverfahren für Randwertprobleme bisher nicht gebräuchlich.

Als neue Kriterien zur statischen Gitteranpassung speziell bei direkten Kollokationsverfahren für optimale Steuerungsprobleme werden

- die Reduzierung des lokalen Optimalitätsfehlers (nach der Abschätzung aus Kapitel 5.4)

sowie zwei Kriterien zur beweglichen Wahl der Gitterpunkte, nämlich

- die Minimierung des Kollokationsfehlers und
- die Gleichverteilung der Variation in x,

hergeleitet.

Die beiden Strategien für bewegliche Gitterpunkte sind prinzipiell auch bei Kollokationsverfahren für Randwertprobleme anwendbar, falls die dort verwendeten Verfahren zur Lösung nichtlinearer Gleichungssysteme geeignet erweitert werden (um die Einhaltung der linearen Nebenbedingungen $\tau_i < \tau_{i+1}$ zu garantieren). Zur Berücksichtigung aktiver Zustandsbeschränkungen im optimalen Steuerungsproblem bei der Gitteranpassung ist jedoch noch eine geeignete Erweiterung dieser beiden Kriterien notwendig. Die genannten Möglichkeiten zur Gitteranpassung sind im Programm DIRCOL implementiert worden.

6.2 Bewegliche Gitteranpassung zur Minimierung des Kollokationsfehlers

6.2.1 Abschätzung des Kollokationsfehlers

Vernachlässigt man den Fehler in den Ungleichungsbeschränkungen von (P) und nimmt man die optimale Steuerung $u = u_{\text{opt}}$ und $t_f = t_{f\text{opt}}$ als gegeben und die Randbedingungen als geeignet an, so erfüllt $x(t)$ als Lösung des optimalen Steuerungsproblems auch ein gewöhnliches Randwertproblem, nämlich (mit $f^\star(x(t),t) = f(x(t), u_{\text{opt}}(t), t)$)

$$\dot{x}(t) = f^\star(x(t),t), \qquad t \in [t_0, t_f], \tag{6.1}$$

$$0 = r^\star(x(t_0), x(t_f)) \tag{6.2}$$

(das Symbol $\star$ wird im folgenden wieder unterdrückt).
Zur Ermittlung des Fehlers einer diskretisierten Lösung dieses Randwertproblems betrachten wir zunächst den Fall einer einzelnen Differentialgleichung. Es sei r ein stetiges lineares Funktional auf $\mathrm{C}^0([t_0, t_f])$ und linear unabhängig über $\mathbb{P}_1$ [15]. Gegeben sei außerdem eine strikte Unterteilung $\Delta := (t_i)_{i=1}^N$, $N \geq 3$, von $[t_0, t_f]$, d. h.

$$t_0 = t_1 < t_2 < \ldots < t_{N-1} < t_N = t_f. \tag{6.3}$$

Folgende Bezeichnungen werden verwendet:

$$||x||_\infty := \sup_{t_0 \leq t \leq t_f} |x(t)|, \quad x \in \mathrm{C}^0([t_0, t_f]),$$

und $\mathbb{P}_{k,\Delta}$, $k = 2, 4$, sei die Menge aller Funktionen auf $[t_0, t_f]$ mit der Eigenschaft, daß eine Funktion aus dieser Menge auf jedem Teilintervall $[t_i, t_{i+1}]$, $i = 1, \ldots, N-1$, lokal ein Polynom vom Grade echt kleiner als k ist (vgl. Kapitel 3.1 und 3.2).

Im folgenden bezeichne x die Lösung des Randwertproblems, und es werden die Notationen aus Kapitel 3 verwendet. Es gelten die folgenden Sätze:

Satz 6.1 *Es sei $x_{\text{app}} \in \mathrm{S}^2_\Delta := \mathbb{P}_{2,\Delta} \cap \mathrm{C}^0([t_0, t_f])$ eine isolierte Lösung von*

$$\dot{x}_{\text{app}}(t) = f(x_{\text{app}}(t), t), \ \textit{für } t = t_{1+1/2}, t_{2+1/2}, \ldots, t_{N-1/2}, \ \textit{und}$$

$$0 = r(x_{\text{app}}(t_0), x_{\text{app}}(t_f)),$$

und f sei hinreichend glatt. Dann gelten:

(i) Für jedes Gitter Δ mit hinreichend kleinem h existiert genau eine Lösung $x_{\text{app}}(\Delta)$ „nahe bei“ x.

(ii) Es gilt die Fehlerabschätzung

$$||(x - x_{\text{app}})^{(j)}||_\infty \leq C_1 h ||x^{(2)}||_\infty, \ j = 0, 1. \tag{6.4}$$

(iii) Es gilt die verschärfte Fehlerabschätzung

$$||x - x_{\text{app}}||_\infty \leq C_2 \max\{h_i \sup_{t_i \leq t \leq t_{i+1}} |x^{(2)}(t)| : \ i = 1, \ldots, N-1\} + o(h). \tag{6.5}$$

Dabei sind C_1 und C_2 Konstanten, die nur von f und r abhängen.

Satz 6.2 *Es sei $x_{\text{app}} \in \mathrm{S}^4_\Delta := \mathrm{I\!P}_{4,\Delta} \cap \mathrm{C}^0([t_0, t_f])$ eine isolierte Lösung von*

$$\begin{aligned} \dot{x}_{\text{app}}(t) &= f(x_{\text{app}}(t), t), \ \text{für } t = t_1, t_{1+1/2}, t_2, t_{2+1/2}, \ldots, t_{N-1/2}, t_N, \ \text{und} \\ 0 &= r(x_{\text{app}}(t_0), x_{\text{app}}(t_f)), \end{aligned}$$

und f sei hinreichend glatt. Dann gelten:

(i) Für jedes Gitter Δ mit hinreichend kleinem h existiert genau eine Lösung $x_{\text{app}}(\Delta)$ „nahe bei“ x.

(ii) Es gilt die Fehlerabschätzung

$$||(x - x_{\text{app}})^{(j)}||_\infty \leq C_3 h^3 ||x^{(4)}||_\infty, \ j = 0, 1. \tag{6.6}$$

(iii) Es gilt die verschärfte Fehlerabschätzung

$$||x - x_{\text{app}}||_\infty \leq C_4 \max\{h_i^3 \sup_{t_i \leq t \leq t_{i+1}} |x^{(4)}(t)| : \ i = 1, \ldots, N-1\} + o(h^3). \tag{6.7}$$

Dabei sind C_3 und C_4 Konstanten, die nur von f und r abhängen.

Beweise: Spezialisierung der Resultate von Russell, Shampine [117], de Boor, Swartz [15] und de Boor [14]. □

Bemerkung 6.3 Die entsprechenden Fehlerabschätzungen bei Systemen von n Differentialgleichungen $\dot{x}_i(t) = f_i(x(t), t), \ i = 1, \ldots, n$, erhält man, indem man die obige Definition von $||.||_\infty$ ersetzt durch (vgl. [118])

$$||x||_\infty = \max_{1 \leq i \leq n} \sup_{t_0 \leq t \leq t_f} |x_i(t)|. \tag{6.8}$$

Bemerkung 6.4 Die Konstanten der obigen Fehlerabschätzungen sind nicht invariant gegen Skalierungen der Form

$$x \longrightarrow \hat{x} = Dx, \ \text{mit } D = \operatorname{diag}(d_1, \ldots, d_n), \ d_i \in \mathrm{I\!R}^+. \tag{6.9}$$

Dies ist ein (weiterer) Hinweis auf die große praktische Bedeutung einer „vernünftigen“ Skalierung aller auftretenden abhängigen und unabhängigen Variablen (vgl. Kapitel 7.1.5 und Kapitel 8).

6.2.2 Minimierung des Kollokationsfehlers

In diesem Abschnitt gelten dieselben Annahmen wie zu Beginn des vorherigen Abschnittes 6.2.1.

Eine Strategie zur Anpassung der Gitterpunkte

Annahme: *h sei hinreichend klein, so daß $o(h)$ in Satz 6.1 bzw. $o(h^3)$ in Satz 6.2 vernachlässigbar sind, d. h. für $x_{\text{app}} \in \mathrm{S}_\Delta^{2+k},\ k = 0$ bzw. 2, gelte*

$$||x - x_{\text{app}}||_\infty \leq C_{2+k} \max\{h_i^{1+k} \sup_{t_i \leq t \leq t_{i+1}} |x^{(2+k)}(t)| :\ i = 1, \ldots, N-1\}. \tag{6.10}$$

Die inneren Gitterpunkte $t_2, \ldots, t_{N-1}$ sind nun so zu plazieren, daß die rechte Seite dieser Fehlerabschätzung minimal wird, also

$$\min_{t_2,\ldots,t_{N-1}} \max\{h_i^{1+k} \sup_{t_i \leq t \leq t_{i+1}} |x^{(2+k)}(t)| :\ i = 1, \ldots, N-1\}\ ! \tag{6.11}$$

Ein Minimum liegt vor, wenn für $t_2, \ldots, t_{N-1}$ gilt [14]

$$h_i^{1+k} \sup_{t_i \leq t \leq t_{i+1}} |x^{(2+k)}(t)| \ = \ \text{const}_k, \quad i = 1, \ldots, N-1, \tag{6.12}$$

(für $x_{\text{app}} \in \mathrm{S}_\Delta^{2+k},\ k = 0$ bzw. 2). Mit

$$r_{k,i} \ := \ \sup_{t_i \leq t \leq t_{i+1}} |x^{(2+k)}(t)|,\ k = 0 \text{ bzw. } 2, \tag{6.13}$$

können also die inneren Gitterpunkte $t_i,\ i = 2, \ldots, N-2$, durch die Gleichungen

$$h_{i+1}^{1+k} r_{k,i+1} - h_i^{1+k} r_{k,i} \ = \ 0,\ i = 1, \ldots, N-2, \tag{6.14}$$

festgelegt werden (für $x_{\text{app}} \in \mathrm{S}_\Delta^{2+k},\ k = 0$ bzw. 2). Nimmt man also im diskretisierten Problem die Lage der $N-2$ inneren Gitterpunkte als zusätzliche Freiheitsgrade hinzu, so sind diese durch die zusätzlichen $N-2$ Gleichungen (6.14) formal bestimmt.

Bemerkung 6.5 Um diese Gleichungen als Nebenbedingungen in einem Optimierungsverfahren verwenden zu können, muß die unbekannte Funktion $x^{(2+k)}$ noch geeignet approximiert werden. Auch ist darauf zu achten, daß die Gleichungen (6.14) stetig differenzierbar von allen unabhängigen Veränderlichen des diskretisierten Problems abhängen, da die verwendeten numerischen Algorithmen die mindestens einmalige Differenzierbarkeit aller Funktionen benützen.

Approximation der Restglieder

Da die gesuchte Lösung x des Randwertproblems unbekannt ist, wird nun im folgenden mit Hilfe einer Näherungslösung $x_{\text{app}} \in \mathrm{S}_\Delta^{2+k}$ ($k = 0$ oder 2) eine Näherung für die unbekannten Funktionen $x^{(2+k)}(t)$ konstruiert, um die Gleichungen (6.14) zur optimalen Bestimmung der Gitterpunkte verwenden zu können.

Dazu betrachten wir zunächst die folgende Interpolationsaufgabe: Gesucht ist ein Polynom 2-ten Grades $p(x)$, das die 3 Punkte $(x_i, y_i),\ i = 1, 2, 3,\ x_1 < x_2 < x_3$, interpoliert. Mit Hilfe der Lagrangeschen Darstellung des Interpolationpolynoms erhält man sofort

$$\begin{aligned} p(x) &= y_1 L_1(x) + y_2 L_2(x) + y_3 L_3(x) \qquad (6.15)\\ \text{mit} \quad L_i(x) &= \frac{(x - x_k)(x - x_l)}{(x_i - x_k)(x_i - x_l)}, \quad i, k, l \in \{1, 2, 3\},\ k \neq i,\ l \neq i,\ k \neq l. \end{aligned}$$

Die Auswertung von $p'(x) = \mathrm{d}p/\mathrm{d}x$ bei x_i liefert

$$p'(x_1) = y_1 \frac{(2x_1 - x_2 - x_3)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(x_1 - x_3)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(x_1 - x_2)}{(x_3 - x_1)(x_3 - x_2)}, \tag{6.16}$$

$$p'(x_2) = y_1 \frac{(x_2 - x_3)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(2x_2 - x_1 - x_3)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(x_2 - x_1)}{(x_3 - x_1)(x_3 - x_2)}, \tag{6.17}$$

$$p'(x_3) = y_1 \frac{(x_3 - x_2)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(x_3 - x_1)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(2x_3 - x_1 - x_2)}{(x_3 - x_1)(x_3 - x_2)}. \tag{6.18}$$

Mit Hilfe dieser Beziehungen wollen wir nun eine Approximation von $x^{(2+k)}$, $k = 0, 2$, bestimmen. Falls $x_{\mathrm{app}} \in \mathrm{S}_\Delta^{2+k}$ gewählt wurde, so ist $x_{\mathrm{app}}^{(1+k)}$ eine stückweise konstante Funktion

$$x_{\mathrm{app}}^{(1+k)}(t) = x_{\mathrm{app}}^{(1+k)}(t_{i+1/2}) = \mathrm{const}_i =: x_{\mathrm{app},i}^{(1+k)}, \quad t_i \le t < t_{i+1}, \quad i = 1, \ldots, N-1. \tag{6.19}$$

Mit $x_1 = t_{i-1/2}$, $x_2 = t_{i+1/2}$, $x_3 = t_{i+3/2}$ und $y_1 = x_{\mathrm{app},i-1}^{(1+k)}$, $y_2 = x_{\mathrm{app},i}^{(1+k)}$, $y_3 = x_{\mathrm{app},i+1}^{(1+k)}$ bestimmen wir die interpolierende Parabel $p(x)$ und setzen in den inneren Intervallen (d. h. für $i = 2, \ldots, N-2$)

$$x^{(2+k)}(t) = x^{(2+k)}(t_{i+1/2}) + \mathcal{O}\left(h_i/2\right) \doteq x^{(2+k)}(t_{i+1/2}) \doteq p'(t_{i+1/2}), \quad t_i \le t < t_{i+1}. \tag{6.20}$$

Im linken Randintervall setzen wir $x_1 = t_{1+1/2}$, $x_2 = t_{2+1/2}$, $x_3 = t_{3+1/2}$ und $y_1 = x_{\mathrm{app},1}^{(1+k)}$, $y_2 = x_{\mathrm{app},2}^{(1+k)}$, $y_3 = x_{\mathrm{app},3}^{(1+k)}$ und werten die Ableitung der interpolierenden Parabel bei x_1 aus

$$x^{(2+k)}(t) \doteq x^{(2+k)}(t_{1+1/2}) \doteq p'(t_{1+1/2}), \quad t_1 \le t \le t_2. \tag{6.21}$$

Im rechten Randintervall setzen wir $x_1 = t_{N-5/2}$, $x_2 = t_{N-3/2}$, $x_3 = t_{N-1/2}$ und $y_1 = x_{\mathrm{app},N-3}^{(1+k)}$, $y_2 = x_{\mathrm{app},N-2}^{(1+k)}$, $y_3 = x_{\mathrm{app},N-1}^{(1+k)}$, werten die Ableitung der interpolierenden Parabel bei x_3 aus und erhalten

$$x^{(2+k)}(t) \doteq x^{(2+k)}(t_{N-1/2}) \doteq p'(t_{N-1/2}), \quad t_{N-1} \le t < t_N. \tag{6.22}$$

Bemerkung 6.6 Um die interpolierenden Parabeln definieren zu können, werden mindestens 3 Diskretisierungsintervalle benötigt. Daher muß jetzt für die Anzahl N der Gitterpunkte $N \ge 4$ vorausgesetzt werden.

Mit Hilfe der Beziehungen

$$x_2 - x_1 = t_{i+1/2} - t_{i-1/2} = \frac{1}{2}(h_{i-1} + h_i), \tag{6.23}$$

$$x_3 - x_2 = t_{i+3/2} - t_{i+1/2} = \frac{1}{2}(h_i + h_{i+1}), \tag{6.24}$$

$$x_3 - x_1 = t_{i+3/2} - t_{i-1/2} = \frac{1}{2}(h_{i-1} + 2h_i + h_{i+1}), \tag{6.25}$$

$$2x_2 - x_1 - x_3 = \frac{1}{2}(h_{i-1} - h_{i+1}), \tag{6.26}$$

$$2x_1 - x_2 - x_3 = -(h_{i-1} + \frac{3}{2}h_i + \frac{1}{2}h_{i+1}), \tag{6.27}$$

$$2x_3 - x_1 - x_2 = \frac{1}{2}h_{i-1} + \frac{3}{2}h_i + h_{i+1}, \tag{6.28}$$

erhält man in den inneren Intervallen

$$\begin{aligned} x^{(2+k)}(t_{i+1/2}) &\doteq -2x^{(1+k)}_{\mathrm{app},i-1}\frac{h_i+h_{i+1}}{(h_{i-1}+h_i)(h_{i-1}+2h_i+h_{i+1})} \\ &\quad +2x^{(1+k)}_{\mathrm{app},i}\frac{h_{i+1}-h_{i-1}}{(h_{i-1}+h_i)(h_i+h_{i+1})}+2x^{(1+k)}_{\mathrm{app},i+1}\frac{h_{i-1}+h_i}{(h_{i-1}+2h_i+h_{i+1})(h_i+h_{i+1})}, \\ &\quad i=2,\ldots,N-2. \end{aligned} \tag{6.29}$$

Im linken Randintervall erhält man

$$\begin{aligned} x^{(2+k)}(t_{1+1/2}) &\doteq -4x^{(1+k)}_{\mathrm{app},1}\frac{h_1+\frac{3}{2}h_2+\frac{1}{2}h_3}{(h_1+h_2)(h_1+2h_2+h_3)} \\ &\quad +2x^{(1+k)}_{\mathrm{app},2}\frac{h_1+2h_2+h_3}{(h_1+h_2)(h_2+h_3)}-2x^{(1+k)}_{\mathrm{app},3}\frac{h_1+h_2}{(h_1+2h_2+h_3)(h_2+h_3)} \end{aligned} \tag{6.30}$$

und im rechten Randintervall

$$\begin{aligned} x^{(2+k)}(t_{N-1/2}) &\doteq +2x^{(1+k)}_{\mathrm{app},N-3}\frac{h_{N-2}+h_{N-1}}{(h_{N-3}+h_{N-2})(h_{N-3}+2h_{N-2}+h_{N-1})} \\ &\quad -2x^{(1+k)}_{\mathrm{app},N-2}\frac{h_{N-3}+2h_{N-2}+h_{N-1}}{(h_{N-3}+h_{N-2})(h_{N-2}+h_{N-1})} \\ &\quad +4x^{(1+k)}_{\mathrm{app},N-1}\frac{\frac{1}{2}h_{N-3}+\frac{3}{2}h_{N-2}+h_{N-1}}{(h_{N-3}+2h_{N-2}+h_{N-1})(h_{N-2}+h_{N-1})}. \end{aligned} \tag{6.31}$$

Nach Kapitel 3 ist für $x \in S^2_\Delta$

$$x^1_{\mathrm{app},i} = \frac{x_{i+1}-x_i}{h_i} \quad \text{mit} \quad x_i = x_{\mathrm{app}}(t_i) \tag{6.32}$$

und für $x \in S^4_\Delta$ ist

$$\begin{aligned} x^{(3)}_{\mathrm{app},i} &= 6\frac{c_{i,3}}{h_i^3}, \quad i=1,\ldots,N-1, \qquad (6.33) \\ \text{mit} \quad c_{i,3} &= 2x_i + h_i f_i - 2x_{i+1} + h_i f_{i+1}, \\ x_i &= x_{\mathrm{app}}(t_i), \quad u_i = u_{\mathrm{app}}(t_i) \quad \text{und} \quad f_i = f(x_i,u_i,t_i). \end{aligned}$$

Damit liegen nun Approximationen der Restglieder (6.13) vor, und die Auswertung der Gleichungen (6.14) ist mit diesen nun im Prinzip möglich.

Erweiterung auf Systeme von Differentialgleichungen

Da im allgemeinen die betrachteten dynamischen Systeme mehrere Differentialgleichungen enthalten, ist eine Erweiterung der Gleichungen (6.14) notwendig. Dazu werden die folgenden Beziehungen benötigt:

(i) Es seien $y \in \mathbb{R}^n$ und

$$||y||_\infty := \max_{1\le i\le n}|y_i|, \qquad ||y||_p := \Big(\sum_{i=1}^n |y_i|^p\Big)^{\frac{1}{p}},\ p\in\mathbb{N}.$$

Dann gelten $\forall p \in \mathbb{N}$ und $\forall y \in \mathbb{R}$

$$||y||_\infty \le ||y||_{p+1} \le ||y||_p, \tag{6.34}$$

und

$$\lim_{p\to\infty}||y||_p = ||y||_\infty. \tag{6.35}$$

(ii) Es seien $x \in \mathrm{C}^0([a,b]),\ a < b,\ a, b \in \mathbb{R}$, und

$$||x||_\infty \ := \ \sup_{a \le t \le b} |x(t)|, \qquad ||x||_q \ := \ \Big(\int_a^b |x(t)|^q \mathrm{d}t\Big)^{\frac{1}{q}},\ q \in \mathbb{N}.$$

Dann gelten $\forall q \in \mathbb{N}$ und $\forall x \in \mathrm{C}^0([a,b])$

$$||x||_\infty \ \le \ ||x||_{q+1} \ \le \ ||x||_q \tag{6.36}$$

und

$$\lim_{q \to \infty} ||x||_q \ = \ ||x||_\infty. \tag{6.37}$$

Es sei jetzt $x = (x_1, \ldots, x_n)^{\mathrm{T}}$ mit $x_i \in C^0([a,b])$ für $1 \le i \le n$ und $i, n \in \mathbb{N}$. Dann folgt aus (i) und (ii) die Abschätzung

$$\begin{aligned} ||x||_\infty \ &= \ \max_{1 \le i \le n} \sup_{t_0 \le t \le t_f} |x_i(t)| \ \le \ \Big(\sum_{i=1}^n \sup_{a \le t \le b} |x_i(t)|^p\Big)^{\frac{1}{p}}, \qquad p \in \mathbb{N}, \\ &\le \ \Big(\sum_{i=1}^n \big(\int_a^b |x_i(t)|^q \mathrm{d}t\big)^{\frac{p}{q}}\Big)^{\frac{1}{p}},\ p,\ q \in \mathbb{N}. \end{aligned} \tag{6.38}$$

Mit $p = q$ folgt weiter

$$||x||_\infty \ \le \ \Big(\sum_{i=1}^n \int_a^b |x_i(t)|^q \mathrm{d}t\Big)^{\frac{1}{q}}, \quad q \in \mathbb{N}. \tag{6.39}$$

Da die Betragsfunktion nicht differenzierbar ist, betrachten wir nur geradzahlige q, d. h. $q = 2\tilde{q},\ \tilde{q} \in \mathbb{N}$, notieren im folgenden aber wieder q statt $\tilde{q}$.

Für Systeme von n Differentialgleichungen und für $x_{\mathrm{app}} \in \mathrm{S}_\Delta^{2+k},\ k = 0$ bzw. 2, können also die Terme der Restglieder in (6.13) angenähert werden durch

$$r_{k,i} \ = \ \max_{1 \le j \le n} \sup_{t_i \le t \le t_{i+1}} |x^{(2+k)}(t)| \ \doteq \ \Big(\sum_{j=1}^n \int_{t_i}^{t_{i+1}} (x_j^{(2+k)}(t))^{2q} \mathrm{d}t\Big)^{\frac{1}{2q}},\ q \in \mathbb{N}. \tag{6.40}$$

Mit den Beziehungen aus den vorherigen Abschnitten und wegen

$$r_{k,i} \ \doteq \ \Big(\sum_{j=1}^n \int_{t_i}^{t_{i+1}} (x_j^{(2+k)}(t))^{2q} \mathrm{d}t\Big)^{\frac{1}{2q}} \qquad \doteq \ \Big(\sum_{j=1}^n \int_{t_i}^{t_{i+1}} (x_j^{(2+k)}(t_{i+1/2}))^{2q} \mathrm{d}t\Big)^{\frac{1}{2q}} \tag{6.41}$$

$$= \ \Big(\sum_{j=1}^n (x_j^{(2+k)}(t_{i+1/2}))^{2q} \int_{t_i}^{t_{i+1}} \mathrm{d}t\Big)^{\frac{1}{2q}} \ = \ \Big(\sum_{j=1}^n (x_j^{(2+k)}(t_{i+1/2}))^{2q} h_i\Big)^{\frac{1}{2q}},\ q \in \mathbb{N}, \tag{6.42}$$

können nun die Gleichungen (6.14), die die optimale Lage der inneren Gitterpunkte $(t_i)_{i=2}^{N-1}$ für $x_{\mathrm{app}} \in \mathrm{S}_\Delta^{2+k},\ k = 0$ bzw. 2, festlegen, durch folgende Gleichungen angenähert werden:

$$\begin{aligned} 0 \ &= \ h_{i+1}^{1+k} r_{k,i+1} - h_i^{1+k} r_{k,i}, \quad i = 1, \ldots, N-2 \\ &\doteq \ h_{i+1}^{1+k} \Big(\sum_{j=1}^n h_{i+1} (x_j^{(2+k)}(t_{i+3/2}))^{2q}\Big)^{\frac{1}{2q}} - h_i^{1+k} \Big(\sum_{j=1}^n h_i (x_j^{(2+k)}(t_{i+1/2}))^{2q}\Big)^{\frac{1}{2q}}. \end{aligned} \tag{6.43}$$

Dabei werden die Werte von $x^{(2+k)}(t_{i+1/2}),\ i = 1, \ldots, N-1$, aus den Gleichungen (6.29) bzw. (6.30) und (6.31) bestimmt.

Da die Ausdrücke in $(.)^{1/2q}$ von den Parametern der jeweiligen Diskretisierungen abhängen, und da für $h \to 0$ auch $(.)^{1/2q} \to 0$ geht, ist $(.)^{1/2q}$ im Grenzfall nicht differenzierbar. In der Praxis wirkt sich Nichtdifferenzierbarkeit bereits in einer „kleinen" Umgebung der Null aus ($(.)^{1/2q} \approx 0$). Es ist daher numerisch günstiger, den folgenden äquivalenten Satz von Gleichungen zu verwenden:

$$\boxed{\begin{array}{rcl} 0 & = & h_{i+1}^{2q(1+k)+1}\left(\sum_{j=1}^{n}\left(x_j^{(2+k)}(t_{i+3/2})\right)^{2q}\right) - h_i^{2q(1+k)+1}\left(\sum_{j=1}^{n}\left(x_j^{(2+k)}(t_{i+1/2})\right)^{2q}\right) \\ i & = & 1,\ldots,N-2. \end{array}} \tag{6.44}$$

Bemerkung 6.7 Für die praktische Anwendung in einem Optimierungsverfahren kommen im wesentlichen $q = 1, 2, 3$ in Frage. Dabei ergibt sich für $q = 3$ zwar eine bessere Approximation der ursprünglichen Gleichungen, für $q = 3$ sind die Gleichungen jedoch viel „stärker" nichtlinear als für $q = 1$. Dies kann ein erheblich schlechteres Konvergenzverhalten des Optimierungsverfahrens nach sich ziehen, insbesondere dann, wenn, wie in unserem Fall, partielle Ableitungen dieser Gleichungen durch numerische Differenzenquotienten approximiert werden (s. Kapitel 7.2).

Bemerkung 6.8 In der Praxis ist es wichtig sicherzustellen, daß während einzelner Iterationsschritte im Optimierungsverfahren die geordnete Reihenfolge der t_j nicht verändert wird. Dies kann durch die Hinzunahme von zusätzlichen Nebenbedingungen wie

$$0 \;\le\; h_{\min} \;\le\; h_j = t_{j+1} - t_j \;\le\; h_{\max}, \quad j = 1, \ldots, N-1, \tag{6.45}$$

geschehen, wobei $h_{\min}(N)$ und $h_{\max}(N)$ so zu wählen sind, daß

$$\frac{h_{\max}}{h_{\min}} \;\le\; K = \text{const.} < +\infty \quad \text{und} \quad h_{\max} \to 0 \;\text{ für }\; N \to \infty.$$

Diese linearen Nebenbedingungen müssen in *jedem* Iterationsschritt eingehalten werden. Der verwendete Algorithmus NPSOL (s. Kapitel 7.3.3) garantiert das.

6.3 Bewegliche Gitteranpassung zur Gleichverteilung der Variation

Bei der numerischen Lösung von Randwertproblemen gewöhnlicher Differentialgleichungen mit Differenzenverfahren konnte die Plazierung der Gitterpunkte zur Gleichverteilung der „Variation“[1], d. h. (für $n = 1$)

$$\int_{t_i}^{t_{i+1}} |\dot{x}(t)| \, \mathrm{d}t \;=\; \text{const.}, \quad i = 1, \ldots, N-1, \tag{6.46}$$

mit gewissem Erfolg verwendet werden (Keller zitiert in [14]). Dabei bezeichnet x die Lösung des Randwertproblems (6.1), (6.2). Bei der Diskretisierung (DP2) des Optimalsteuerungsproblems (P) kann aus Gleichung (1.1) die Näherung

$$\int_{t_i}^{t_{i+1}} \sum_{j=1}^{n} |\dot{x}_j(t)| \, \mathrm{d}t \;\doteq\; \int_{t_i}^{t_{i+1}} \sum_{j=1}^{n} |f_j(x_{\text{app}}(t), u_{\text{app}}(t), t)| \, \mathrm{d}t \tag{6.47}$$

gewonnen werden. Die rechte Seite von (6.47) kann durch

$$\sqrt{\sum_{j=1}^{n} \int_{t_i}^{t_{i+1}} (f_j(x_{\text{app}}(t), u_{\text{app}}(t), t))^2 \, \mathrm{d}t} \;\doteq\; \sqrt{\sum_{j=1}^{n} \int_{t_i}^{t_{i+1}} \frac{1}{4} \left(f_j|_{t=t_i} + f_j|_{t=t_{i+1}} \right)^2 \mathrm{d}t} \tag{6.48}$$

$$\doteq\; \frac{1}{2} \sqrt{h_i \sum_{j=1}^{n} \left(f_j|_{t=t_i} + f_j|_{t=t_{i+1}} \right)^2} \;=:\; v_i \tag{6.49}$$

approximiert werden. Die Lage der inneren Gitterpunkte $(\tau_i)_{i=2}^{N-1}$ kann in der Optimierung nun formal durch die Gleichungsbeschränkungen

$$\boxed{v_{i+1}^2 - v_i^2 \;=\; 0, \quad i = 1, \ldots, N-2} \tag{6.50}$$

festgelegt werden. Diese sind differenzierbar nach den Diskretisierungsparametern.

Bemerkung 6.9 Die Auswertung der Bedingungen (6.50) läßt sich effizient implementieren: Die Differentialgleichungen f werden nämlich bereits zur Formulierung der Kollokationsbedingungen im diskretisierten Problem (DP2) bei t_i berechnet.

[1] Bezeichnung nach [14]

6.4 Beispiel: Optimale Bahnverfolgung für einen zweiachsigen Modellroboter

Der Greifer eines zweiachsigen Roboters mit vereinfachter Dynamik, im folgenden Modellroboter genannt, soll mit einer konstanten Geschwindigkeit eine vorgeschriebene Bahn abfahren, die eine nichtglatte Ecke enthält. Dabei entsteht ein unvermeidlicher Bahnfehler, da der Greifer dem Verlauf der Ecke nur dann exakt folgen kann, wenn er dort zur Ruhe kommt. Der Bahnfehler soll minimiert werden.

Das dynamische Verhalten des Modellroboters wird durch

$$\begin{array}{rclrclrclrclrcl} \ddot{q}_1(t) &=& u_1(t), & q_1(0) &=& 0, & \dot{q}_1(0) &=& 0.5, & q_1(2) &=& 0.5, & \dot{q}_1(2) &=& 0, \\ \ddot{q}_2(t) &=& u_2(t), & q_2(0) &=& 0, & \dot{q}_2(0) &=& 0, & q_1(2) &=& 0.5, & \dot{q}_1(2) &=& 0.5, \end{array} \tag{6.51}$$

beschrieben. Der zu minimierende Bahnfehler ist

$$J[u] \;:=\; \int_0^2 \sum_{i=1}^{2} (q_i(t) - q_{i,\mathrm{soll}}(t))^2 + \sum_{i=1}^{2} (\dot{q}_i(t) - \dot{q}_{i,\mathrm{soll}}(t))^2 \mathrm{d}t \;\to\; \min! \tag{6.52}$$

Als Sollbahn wird

$$\begin{array}{rclrcl} q_{1,\mathrm{soll}}(t) &=& \begin{cases} t/2, & 0 \le t \le 1, \\ 1/2, & 1 < t \le 2, \end{cases} & q_{2,\mathrm{soll}}(t) &=& \begin{cases} 0, & 0 \le t \le 1, \\ (t-1)/2, & 1 < t \le 2, \end{cases} \\ \dot{q}_{1,\mathrm{soll}}(t) &=& \begin{cases} 1/2, & 0 \le t \le 1, \\ 0, & 1 < t \le 2, \end{cases} & \dot{q}_{2,\mathrm{soll}}(t) &=& \begin{cases} 0, & 0 \le t \le 1, \\ 1/2, & 1 < t \le 2 \end{cases} \end{array} \tag{6.53}$$

verwendet. Als Startwerte für DIRCOL sind für q und $\dot{q}$ lineare Interpolationen der vorgeschriebenen Randwerte ausreichend. Die Anfangsschätzungen der Steuerungen werden auf null gesetzt. Ausgehend von diesen Startwerten werden zwei mit jeweils 7 Gitterpunkten diskretisierte Probleme (DP2) gelöst. Das erste Mal werden die Gitterpunkte äquidistant und fest gewählt, das zweite Mal wird die Lage der Gitterpunkte durch Gleichverteilung der Variation mitoptimiert. In Tabelle 6.1 ist das Konvergenzverhalten wiedergegeben. Um den Preis einer höheren Rechenzeit kann mit variablen Gitterpunkten eine deutlich verbesserte Approximation der Sollbahn erreicht werden (Abbildung 6.1). Drei Kriterien zur Gitteranpassung werden in Abbildung 6.2 gegenübergestellt

1. die lokale Variation v_i^2 (Abschnitt 6.3),
2. der lokale Optimalitätsfehler ω_i (Abschnitt 5.4) und
3. der Defekt d, eine stückweise lineare Funktion, die aus den Werten von $d(t) \;=\; \max_j\{|f_j(x_{\mathrm{app}}(t), u_{\mathrm{app}}(t), t) - \dot{x}_{j,\mathrm{app}}(t)|,\; j = 1, \ldots, n\}$ bei $t = t_{i+k/4}$, $k = 0, 1, 2, 3, 4$, $i = 1, \ldots, N-1$ gebildet wird [137].

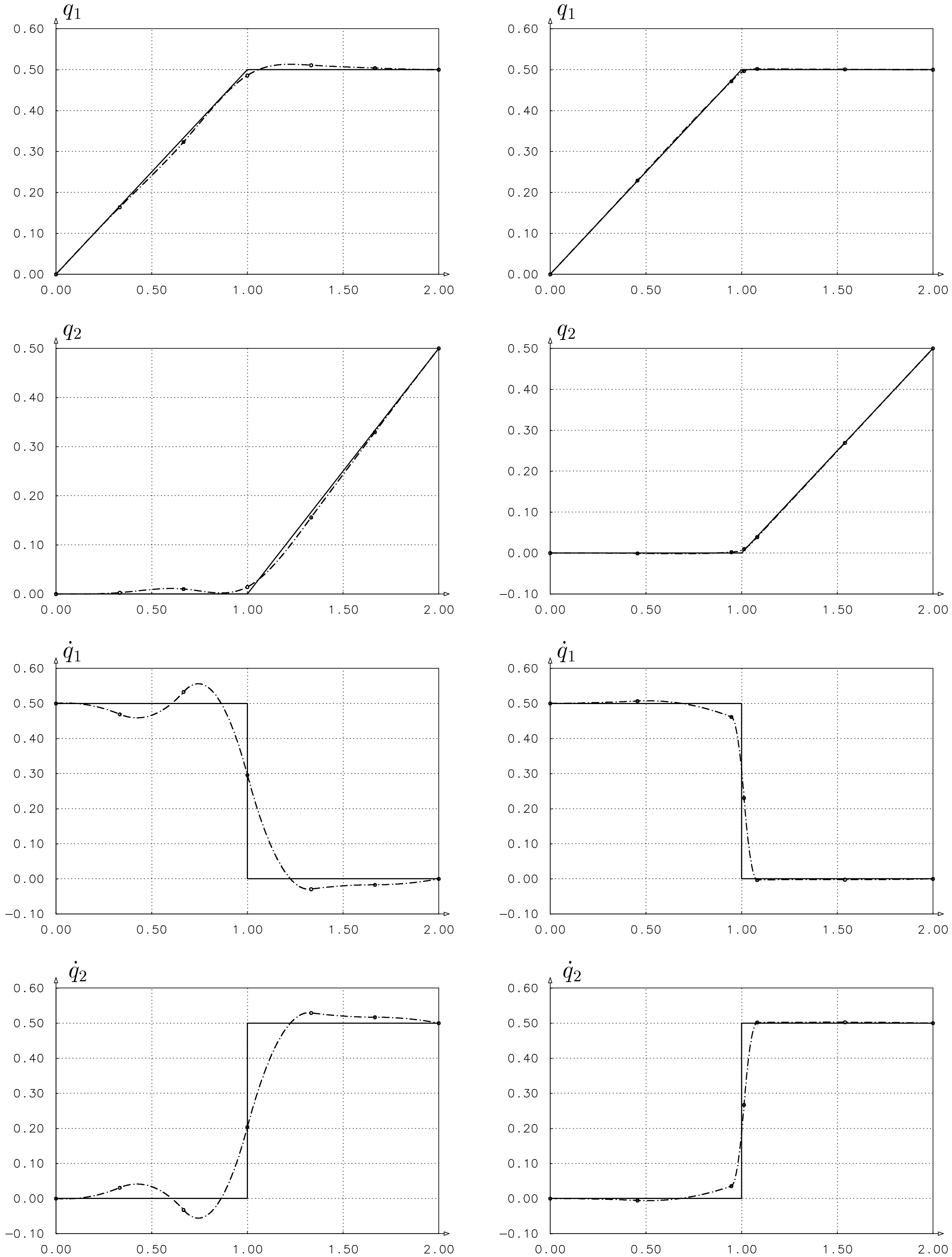

Abbildung 6.1: Vergleich der Lösungen für 7 feste äquidistante (links) und 7 variable, die Variation gleichverteilende (rechts) Gitterpunkte (– – –) mit der Sollbahn (———) des Modellroboters.

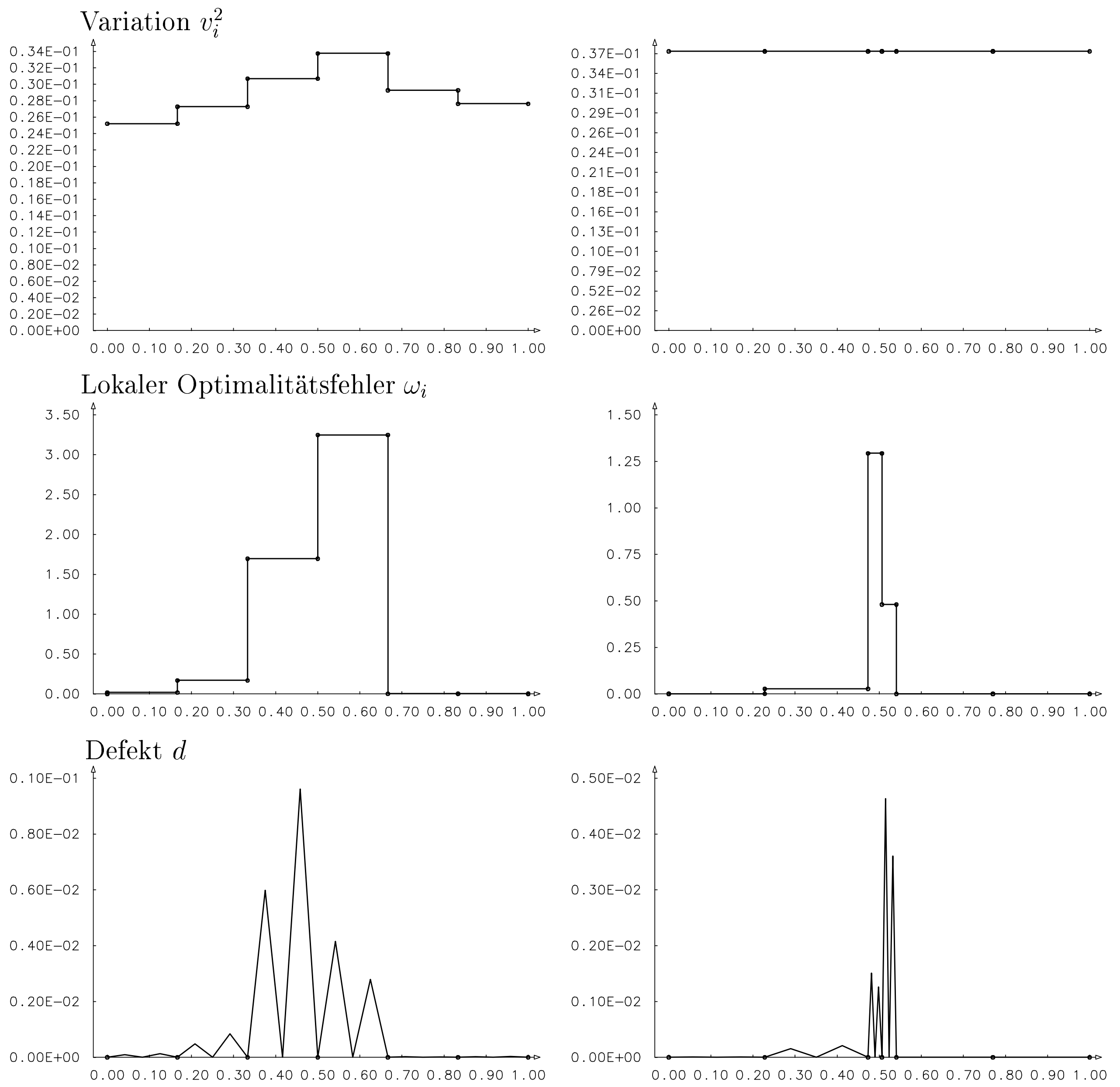

Abbildung 6.2: Vergleich der (skalierten) Fehlerkriterien für 7 feste, äquidistante (links) und 7 variable, die Variation gleichverteilende (rechts) Gitterpunkte beim Bahnverfolgungsproblem für den Modellroboter (Die Ordinaten sind unterschiedlich skaliert).

Tabelle 6.1: Konvergenzverhalten für 7 feste und 7 variable, die Variation gleichverteilende Gitterpunkte ($\epsilon_{\mathrm{OPT}} = 10^{-5}$, $\epsilon_{\mathrm{NFT}} = 10^{-5}$, Abkürzungen s. Kapitel 8.1).

NG	NY	NEQ	NIT	CPU	J
7 (fest)	41	30	41	8	5.973
7 (variabel)	46	35	58	17	1.917

6.5 Anpassung der Gitterpunkte in der Literatur

6.5.1 Statische Gitteranpassung bei Kollokationsverfahren

Eine sehr verbreitete Strategie zur Gitteranpassung beruht auf folgendem:

- Ein festes Gitter $(t_i)_{i=1}^{N}$ wird vorgegeben,
- das zugehörige diskretisierte Problem wird gelöst,
- die berechnete Lösung wird analysiert und dementsprechend ein neues, festes Gitter $(t_i^\star)_{i=1}^{N^\star}$ gewählt, mit dem
- das neue diskretisierte Problem gelöst wird.

Dieses Schema wird solange wiederholt, bis die geforderten Genauigkeiten erreicht sind. Dieses Vorgehen ist wohl zum ersten Mal in [14] beschrieben worden. Es wird u. a. in [7], [119] und [120] ausführlicher untersucht und ist im Kollokationsverfahren COLSYS [6] bzw. COLNEW [8] zur Lösung von Randwertproblemen erfolgreich implementiert worden.

Eine andere Strategie zur Gitteranpassung, die vom Interpolationsfehler ausgeht, hat Maier [89] mit Erfolg verwendet: Mit einem festen vorgegebenen Gitter werden einige wenige Iterationen des Newton-Verfahrens durchgeführt, bevor das Gitter endgültig festgelegt und die Iterationen zu Ende geführt werden. Dieses Verfahren funktioniert umso besser, je näher man bereits der Lösung des diskretisierten Problems ist. Zu Näherungswerten, die noch weit von der Lösung entfernt sind, wird meist kein optimales Gitter produziert.

6.5.2 Bewegliche Gitteranpassung

In den gebräuchlichen Kollokationsverfahren zur Lösung von Randwertproblemen werden meist Gitteranpassungsstrategien für statische Gitterpunkte verwendet.

Bei Kollokationsverfahren lassen sich bewegliche Gitterpunkte durch Koordinatentransformation der unabhängigen Veränderlichen einführen. Die Anzahl der Differentialgleichungen wird um eins erhöht. Das neue Gitter ist äquidistant. Die Lage des Gitters in der alten unabhängigen Veränderlichen wird durch die Lösung des Randwertproblems bestimmt. Dieses Vorgehen wird u. a. von Ascher et al. [7] und Russell [119] ausführlicher beschrieben. Es finden sich dort auch weitere Verweise.

Zur beweglichen Gitterpunktwahl bei diskretisierten optimalen Steuerungsproblemen gibt es nur wenige Arbeiten. Jänsch und Paus [69] beispielsweise gehen zunächst von der Diskretisierung von Hargraves und Paris [57] aus und verwenden die Lagen der inneren Gitterpunkte $t_2, \ldots, t_{N-1}$ als zusätzliche Variablen, *ohne* sie allerdings durch zusätzliche Nebenbedingungen in Gleichungsform formal festzulegen. Sie beobachten, daß sich bei der numerischen Lösung die Gitterpunkte so verschieben, daß sie den Ungleichungsbeschränkungen ausweichen. Um diesen Effekt zu beheben, modifizieren sie die im ursprünglichen Ansatz punktweisen Ungleichungsbeschränkungen. Dieses Vorgehen enthält aber zwei signifikante Nachteile, die die Konvergenz der Lösung des diskretisierten Problems gegen die Lösung der notwendigen Bedingungen des kontinuierlichen Optimalsteuerungsproblems für $h \to 0$ verhindern und die in

[69] nicht berücksichtigt wurden: Es ist unklar, ob und gegen was die zusätzlichen Freiheitsgrade der inneren Gitterpunkte für $h \to 0$ konvergieren, bzw. was diesen im kontinuierlichen Problem entspricht. Die Umformulierung der Diskretisierung von Ungleichungsbeschränkungen zerstört die entsprechenden Konvergenzeigenschaften der ursprünglichen Diskretisierung bei beschränkten Steuerungsproblemen: Aus der Theorie der optimalen Steuerungen ist bekannt, daß eine aktive Zustandsbeschränkung längs eines Randstückes eine Steuerung festlegt. Diese Eigenschaft muß auch das diskretisierte Problem durch *geeignete Verknüpfung der Diskretisierung der Zustandsbeschränkungen mit der Diskretisierung der Steuerungen* widerspiegeln. Die numerische Lösung dieses diskretisierten Problems *kann* also möglicherweise die Lösung des kontinuierlichen Problems *nicht* mehr hinreichend genau approximieren!

Zur Diskretisierung von optimalen Steuerungsproblemen bei differential-algebraischen Systemen aus der Chemie verwenden Cuthrell und Biegler [39] orthogonale Polynome p-ten Grades für $x(t)$ und $(p-1)$-ten Grades für $u(t)$. Die Gitterpunkte werden als zusätzliche Variablen hinzugenommen und sollen durch zusätzliche Gleichungsnebenbedingungen bestimmt werden. Zur Herleitung dieser Gleichungen verwenden sie die Fehlerabschätzungen von de Boor [14] und die dort angegebene Approximation der Restglieder $q_i,\ i = 1, \ldots, N-1$. Außerdem erweitern sie die so erhaltenen Gleichungen auf Systeme von Differentialgleichungen und erhalten schließlich als bestimmende Gleichungen

$$h_{i+1}\Big(\sum_{j=1}^{n}(q_{j,i+1}^{\frac{1}{p+1}})^2\Big)^{\frac{1}{2}} - h_i\Big(\sum_{j=1}^{n}(q_{j,i}^{\frac{1}{p+1}})^2\Big)^{\frac{1}{2}} \;=\; 0, \quad i = 1, \ldots, N-1. \tag{6.54}$$

In den in [39] angegebenen Beispielen wurde $p = 2$ verwendet. Dieses Vorgehen hat 3 wichtige Nachteile:

- Die gewählte Approximation der Restglieder $q_i,\ i = 1, \ldots, N-1$, ist nach den Diskretisierungsparametern nicht differenzierbar.
- Die durchgeführte Erweiterung (6.54) auf Systeme von Differentialgleichungen ist bei $q_{j,i} = 0$ nicht differenzierbar. Praktische Schwierigkeiten treten bereits auf, falls $q_{j,i} \approx 0$.
- Fehler in den nichtlinearen Gleichungsbeschränkungen werden nicht berücksichtigt.

Dieses Vorgehen ist in dieser Form nicht geeignet. Im übrigen sind Konvergenzeigenschaften der Diskretisierungen für $h \to 0$ nicht untersucht worden.