Quadratmittelapproximation durch B-Splines mit freien Knoten und Ungleichheitsnebenbedingungen an Ableitungen

Torsten Schütze*
Institut für Numerische Mathematik
Technische Universität Dresden

25. Juni 1995

*Forschungsbericht, angefertigt im Rahmen des DFG-Projektes Schm 968/2-1

Inhaltsverzeichnis

1 Einleitung **1**

2 B-Splines – Definitionen, Grundlagen **4**
2.1 B-Splines auf unendlichem Gitter 4
2.2 Spezialisierung auf ein endliches Intervall 5
2.3 Ableitung eines Splines nach seinem Argument 6
2.4 Ableitung eines B-Splines nach den Knoten 7
2.5 Ableitung eines Splines nach den Knoten 8

3 Problemstellung **11**
3.1 Formulierung des Funktionals 11
3.2 Die Anordnungsbedingung für die freien Knoten 15
3.2.1 Wahl eines absoluten Distanzmaßes 15
3.2.2 Wahl eines relativen Distanzmaßes 16
3.3 Nebenbedingungen an Ableitungen des Splines 17
3.3.1 Positivitätsforderung an die Ableitung des Splines 18
3.3.2 Einfache Schranken für die Ableitung des Splines 18
3.3.3 Strikte Konsistenz 20
3.4 Formulierung der Nebenbedingungen 23
3.4.1 Nebenbedingungen an freie Knoten 23
3.4.2 Nebenbedingungen an Ableitungen des Splines 23
3.5 Formulierung des kompletten Problems 23

4 Die Struktur der Quadratmittelprobleme **25**
4.1 Ausgangsproblem, reduziertes Problem, Subproblem (A) 25
4.2 Erste Bemerkungen zur Lösung des Problems 26
4.3 Qualitative Analyse von Subproblem (A) 27
4.4 Quantitative Analyse von Subproblem (A) 28
4.5 Analyse des reduzierten Problems 30
4.6 Die Beziehung zwischen Ausgangs- und reduziertem Problem 31

5 Restringierte semilineare Quadratmittelprobleme **33**
5.1 Der Gradient des reduzierten Funktionals 34
5.2 Die Hesse-Matrix des reduzierten Funktionals 34
5.3 Die Residiumsfunktion des reduzierten Funktionals 36
5.4 Die Jacobi-Matrix des reduzierten Funktionals 37
5.5 Struktur der Residiumsfunktion und der Jacobi-Matrix 37

6 Anwendung auf Spline-Approximation und Spline-Glättung **40**
6.1 Die Voraussetzungen des Hauptsatzes der Sensitivitätstheorie 41
6.1.1 Differenzierbarkeit der Problemfunktionen 42
6.1.2 Hinreichende Optimalitätsbedingungen zweiter Ordnung . . . 43
6.1.3 Lineare Unabhängigkeit der Gradienten der aktiven Restriktionen 43
6.1.4 Strikte Komplementarität 45
6.2 Existenz einer Lösung des reduzierten Problems 46
6.3 Äquivalenz von Ausgangsproblem und reduziertem Problem 47

7 Numerische Lösung des reduzierten Problems im Fall der Spline-Approximation 51
7.1 Die Kaufman-Jacobi-Matrix im Fall der Spline-Approximation . . . 51
7.2 Berechnung der Matrizen, in denen Ableitungen auftreten 52
7.2.1 Ableitung der Beobachtungsmatrix nach den Knoten 52
7.2.2 Ableitung der Nebenbedingungen nach den Knoten 53
7.3 Berechnung einer Nullraumbasis $\mathbf{N}$ bzw. von Vektoren $\bar{\mathbf{R}}^{+}\mathbf{y}$ 55
7.4 Berechnung des Projektors $\mathbf{P}_{BN}^{\perp}$ bzw. von $\mathbf{P}_{BN}^{\perp}\mathbf{y}$ 58
7.5 Ein Algorithmus zur Berechnung der Kaufman-Jacobi-Matrix 60
7.6 Ein Algorithmus zur Lösung des reduzierten Problems 61

8 Numerische Lösung des reduzierten Problems im Fall der Spline-Glättung 64
8.1 Die Kaufman-Jacobi-Matrix im Fall der Spline-Glättung 64
8.2 Berechnung der Matrizen, in denen Ableitungen auftreten 65
8.2.1 Ableitung der Glättungsmatrix nach den Knoten 65
8.3 Berechnung einer Nullraumbasis $\mathbf{N}$ bzw. von Vektoren $\bar{\mathbf{R}}^{+}\mathbf{y}$ 67
8.4 Berechnung des Projektors $\mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N}$ bzw. von $\mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N}\mathbf{y}$ 67
8.5 Ein Algorithmus zur Berechnung der Kaufman-Jacobi-Matrix 70
8.6 Ein Algorithmus zur Lösung des reduzierten Problems 72

9 Numerische Tests 74
9.1 Titan-Daten 74
9.2 Arctan-Daten 77

Literatur 81

1 Einleitung

Gegeben seien Meßdaten $\{x_i, y_i\}$ $(i = 1, \ldots, m)$, wobei $\{x_i\}$ Abszissen mit

$$a \leq x_1 < \cdots < x_i < x_{i+1} < \cdots < x_m \leq b$$

und $\{y_i\}$ fehlerbehaftete Meßwerte einer unbekannten glatten Funktion $g \in W_2^q[a, b]$ sind, d.h.

$$y_i = g(x_i) + \epsilon_i \qquad (i = 1, \ldots, m).$$

Der Fehler der i-ten Messung sei mit ϵ_i bezeichnet.

Diese Funktion g soll durch eine Funktion s aus einem geeignet gewählten Teilraum $\mathcal{S} \subset W_2^q[a, b]$ approximiert werden. An die Funktion s stellen wir folgende Forderungen:

- s kann leicht berechnet werden,
- s kann durch wenige Parameter beschrieben werden (Datenreduktion!),
- s ist eine glatte Funktion, welche die Funktion g innerhalb des Fehlerniveaus der Daten repräsentiert,
- s ist eine formerhaltende Approximation an die Funktion g, z.B. $s^{(p)}(x) \geq 0$ für alle $x \in [a, b]$ falls $g^{(p)}(x) \geq 0$ für alle $x \in [a, b]$, $p \in \{0, \ldots, q\}$.

Wenn die Fehler ϵ_i stochastisch unabhängig und normalverteilt sind, so stellt der Quadratmittelfehler

$$\varphi(s) := \frac{1}{2} \sum_{i=1}^{m} [y_i - s(x_i)]^2$$

ein geeignetes Maß für die Approximationsgüte dar. Den Term $\varphi(s)$ bezeichnen wir im folgenden als *Approximationsterm.*

Bei der praktischen Anwendung von Approximationsschemata ist oftmals ein Kompromiß zwischen der Güte der Approximation und der Glattheit gefragt. Wir betrachten daher zusätzlich den *Glättungsterm*

$$\rho(s) := \frac{1}{2} \left\| s^{(r)} \right\|_{L_2}^2 = \frac{1}{2} \int_a^b \left[s^{(r)}(x) \right]^2 dx$$

mit festem $r \in \{0, \ldots, q\}$ und das zugehörige Minimierungsproblem

$$\phi(s) := \varphi(s) + \mu\rho(s) = \frac{1}{2} \sum_{i=1}^{m} [y_i - s(x_i)]^2 + \mu \frac{1}{2} \int_a^b \left[s^{(r)}(x) \right]^2 dx \longrightarrow \min$$

mit dem *Glättungsparameter* $\mu > 0$.

Als endlichdimensionalen Teilraum $\mathcal{S}$ von $W_2^q[a, b]$ wählen wir den n-dimensionalen Raum $\mathcal{S}_{k,\mathbf{t}}$ der Polynomsplines der Ordnung $k \geq 1$ zu der Knotenfolge $\mathbf{t} = \{t_j\}$ mit

$$t_1 = \cdots = t_k = a < t_{k+1} \leq \cdots \leq t_n < b = t_{n+1} = \cdots = t_{n+k}.$$

Es sei $m \geq n$. Für praktisch relevante Aufgabenstellungen gilt gewöhnlich $m \gg n$.

Zur formerhaltenden Approximation mit festen Knoten existieren eine Reihe von Arbeiten und effizienten numerischen Verfahren: In Micchelli/Utreras [MU88] werden Existenz- und Charakterisierungsaussagen für das Problem der Splineinterpolation und -glättung in einer konvexen Teilmenge eines Hilberraumes behandelt. Elfving/Andersson [EA88] betrachten den Fall $r = 2$ und Konvexitätsnebenbedingungen $s''(x) \geq \delta(x)$, δ gegeben. Schmidt/Scholz [SS90] entwickeln ein effizientes Verfahren für $r = 2$ und die verallgemeinerten Konvexitätsnebenbedingungen $\delta(x) \leq s''(x) \leq \epsilon(x)$ (δ, ϵ - lineare $\mathbb{C}^0$-Splines), indem sie an Stelle des partiell separablen primalen Problems das unrestringierte duale Problem behandeln. In Schwetlick/Kunert [SK93]) wird schließlich das allgemeine Problem $\delta(x) \leq s^{(p)}(x) \leq \epsilon(x)$ (δ, ϵ - lineare $\mathbb{C}^0$-Splines; $p, r \in \{0, \ldots, q\}$ beliebig) mittels Orthogonalisierungstechniken gelöst.

Betrachtet man die Lage der Knoten als zusätzliche freie Parameter des Optimierungsprozesses, so kann die Approximation i.allg. wesentlich verbessert werden. Wir betrachten in dieser Arbeit ausschließlich die l_2-Approximation. Es existiert eine Fülle von Literatur zur L_∞-Approximation (siehe [Bra86], [Nür89] und die dort zitierte Literatur).

Eine wesentliche Schwierigkeit bei der Approximation mit freien Knoten stellt das Zusammenfallen von Knoten im Laufe des Optimierungsprozesses dar. Aus analytischen und numerischen Gründen (Differenzierbarkeit, Lethargiesyndrom) muß man das Zusammenfallen von Splineknoten verhindern. In der Einbeziehung dieser Nebenbedingung ($t_{j+1} - t_j > 0$) unterscheiden sich die verschiedenen entwickelten Verfahren.

In einer ersten Arbeit [dBR68] minimieren die Autoren zyklisch den L_2-Fehler im Intervall $[t_{j-1} + \epsilon(t_{j+1} - t_{j-1}), t_{j+1} - \epsilon(t_{j+1} - t_{j-1})]$ als Funktion des Knoten t_j. Jupp [Jup78] schlägt eine nichtlineare Transformation $\mu_j = \log \frac{t_{j+1} - t_j}{t_j - t_{j-1}}$ in ein unrestringiertes Problem vor. Nach Elimination der Splinekoeffizienten, welche als lineare Variable auftreten, wendet er verschiedene Standardverfahren der nichtlinearen Optimierung auf das reduzierte Problem an. Dierckx [Die79] schlägt eine Penalty-Technik mit dem Penalty-Term $\sum_{j=k}^{n} (t_{j+1} - t_j)^{-1}$ vor und wendet auf das reduzierte Problem den cg-Algorithmus von Fletcher/Reeves an.

Bei den obigen Verfahren werden stets alle Splineknoten in den Optimierungsprozeß einbezogen. Außerdem ist die erwähnte Reduktion auf ein Problem, in welchem nur die nichtlinearen Parameter $\mathbf{t}$ auftreten, nur zulässig, falls eine gewisse Regularitätsbedingung (Schoenberg-Whitney-Bedingung) in einer Umgebung der betrachteten Knotenfolgen erfüllt sind.

In der Arbeit Schwetlick/Schütze [SS95] wurde deshalb das Problem der Approximation mit freien Knoten mittels eines regularisierenden Glättungstermes behandelt. Das Problem wurde wiederum auf ein Problem in den freien Knoten reduziert, wobei die freien Knoten jetzt eine echte Teilmenge der inneren Splineknoten sein können. Das reduzierte Problem wurde mittels Verfahren für separable Quadratmittelprobleme gelöst. Die linearen Nebenbedingungen an die freien Knoten wurden direkt einbezogen in den Lösungsalgorithmus.

Suchomski [Suc91] formuliert die Regularitätsbedingung (Schoenberg-Whitney-Bedingung) im Falle der Splineinterpolation direkt als Nebenbedingung an die freien Knoten. Mittels einer nichtlinearen Variablentransformation ähnlich der von Jupp wird das Problem in ein unrestringiertes Quadratmittelproblem überführt. Die Separabilität des Quadratmittelproblems wird in dem entwickelten Verfahren ebenfalls ausgenutzt.

Das Ziel der vorliegenden Arbeit ist die Verbindung zwischen formerhaltender Approximation und Quadratmittelapproximation mit freien Knoten. Eine Teilmenge $\tilde{\mathbf{t}}$ der Knotenfolge $\mathbf{t}$ wird in den Optimierungsprozeß einbezogen, woraus ein nichtlineares Quadratmittelproblem in den Splinekoeffizienten und den freien Knoten resultiert. Man erhält lineare Ungleichheitsnebenbedingungen an die freien Knoten und lineare Ungleichheitsnebenbedingungen an die Splinekoeffizienten bzgl. fester $\tilde{\mathbf{t}}$. Die Arbeit beschäftigt sich mit Bedingungen dafür, daß man das Originalproblem auf ein Optimierungsproblem reduzieren kann, in welchem nur die freien Knoten als Variable auftreten. Im Gegensatz zu den nichtlinearen Nebenbedingugen des Originalproblems besitzt das reduzierte Problem lediglich lineare Ungleichheitsnebenbedingungen. Bei der Betrachtung des reduzierten Problems für eine feste Knotenfolge $\mathbf{t}$ können Ergebnisse aus [SK93] benutzt werden.

Zur Lösung des reduzierten Problems wird ein effizienter Algorithmus entwickelt. Ein Hauptanliegen bei der Entwicklung des Algorithmus war die weitgehende Ausnutzung spezieller Schwachbesetztheitsstrukturen im Rahmen der Approximation mit B-Splines sowie die Verwendung numerisch gutartiger Orthogonalisierungsverfahren.

Die Verbindung von formerhaltender Approximation und freien Knoten scheint neu zu sein.

2 B-Splines – Definitionen, Grundlagen

2.1 B-Splines auf unendlichem Gitter

Definition 1 (Knotenfolge)
Eine Folge $\mathbf{t} = (\ldots, t_{-1}, t_0, t_1, \ldots)$ *mit* $t_j \leq t_{j+1}$ *für alle* $j \in \mathbb{Z}$ *und* $\lim\limits_{j \to \pm\infty} t_j = \pm\infty$ *heißt (biinfinite) Knotenfolge.*

Definition 2 (Normalisierte polynomiale B-Splines)
Der j-te B-Spline der Ordnung k für die Knotenfolge $\mathbf{t}$ *wird mit* $B_{j,k}$ $(B_{j,k,\mathbf{t}})$ *bezeichnet und definiert durch:*

$$(2.1) \quad B_{j,1}(x) := \begin{cases} 1 & \textit{falls } t_j \leq x < t_{j+1}, \\ 0 & \textit{sonst} \end{cases}$$

$$(2.2) \quad B_{j,k}(x) := \omega_{j,k}(x) \cdot B_{j,k-1}(x) + (1 - \omega_{j+1,k}(x)) \cdot B_{j+1,k-1}(x) \quad \text{für} \quad k > 1$$

mit

$$(2.3) \quad \omega_{j,k}(x) := \begin{cases} \frac{x - t_j}{t_{j+k-1} - t_j} & \textit{falls } t_j < t_{j+k-1}, \\ 0 & \textit{sonst.} \end{cases}$$

Bemerkung 1
Die Bezeichnung $B_{j,k,\mathbf{t}}$ *bzw.* $B_{j,k}(x, \mathbf{t})$ *deutet dabei an, daß die B-Splines von der Knotenfolge* $\mathbf{t}$ *abhängig sind. Wenn aus dem Zusammenhang deutlich wird, auf welche Knotenfolge sich der B-Spline bezieht, verzichten wir gegebenenfalls auf die gesonderte Kennzeichnung der Knotenfolge.*

Für $t_j = t_{j+k}$ erkennt man sofort $B_{j,k} \equiv 0$. Es ist deshalb vernünftig,

$$(2.4) \qquad t_j < t_{j+k} \qquad \forall j \in \mathbb{Z}$$

zu fordern, um die trivialen B-Splines $B_{j,k} \equiv 0$ von vornherein von der Betrachtung auszuschließen.

Definition 3 (Knotenvielfachheit)
Der Knoten t_j *hat die Vielfachheit* $\#t_j$*, wenn*

$$\exists i : \qquad t_i < t_j = t_{i+1} = \cdots = t_{i+\#t_j} < t_{i+\#t_j+1}.$$

Definition 4 (Spline, Splineraum)
Die Menge $\mathcal{S}_{k,\mathbf{t}} := \left\{ \sum_j B_{j,k,\mathbf{t}} \alpha_j \, , \, \alpha_j \in \mathbb{R} \right\}$ *heißt Splineraum der Ordnung k bezüglich der Knotenfolge* $\mathbf{t}$*. Die Elemente s des Splineraumes* $\mathcal{S}_{k,\mathbf{t}}$ *heißen Splines.*

Den angekündigten Zusammenhang zwischen Knotenvielfachheit und Glattheit sowie die lineare Unabhängigkeit der B-Splines zeigt das folgende Theorem:

Theorem 1 (Curry, Schoenberg)
Wenn $t_j < t_{j+k}$ *für alle j gilt, dann sind die B-Splines* $B_{j,k,\mathbf{t}}$ *linear unabhängig und bilden eine Basis für den Raum* $\tilde{\mathcal{S}}_{k,\mathbf{t}}$ *aller stückweisen Polynome vom Grad* $k-1$ *mit Breakpoints* t_j*, welche in den Breakpoints* $(k - 1 - \#t_j)$*-mal stetig differenzierbar sind.*
Anzahl der Glattheitsbedingungen in t_j + *Knotenvielfachheit* $\#t_j$ = *Ordnung k*

2.2 Spezialisierung auf ein endliches Intervall

Da jeder B-Spline nur einen endlichen Träger besitzt, genügt es, sich auf ein endliches Intervall zu beschränken. Wir betrachten daher im folgenden ausschließlich die (finite) Knotenfolge

$$\mathbf{t} = (t_1, \ldots, t_{n+k}). \tag{2.5}$$

Dabei sei n die Anzahl der B-Splines.

Später werden wir B-Splines benutzen, um gegebene Daten $\{x_i, y_i\}$ $(i = 1, \ldots, m)$, zu approximieren. Die *Meßstellen* x_i befinden sich in einem endlichen Intervall $x_i \in [a, b] \subset \mathbb{R}$. Wir beschränken uns daher auf die Knotenfolge

$$\mathbf{t} = (t_1, \ldots, t_{n+k}) \tag{2.6}$$

mit

$$t_1 \leq \cdots \leq t_k \leq a \quad < \quad b \leq t_{n+1} \leq \cdots \leq t_{n+k} \tag{2.7}$$

und $t_{k+1} \leq \cdots \leq t_n \in (a, b)$. Die Knoten $t_{k+1}, \ldots, t_n$ werden als innere Knoten bezeichnet. Es gelte stets $t_j < t_{j+k}$ $(j = 1, \ldots, n)$, d.h. $\#t_j < k$ für alle j. Stärkere Glattheitsforderungen (etwa einfache innere Knoten, $\#t_j = 1$, $j = k+1, \ldots, n$) werden an späterer Stelle diskutiert.

Für die äußeren Knoten $t_1, \ldots, t_k$ und $t_{n+1}, \ldots, t_{n+k}$ gelte

$$t_1 = t_2 = \cdots = t_k = a \qquad \text{und} \qquad b = t_{n+1} = t_{n+2} = \cdots = t_{n+k}. \tag{2.8}$$

Wir betrachten jetzt die Einschränkung von $\mathcal{S}_{k,\mathbf{t}}$ auf $[a, b]$.

$$\mathcal{S}_{k,\mathbf{t}_{|[a,b]}} = \left\{ \sum_{j=1}^{n} B_{j,k,\mathbf{t}} \alpha_j \,,\ \alpha_j \in \mathbb{R} \right\}$$

Statt $\mathcal{S}_{k,\mathbf{t}_{|[a,b]}}$ werden wir im folgenden kurz $\mathcal{S}_{k,\mathbf{t}}$ schreiben. Für einen Spline $s \in \mathcal{S}_{k,\mathbf{t}}$ gibt es eine Darstellung

$$s(x) = \sum_{j=1}^{n} B_{j,k}(x, \mathbf{t}) \alpha_j \tag{2.9}$$

oder in Vektorschreibweise

$$s(x) = \boldsymbol{\beta}(x, \mathbf{t})^T \boldsymbol{\alpha} \tag{2.10}$$

mit

$$\boldsymbol{\beta}(x, \mathbf{t}) := (B_{1,k}(x, \mathbf{t}), \ldots, B_{n,k}(x, \mathbf{t}))^T \in \mathbb{R}^n \tag{2.11}$$

und

$$\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{R}^n. \tag{2.12}$$

Die Zahlen $\alpha_1, \ldots, \alpha_n$ heißen Splinekoeffizienten.

Bemerkung 2

(i) *Unter der Voraussetzung $t_j < t_{j+k}$ $(j = 1, \ldots, n)$ ist die Dimension von $\mathcal{S}_{k,\mathbf{t}}$ gleich n und die Darstellung (2.9) ist eindeutig.*

(ii) Nach Definition 2 sind die B-Splines rechtsstetig. Am rechten Intervallende definieren wir daher $B_{j,k}(b) := B_{j,k}(b-)$.

Bei der Quadratmittelapproximation mit B-Splines spielt die Struktur folgender Matrix, der sogenannten *Beobachtungsmatrix*, eine Rolle:

$$\mathbf{B} = \mathbf{B}(\mathbf{t}) = \begin{bmatrix} B_{1,k}(x_1, \mathbf{t}) & \dots & B_{n,k}(x_1, \mathbf{t}) \\ \vdots & & \vdots \\ B_{1,k}(x_m, \mathbf{t}) & \dots & B_{n,k}(x_m, \mathbf{t}) \end{bmatrix} \in \mathbb{R}^{m,n}$$

Satz 1 (Schoenberg, Whitney)
Die Matrix $\mathbf{B} \in \mathbb{R}^{m,n}$ hat Vollrang n genau dann, wenn es eine Teilfolge der Datenabszissen (x_{i_j}) gibt, so daß

$$B_{j,k}\left(x_{i_j}, \mathbf{t}\right) \neq 0 \qquad j = 1, \dots, n. \tag{2.13}$$

Bemerkung 3
(i) Eine hinreichende Bedingung (und bei $t_j < t_{j+k}$ auch notwendige) Bedingung für (2.13) ist die Schoenberg-Whitney-Bedingung

$$t_j < x_{i_j} < t_{j+k}. \tag{2.14}$$

(ii) Die Matrix $\mathbf{B}$ hat eine (Zeilen-) Bandbreite kleiner oder gleich k.

2.3 Ableitung eines Splines nach seinem Argument

Lemma 2 (Ableitung eines Splines nach seinem Argument)
Gilt für die Knotenfolge $t_j < t_{j+k-1}$ $(j = 2, \dots, n)$, so ist die Ableitung eines Splines aus $\mathcal{S}_{k,\mathbf{t}}$ nach seinem Argument ein Spline der Ordnung $k-1$ bezüglich derselben Knotenfolge. Gilt $s(x) = \sum_{j=1}^{n} B_{j,k}(x, \mathbf{t})\alpha_j$, so besitzt s' die Darstellung

$$s'(x) = \sum_{j=2}^{n} B_{j,k-1}(x, \mathbf{t})\alpha'_j \tag{2.15}$$

mit

$$\alpha'_j := \frac{\alpha_j - \alpha_{j-1}}{\frac{t_{j+k-1} - t_j}{k-1}} \qquad (j = 2, \dots, n). \tag{2.16}$$

Durch wiederholte Anwendung von Lemma 2 erhält man die r-te Ableitung eines Splines. Wir setzen $t_j < t_{j+k-r}$ $(j = r+1, \dots, n)$ mit $0 \leq r < k$ voraus. Nach Lemma 2 existiert dann $s^{(r)}$ und hat die Darstellung

$$s^{(r)}(x) = \sum_{j=r+1}^{n} B_{j,k-r}(x, \mathbf{t})\alpha_j^{(r)} \tag{2.17}$$

$$= \boldsymbol{\beta}_r(x, \mathbf{t})^T \boldsymbol{\alpha}^{(r)} \tag{2.18}$$

mit

$$\boldsymbol{\beta}_r(x, \mathbf{t}) := \left(B_{r+1,k-r}(x, \mathbf{t}), \dots, B_{n,k-r}(x, \mathbf{t})\right)^T \in \mathbb{R}^{n-r} \tag{2.19}$$

und rekursiv bestimmbaren Splinekoeffizienten

$$\boldsymbol{\alpha}^{(r)} := \left(\alpha_{r+1}^{(r)}, \dots, \alpha_n^{(r)}\right)^T \in \mathbb{R}^{n-r}. \tag{2.20}$$

Die Splinekoeffizienten $\boldsymbol{\alpha}^{(r)}$ sind durch folgende Beziehung mit den Splinekoeffizienten $\boldsymbol{\alpha}$ verbunden:

$$\boldsymbol{\alpha}^{(r)} := \mathbf{D}_r \boldsymbol{\alpha} \tag{2.21}$$

mit

$$\mathbf{D}_0 := \mathbf{I} \in \mathbb{R}^{n,n} \tag{2.22}$$

$$\mathbf{D}_r := \mathbf{H}_r \mathbf{L}_r \dots \mathbf{H}_1 \mathbf{L}_1 \in \mathbb{R}^{n-r,n} \qquad r \geq 1. \tag{2.23}$$

Die Matrizen $\mathbf{H}_\nu$ und $\mathbf{L}_\nu$ sind definiert durch

$$\mathbf{H}_\nu := (k-\nu) \operatorname{diag} \left(\frac{1}{t_{k+j} - t_{\nu+j}} \right)_{j=1,\dots,n-\nu} \in \mathbb{R}^{n-\nu,n-\nu} \tag{2.24}$$

$$\mathbf{L}_\nu := \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n-\nu,n-\nu+1} \tag{2.25}$$

für $\nu = 1, \dots, r$.

In Vektorform erhalten wir also für die r-te Ableitung eines Splines

$$s^{(r)}(x) = \boldsymbol{\beta}_r(x, \mathbf{t})^T \mathbf{D}_r(\mathbf{t}) \boldsymbol{\alpha} \tag{2.26}$$

mit der, von der Knotenfolge $\mathbf{t}$ abhängigen, Matrix $\mathbf{D}_r(\mathbf{t}) = \mathbf{D}_r$.

2.4 Ableitung eines B-Splines nach den Knoten

Eine zu Definition 2 äquivalente Definition der B-Splines kann mittels dividierter Differenzen angegeben werden

$$B_{j,k}(x) := \begin{cases} (t_{j+k} - t_j)\,(-1)^k\,[t_j, \dots, t_{j+k}]\,(x-y)_+^{k-1} & \text{falls} \quad t_j < t_{j+k}, \\ 0 & \text{sonst.} \end{cases}$$

Sei $\boldsymbol{\tau}$ die feinste streng monoton wachsende Unterfolge von $\{t_j, \dots, t_{j+k}\}$

$$t_j \leq t_{j+1} \leq \dots \leq t_{j+k} = \underbrace{\tau_1}_{\nu_1} < \underbrace{\tau_2}_{\nu_2} < \dots < \underbrace{\tau_d}_{\nu_d},$$

d.h. τ_1 kommt genau $\nu_1 = \#t_j$-mal vor,

$\vdots$

τ_d kommt genau $\nu_d = \#t_{j+k}$-mal vor.

Jedem τ_i $(i = 1, \dots, d)$ der streng monoton wachsenden Folge ist also die entsprechende Vielfachheit ν_i zugeordnet.

Theorem 2 (Ableitung eines B-Splines nach den Knoten, Schumaker)
Sei $t_j \leq t_{j+1} \leq \cdots \leq t_{j+k} = \underbrace{\tau_1}_{\nu_1} < \underbrace{\tau_2}_{\nu_2} < \cdots < \underbrace{\tau_d}_{\nu_d}$. *Sei* $1 \leq i \leq d$ *fest und gelte* $\nu_i \leq k-2$. *Dann gilt*

$$(2.27) \quad \frac{\partial B_{j,k}(x)}{\partial \tau_i} =$$

$$\begin{cases} (-1)^{k-1} \begin{bmatrix} \nu_1+1, \nu_2, ..., \nu_d - 1 \\ \tau_1, \tau_2, ..., \tau_d \end{bmatrix} (x-y)_+^{k-1}, & \text{für } i = 1 \text{ und } \nu_1 = 1 \\ (-1)^{k} \begin{bmatrix} \nu_1 - 1, ..., \nu_{d-1}, \nu_d + 1 \\ \tau_1, ..., \tau_{d-1}, \tau_d \end{bmatrix} (x-y)_+^{k-1}, & \text{für } i = d \text{ und } \nu_d = 1 \\ \nu_i \left(t_{j+k} - t_j\right) (-1)^k \begin{bmatrix} \nu_1, ..., \nu_i + 1, ..., \nu_d \\ \tau_1, ..., \tau_i, ..., \tau_d \end{bmatrix} (x-y)_+^{k-1}, & \text{sonst.} \end{cases}$$

Die gleichen Formel gelten, falls $\nu_i = k-1$ *oder* $\nu_i = k$ *für alle* x *außer* $x = \tau_i$.

Folgerung 1 (Ableitung eines B-Splines nach den inneren Knoten)
Sei $t_1 = \cdots = t_k = a < t_{k+1} \leq \cdots \leq t_n < b = t_{n+1} = \cdots = t_{n+k}$ *eine Knotenfolge mit* $t_j < t_{j+k-2}$ $(j = 3, \ldots, n)$, *d.h.* $\#t_j \leq k-2$. *Dann existieren die Ableitungen der B-Splines nach den inneren Knoten* $\frac{\partial B_j}{\partial t_i}$ $(j = 1, \ldots, n; i = k+1, \ldots, n)$ *für alle* $x \in [a, b]$.

Bemerkung 4

(i) Die Bedingung $\#t_j \leq k-2$ *impliziert* $k \geq 3$, *d.h. mindestens quadratische B-Splines.*

(ii) Theorem 2 zeigt die Existenz der Ableitungen eines B-Splines nach den Randknoten $t_1 = \cdots = t_k$ *bzw.* $t_{n+1} = \cdots = t_{n+k}$ *an allen Stellen außer* t_1 *bzw.* t_{n+1}. *In den weiteren Ausführungen wird jedoch lediglich die Ableitung eines B-Splines nach den inneren Knoten benötigt.*

(iii) Der letzte Teil der Formel liefert den bekannten Zusammenhang zwischen der Ableitung eines B-Splines der Ordnung k *nach den inneren Knoten und der Ableitung eines B-Splines der Ordnung* $k+1$ *nach dem Argument*

$$\frac{\partial}{\partial \tau_i} B_{j,k} \left(x \,\middle|\, \begin{matrix} \nu_1, ..., \nu_i, ..., \nu_d \\ \tau_1, ..., \tau_i, ..., \tau_d \end{matrix} \right) = -\frac{\nu_i}{k} \frac{\partial}{\partial x} B_{j,k+1} \left(x \,\middle|\, \begin{matrix} \nu_1, ..., \nu_i + 1, ..., \nu_d \\ \tau_1, ..., \tau_i, ..., \tau_d \end{matrix} \right)$$

für $1 < i < d$.

2.5 Ableitung eines Splines nach den Knoten

Sei $s(x) = \sum_{j=1}^{n} B_{j,k}(x, \mathbf{t})\alpha_j$ ein Spline der Ordnung k mit der Knotenfolge $\mathbf{t} = (t_1, \ldots, t_{n+k})^T$. Gesucht ist

$$(2.28) \qquad \frac{\partial s(x)}{\partial t_{kk}} = \sum_{j=1}^{n} \frac{\partial B_{j,k}(x, \mathbf{t})}{\partial t_{kk}} \alpha_j.$$

Es gelte $\#t_{kk} \leq k-2$. Nach Theorem 2 gilt

$$\frac{\partial}{\partial t_{kk}} B_{j,k}(x,\mathbf{t}) = -\frac{\#t_{kk}}{k} \frac{\partial}{\partial x} B_{j,k+1}(x,\mathbf{t}')$$

für $j < kk < j+k$ mit der Knotenfolge

$$\mathbf{t}' = (t_1, \ldots, t_{kk}, t_{kk}, \ldots t_{n+k})^T \in \mathbb{R}^{n+k+1},$$

d.h.

$$\begin{aligned} t'_j &= t_j && j = 1, \ldots, kk \\ t'_j &= t_{j-1} && j = kk+1, \ldots, n+k+1. \end{aligned}$$

Bekanntlich gilt für die Ableitung eines B-Splines nach seinem Argument

$$\frac{\partial}{\partial x} B_{j,k}(x,\mathbf{t}) = (k-1)\left(\frac{B_{j,k-1}(x,\mathbf{t})}{t_{j+k-1}-t_j} - \frac{B_{j+1,k-1}(x,\mathbf{t})}{t_{j+k}-t_{j+1}}\right),$$

also

$$\frac{\partial}{\partial x} B_{j,k+1}(x,\mathbf{t}') = k\left(\frac{B_{j,k}(x,\mathbf{t}')}{t'_{j+k}-t'_j} - \frac{B_{j+1,k}(x,\mathbf{t}')}{t'_{j+k+1}-t'_{j+1}}\right).$$

Im weiteren setzen wir voraus, daß die Vielfachheit von t_{kk} gleich 1 ist ($\#t_{kk} = 1$).

Bemerkung 5
Da wir im folgenden ausschließlich die Ableitung eines Splines nach den freien Knoten benötigen, ist diese Voraussetzung erfüllt, falls die freien Knoten einfache Knoten sind (bzw. einfache Knoten bleiben).

Wir erhalten damit

$$\begin{aligned} \frac{\partial}{\partial t_{kk}} B_{j,k}(x,\mathbf{t}) &= -\frac{1}{k} k \left(\frac{B_{j,k}(x,\mathbf{t}')}{t'_{j+k}-t'_j} - \frac{B_{j+1,k}(x,\mathbf{t}')}{t'_{j+k+1}-t'_{j+1}}\right) \\ &= \frac{B_{j,k}(x,\mathbf{t}')}{t'_{j+k}-t'_j} - \frac{B_{j+1,k}(x,\mathbf{t}')}{t'_{j+k+1}-t'_{j+1}} \qquad \text{für} \quad j < kk < j+k. \end{aligned} \tag{2.29}$$

Zur Berechnung von $\frac{\partial s(x)}{\partial t_{kk}}$ benötigen wir noch $\frac{\partial B_{j,k}(x,\mathbf{t})}{\partial t_{kk}}$ für $kk \leq j$ und $j+k \leq kk$. Wegen der Definition von $B_{j,k} = B_{j,k}(t_j, \ldots, t_{j+k})$ gilt

$$\frac{\partial B_{j,k}}{\partial t_{kk}} = 0 \qquad \text{für} \quad kk < j \quad \text{und} \quad j+k < kk. \tag{2.30}$$

Es bleiben also $\frac{\partial B_{j,k}}{\partial t_j}$ und $\frac{\partial B_{j,k}}{\partial t_{j+k}}$ zu bestimmen. Nach der ersten Formel von (2.27) gilt

(2.31)

$$\begin{aligned} \frac{\partial B_{j,k}(x,\mathbf{t})}{\partial t_j} &= (-1)^{k-1} \begin{bmatrix} \nu_1+1, \nu_2, \ldots, \nu_d - 1 \\ \tau_1, \tau_2, \ldots, \tau_d \end{bmatrix} (x-y)_+^{k-1} \\ &= \frac{-1}{t_{j+k-1}-t_j} (-1)^k (t_{j+k-1}-t_j) \begin{bmatrix} \nu_1+1, \nu_2, \ldots, \nu_d - 1 \\ \tau_1, \tau_2, \ldots, \tau_d \end{bmatrix} (x-y)_+^{k-1} \\ &= -\frac{1}{t'_{j+k}-t'_j} B_{j,k}(x,\mathbf{t}'). \end{aligned}$$

Analog erhalten wir nach der zweiten Formel von (2.27)

(2.32)

$$\frac{\partial B_{j,k}(x,\mathbf{t})}{\partial t_{j+k}} = (-1)^k \left[\begin{array}{l} \nu_1 - 1, ..., \nu_{d-1}, \nu_d + 1 \\ \tau_1, ..., \tau_{d-1}, \tau_d \end{array} \right] (x-y)_+^{k-1}$$

$$= \frac{1}{t_{j+k} - t_{j+1}} (-1)^k (t_{j+k} - t_{j+1}) \left[\begin{array}{l} \nu_1 - 1, ..., \nu_{d-1}, \nu_d + 1 \\ \tau_1, ..., \tau_{d-1}, \tau_d \end{array} \right] (x-y)_+^{k-1}$$

$$= \frac{1}{t'_{j+k+1} - t'_{j+1}} B_{j+1,k}(x,\mathbf{t}').$$

Berücksichtigen wir nun (2.32), so ergibt sich

$$\frac{\partial s}{\partial t_{kk}} = \sum_{j=kk-k}^{kk} \frac{\partial B_{j,k}(x,\mathbf{t})}{\partial t_{kk}} \times \alpha_j$$

und mit (2.29), (2.31) und (2.32)

$$\begin{aligned} \frac{\partial s(x)}{\partial t_{kk}} &= \frac{1}{t'_{kk+1} - t'_{kk-k+1}} B_{kk-k+1,k}(x,\mathbf{t}') \times \alpha_{kk-k} \\ &\quad + \sum_{j=kk-k+1}^{kk-1} \left(\frac{B_{j+1,k}(x,\mathbf{t}')}{t'_{j+k+1} - t'_{j+1}} - \frac{B_{j,k}(x,\mathbf{t}')}{t'_{j+k} - t'_j} \right) \times \alpha_j \\ &\quad - \frac{1}{t'_{kk+k} - t'_{kk}} B_{kk,k}(x,\mathbf{t}') \times \alpha_{kk} \\ &= \sum_{j=kk-k+1}^{kk} \frac{\alpha_{j-1} - \alpha_j}{t'_{j+k} - t'_j} \times B_{j,k}(x,\mathbf{t}'). \end{aligned}$$

Lemma 3 (Ableitung eines Splines nach den Knoten)
Sei $s(x) = \sum_{j=1}^{n} B_{j,k}(x,\mathbf{t})\alpha_j$ *ein Spline der Ordnung* k *mit der Knotenfolge* $\mathbf{t} = (t_1, \ldots, t_{n+k})^T$. *Sei* t_{kk} *ein Splineknoten mit* $\#t_{kk} = 1 \leq k-2$ *und* $k < kk < n+1$. *Dann gilt*

$$\frac{\partial s(x)}{\partial t_{kk}} = \sum_{j=kk-k+1}^{kk} \frac{\alpha_{j-1} - \alpha_j}{t'_{j+k} - t'_j} \times B_{j,k}(x,\mathbf{t}')$$

mit der Knotenfolge $\mathbf{t}' = (t_1, \ldots, t_{kk}, t_{kk}, \ldots t_{n+k})^T \in \mathbb{R}^{n+k+1}$, *d.h.*

$$\begin{array}{ll} t'_j = t_j & j = 1, \ldots, kk \\ t'_j = t_{j-1} & j = kk+1, \ldots, n+k+1. \end{array}$$

Folgerung 2

$$\frac{\partial s(x_i)}{\partial t_{kk}} = 0 \qquad \textit{falls} \quad x_i \leq t_{kk-k+1} \quad \textit{oder} \quad x_i \geq t_{kk+k-1}.$$

3 Problemstellung

3.1 Formulierung des Funktionals

Gegeben seien Meßdaten $\{x_i, y_i\}$ $(i = 1, \ldots, m)$ einer unbekannten glatten Funktion $g \in W_2^q[a, b]$ mit

$$a \leq x_1 < \cdots < x_m \leq b. \tag{3.1}$$

Die mit Fehlern behafteten Meßwerte

$$y_i = g(x_i) + \epsilon_i \qquad (i = 1, \ldots, m) \tag{3.2}$$

$$\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)^T \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right) \tag{3.3}$$

wollen wir durch einen Spline s der Ordnung k zur Knotenfolge $\mathbf{t}$ approximieren, d.h. $s \in \mathcal{S}_{k,\mathbf{t}}$. Unter der Voraussetzung (3.3) stellt der *Approximationsterm*

$$\varphi(s) := \frac{1}{2} \sum_{i=1}^{m} \left[y_i - s(x_i)\right]^2 \tag{3.4}$$

ein geeignetes Maß für die Approximationsgüte dar.
Als Basis des Raumes $\mathcal{S}_{k,\mathbf{t}}$ verwenden wir B-Splines der Ornung k zur Knotenfolge $\mathbf{t}$ mit

$$t_1 = \cdots = t_k = a < t_{k+1} \leq \cdots \leq t_n < b = t_{n+1} = \cdots = t_{n+k}. \tag{3.5}$$

Wir definieren folgende Ansatzfunktion

$$s(x) := \sum_{j=1}^{n} B_{j,k}(x, \mathbf{t}) \alpha_j \tag{3.6}$$

bzw. in Vektorform

$$s(x) := \boldsymbol{\beta}(x, \mathbf{t})^T \boldsymbol{\alpha} \tag{3.7}$$

mit den Splinekoeffizienten

$$\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{R}^n \tag{3.8}$$

und den B-Spline-Werten

$$\boldsymbol{\beta}(x, \mathbf{t}) := (B_{1,k}(x, \mathbf{t}), \ldots, B_{n,k}(x, \mathbf{t}))^T \in \mathbb{R}^n. \tag{3.9}$$

Die Knotenfolge $\mathbf{t}$ erfülle die Bedingung

$$t_j < t_{j+k} \qquad (j = 1, \ldots, n) \tag{3.10}$$

bzw.

$$t_j < t_{j+k-q} \qquad (j = q + 1, \ldots, n) \tag{3.11}$$

mit einer Zahl $q \in \mathbb{Z}$ mit $0 \leq q < k$ sowie $q \leq n - 1$.

Bemerkung 6

(i) Die Bedingung (3.10) schließt die trivialen B-Splines $B_{j,k} \equiv 0$ von der Betrachtung aus.

(ii) Falls die Bedingung (3.11) erfüllt ist, so existiert $s^{(q)}$ und ist selbst wieder ein Spline. Es gilt $s \in W_2^q[a,b]$.

Mittels (3.6) erhalten wir für die Minimierung des Approximationstermes $\varphi(s) \to \min$

$$\frac{1}{2}\sum_{i=1}^{m}\left[y_i - \sum_{j=1}^{n} B_{j,k}(x,\mathbf{t})\alpha_j\right]^2 \longrightarrow \min . \tag{3.12}$$

Mit

$$\mathbf{y} := (y_1, \dots, y_m)^T \in \mathbb{R}^m \tag{3.13}$$

und der Beobachtungsmatrix

$$\mathbf{B}(\mathbf{t}) := \begin{bmatrix} B_{1,k}(x_1,\mathbf{t}) & \dots & B_{n,k}(x_1,\mathbf{t}) \\ \vdots & & \vdots \\ B_{1,k}(x_m,\mathbf{t}) & \dots & B_{n,k}(x_m,\mathbf{t}) \end{bmatrix} \in \mathbb{R}^{m,n} \tag{3.14}$$

erhalten wir die äquivalente Matrixformulierung

$$\varphi(s) := \frac{1}{2}\|\mathbf{y} - \mathbf{B}(\mathbf{t})\boldsymbol{\alpha}\| \longrightarrow \min . \tag{3.15}$$

Bei der praktischen Anwendung von Approximationsschemata ist oftmals ein Kompromiß zwischen der Güte der Approximation und der Glattheit gefragt. Wir betrachten daher zusätzlich den (exakten) *Glättungsterm*

$$\bar{\rho}(s) := \frac{1}{2}\left\|s^{(r)}\right\|_{L_2}^2 = \frac{1}{2}\int_a^b \left[s^{(r)}(x)\right]^2 dx \tag{3.16}$$

mit festem $r \in \{0, \dots, q\}$ und das zugehörige Minimierungsproblem

$$\bar{\phi}(s) := \varphi(s) + \mu\bar{\rho}(s) = \frac{1}{2}\sum_{i=1}^{m}\left[y_i - s(x_i)\right]^2 + \mu\frac{1}{2}\int_a^b \left[s^{(r)}(x)\right]^2 dx \longrightarrow \min \tag{3.17}$$

mit dem *Glättungsparameter* $\mu > 0$. Hierbei erfülle die Knotenfolge die Bedingung (3.11). Dann existiert $s^{(r)}$ und besitzt die Darstellung

$$s^{(r)}(x) = \sum_{j=r+1}^{n} B_{j,k-r}(x,\mathbf{t})\alpha_j^{(r)}$$

mit

$$\boldsymbol{\alpha}^{(r)} := \mathbf{D}_r(\mathbf{t})\boldsymbol{\alpha}$$

(vgl. Abschnitt 2.3). Setzt man dies in den Glättungsterm $\bar{\rho}$ ein, so erhält man

$$\bar{\rho} = \frac{1}{2}\int_a^b \left[\sum_{i=r+1}^{n} B_{i,k-r}(x,\mathbf{t})\alpha_i^{(r)}\right]\left[\sum_{j=r+1}^{n} B_{j,k-r}(x,\mathbf{t})\alpha_j^{(r)}\right] dx \tag{3.18}$$

$$= \frac{1}{2}\sum_{i,j=r+1}^{n} \alpha_i^{(r)}\alpha_j^{(r)} \int_a^b B_{i,k-r}(x,\mathbf{t})B_{j,k-r}(x,\mathbf{t})\, dx \tag{3.19}$$

$$= \frac{1}{2}\boldsymbol{\alpha}^{(r)^T}\bar{\mathbf{M}}_r(\mathbf{t})\boldsymbol{\alpha}^{(r)} \tag{3.20}$$

mit der, von der Knotenfolge $\mathbf{t}$ abhängigen, Matrix

$$\bar{\mathbf{M}}_r(\mathbf{t}) = \left(\bar{m}_{ij}^{(r)}\right) \in \mathbb{R}^{n-r,n-r} \tag{3.21}$$

$$\bar{m}_{ij}^{(r)} := \int_a^b B_{i,k-r}(x,\mathbf{t}) B_{j,k-r}(x,\mathbf{t})\, dx \quad i,j = r+1,\ldots,n. \tag{3.22}$$

Bemerkung 7
Die Matrix $\bar{\mathbf{M}}_r$ ist eine Bandmatrix der Bandbreite $2(k-r)-1$ und als Gramsche Matrix linear unabhängiger Funktionen ($t_j < t_{j+k-r}$, $j = r+1,\ldots,n$!) positiv definit.

Sei

$$\bar{\mathbf{M}}_r(\mathbf{t}) = \bar{\mathbf{F}}_r(\mathbf{t})^T \bar{\mathbf{F}}_r(\mathbf{t}) \tag{3.23}$$

mit der regulären oberen Dreiecksmatrix $\bar{\mathbf{F}}_r(\mathbf{t}) \in \mathbb{R}^{n-r,n-r}$ der Bandbreite $k-r$ die Cholesky-Faktorisierung der symmetrischen, positiv definiten Matrix $\bar{\mathbf{M}}_r(\mathbf{t})$. Damit erhält man für den Glättungsterm

$$\bar{\rho} = \frac{1}{2}\boldsymbol{\alpha}^{(r)^T} \bar{\mathbf{M}}_r(\mathbf{t})\boldsymbol{\alpha}^{(r)} \tag{3.24}$$

$$= \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{D}_r(\mathbf{t})^T \bar{\mathbf{F}}_r(\mathbf{t})^T \bar{\mathbf{F}}_r(\mathbf{t}) \mathbf{D}_r(\mathbf{t})\boldsymbol{\alpha} \tag{3.25}$$

$$= \frac{1}{2}\left\|\bar{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2 \tag{3.26}$$

mit der *Glättungsmatrix*

$$\bar{\mathbf{S}}_r(\mathbf{t}) := \bar{\mathbf{F}}_r(\mathbf{t})\mathbf{D}_r(\mathbf{t}) \in \mathbb{R}^{n-r,n}. \tag{3.27}$$

Die Glättungsmatrix $\bar{\mathbf{S}}_r(\mathbf{t})$ ist eine obere Dreiecksmatrix und besitzt die Bandbreite k.
Die Berechnung der Elemente $\bar{m}_{ij}^{(r)}$ der Gramschen Matrix ist für $k-r > 2$ sehr teuer. Sie kann etwa, wie von [Sch81] empfohlen, über Gaußsche Quadraturformeln erfolgen (vgl. auch [dBLS76] und [VBH92]). Wir verwenden eine billigere Approximation des Glättungstermes, welche sich in den Anwendungen (siehe [SK93] und [SS95]) als günstig erwiesen hat.
Wir ersetzen die L_2-Norm in der Definition des Glättungsterms durch ihr diskretes Äquivalent im Splineraum $\mathcal{S}_{k-r,\mathbf{t}}$

$$\tilde{\rho}(s) := \frac{1}{2}\left\|s^{(r)}\right\|_{l_2}^2 := \frac{1}{2}\sum_{j=r+1}^{n}\left(\alpha_j^{(r)}\right)^2 \frac{t_{j+k-r}-t_j}{k-r}. \tag{3.28}$$

Die Normen sind in folgendem Sinne äquivalent:
Es existiert eine Konstante d_{k-r}, welche nicht von der Knotenfolge $\mathbf{t}$ abhängt, so daß gilt:

$$d_{k-r}^{-1}\left\|s^{(r)}\right\|_{l_2} \le \left\|s^{(r)}\right\|_{L_2} \le \left\|s^{(r)}\right\|_{l_2} \qquad \forall s \in \mathcal{S}_{k,\mathbf{t}}.$$

Die Existenz der Konstanten d_{k-r} für beliebige L_p-Normen und ihr diskretes Äquivalent l_p geht auf de Boor [dB76] zurück. Die Ersetzung der L_p-Norm durch die diskrete l_p Norm wird ebenfalls in [LM88] und [ADLM90] vorgeschlagen. Für diese Approximation erhalten wir

$$\tilde{\rho} = \frac{1}{2}\boldsymbol{\alpha}^{(r)^T}\tilde{\mathbf{F}}_r(\mathbf{t})^T\tilde{\mathbf{F}}_r(\mathbf{t})\boldsymbol{\alpha}^{(r)} \tag{3.29}$$

$$= \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{D}_r(\mathbf{t})^T\tilde{\mathbf{F}}_r(\mathbf{t})^T\tilde{\mathbf{F}}_r(\mathbf{t})\mathbf{D}_r(\mathbf{t})\boldsymbol{\alpha} \tag{3.30}$$

mit

$$\tilde{\mathbf{F}}_r(\mathbf{t}) := \operatorname{diag}\left(\sqrt{\frac{t_{k+j}-t_{r+j}}{k-r}}\right)_{j=1,\ldots,n-r}, \tag{3.31}$$

und schließlich

$$\tilde{\rho} = \frac{1}{2}\left\|\tilde{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2 \tag{3.32}$$

mit der *approximierten Glättungsmatrix*

$$\tilde{\mathbf{S}}_r(\mathbf{t}) := \tilde{\mathbf{F}}_r(\mathbf{t})\mathbf{D}_r(\mathbf{t}) \in \mathbb{R}^{n-r,n}. \tag{3.33}$$

Die Matrix $\tilde{\mathbf{S}}_r(\mathbf{t})$ ist eine obere Dreiecksmatrix der Bandbreite $r+1$ (im Gegensatz zu k bei $\bar{\mathbf{S}}_r$!).

Wegen der Äquivalenz der Normen besitzen die Matrizen $\bar{\mathbf{S}}_r$ und $\tilde{\mathbf{S}}_r$ beide vollen Rang $n-r$ und haben denselben Nullraum

$$\ker\left(\tilde{\mathbf{S}}_r\right) = \ker\left(\bar{\mathbf{S}}_r\right) = \left\{\boldsymbol{\alpha}\in\mathbb{R}^n \,:\, \alpha_j^{(r)} = 0\,,\, j = r+1,\ldots,n\right\},$$

welcher durch alle Polynome der Ordnung r aus $\mathcal{S}_{k,\mathbf{t}}$ gekennzeichnet ist.

Mit $\rho = \bar{\rho} = \frac{1}{2}\left\|\bar{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2$, $\mathbf{S}_r(\mathbf{t}) = \bar{\mathbf{S}}_r(\mathbf{t})$ (exakter Glättungsterm) oder $\rho = \tilde{\rho} = \frac{1}{2}\left\|\tilde{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\right\|^2$, $\mathbf{S}_r(\mathbf{t}) = \tilde{\mathbf{S}}_r(\mathbf{t})$ (approximierter Glättungsterm) erhalten wir das zu minimierende Funktional

$$\varphi(s) := \phi(s) + \mu\rho(s) = \tfrac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\mathbf{t})\boldsymbol{\alpha}\right\|^2 + \mu\tfrac{1}{2}\left\|\mathbf{S}_r(\mathbf{t})\boldsymbol{\alpha}\right]^2 \tag{3.34}$$

$$= \frac{1}{2}\left\|\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix} - \begin{bmatrix}\mathbf{B}(\mathbf{t})\\ \sqrt{\mu}\mathbf{S}_r(\mathbf{t})\end{bmatrix}\boldsymbol{\alpha}\right\|^2 \longrightarrow \min. \tag{3.35}$$

Wir setzen im folgenden voraus, daß

$$m \geq r. \tag{3.36}$$

Dann folgt aus $\mathbf{B}(\mathbf{t})\boldsymbol{\alpha} = \mathbf{0}$ und $\mathbf{S}_r(\mathbf{t})\boldsymbol{\alpha} = \left\{\begin{array}{c}\tilde{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\\ \bar{\mathbf{S}}_r(\mathbf{t})\boldsymbol{\alpha}\end{array}\right\} = \mathbf{0}$, daß $\boldsymbol{\alpha} = \mathbf{0}$, d.h. die Matrix $\begin{bmatrix}\mathbf{B}(\mathbf{t})\\ \sqrt{\mu}\mathbf{S}_r(\mathbf{t})\end{bmatrix}$ hat für jedes $\mu > 0$ Vollrang n und (3.35) besitzt für jede feste Knotenfolge $\mathbf{t}$ eine eindeutige Lösung $\boldsymbol{\alpha}_{opt}(\mathbf{t})$.

Aus der Menge der inneren Splineknoten $\{t_j\}$ $(j = k+1,\ldots,n)$ wählen wir jetzt eine Teilmenge, die *freien Knoten*, aus. Die Anzahl der freien Knoten sei mit l bezeichnet und

$$\tilde{\mathbf{t}} = \left(t_{p(1)},\ldots,t_{p(l)}\right)^T \in \mathbb{R}^l \tag{3.37}$$

sei der Vektor der freien Knoten, wobei

$$\mathbf{p} = (p(1), \ldots, p(l))^T \in \mathbb{Z}^l \tag{3.38}$$

die Indizes der freien Knoten enthält, für welche

$$k < p(1) < p(2) < \cdots < p(l) < n + 1 \tag{3.39}$$

gelten soll. Diese freien Knoten $\tilde{\mathbf{t}}$ sollen zusätzlich zu den Splinekoeffizienten $\boldsymbol{\alpha}$ in den Optimierungsprozeß (3.15) bzw. (3.35) einbezogen werden.

Vereinbarung

Da im weiteren nur die freien Knoten verändert werden, wird die Abhängigkeit der Größen von der Knotenfolge $\mathbf{t}$ durch den Vektor $\tilde{\mathbf{t}}$ der freien Knoten kenntlich gemacht. Falls aus dem Zusammenhang deutlich wird, auf welche Knotenfolge sich die Größen beziehen, verzichten wir gegebenenfalls auch auf die gesonderte Kennzeichnung der freien Knoten, z.B. $\mathbf{B}(\mathbf{t}) \longrightarrow \mathbf{B}(\tilde{\mathbf{t}}) \longrightarrow \mathbf{B}$ und $\mathbf{D}_r(\mathbf{t}) \longrightarrow \mathbf{D}_r(\tilde{\mathbf{t}}) \longrightarrow \mathbf{D}_r$.

Mit dieser Vereinbarung erhält man die folgenden zu minimierenden Funktionale:

Quadratmittelproblem (ohne Glättung)

Sei $t_j < t_{j+k}$ $(j = 1, \ldots, n)$.

$$\varphi(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) := \frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2 \longrightarrow \min_{\tilde{\mathbf{t}}, \boldsymbol{\alpha}} \tag{3.40}$$

Das Quadratmittelproblem (3.40) hat genau dann eine eindeutige Lösung $\boldsymbol{\alpha}$ zu festen $\tilde{\mathbf{t}}$, falls die Knotenfolge $\mathbf{t}$ die Whitney-Schoenberg-Bedingung (2.14) erfüllt.

Quadratmittelproblem (mit Glättung)

Sei $t_j < t_{j+k-q}$ $(j = q + 1, \ldots, n)$ und $r \in \{0, \ldots, q\}$.

(3.41)

$$\phi(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) := \varphi(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) + \mu\rho(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) = \frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2 + \frac{1}{2}\mu \int_a^b \left[s^{(r)}(x) \right]^2 dx \longrightarrow \min_{\tilde{\mathbf{t}}, \boldsymbol{\alpha}}$$

Unter der Voraussetzung $m \geq r$ und $\mu > 0$ hat das Quadratmittelproblem (3.41) genau eine Lösung $\boldsymbol{\alpha}$ (für festes $\tilde{\mathbf{t}}$).

3.2 Die Anordnungsbedingung für die freien Knoten

3.2.1 Wahl eines absoluten Distanzmaßes

Für eine Knotenfolge $\mathbf{t}$ muß gelten

$$t_{j+1} - t_j \geq 0 \qquad j = 1, \ldots, n + k - 1.$$

Da bei der Minimierung von (3.40) bzw. (3.41) nur die freien Knoten $\tilde{\mathbf{t}}$ verändert werden, lautet diese Bedingung

$$\begin{aligned} t_{p(j)} - t_{p(j)-1} &\geq 0 \\ t_{p(j)+1} - t_{p(j)} &\geq 0 \qquad j = 1, \ldots, l. \end{aligned} \tag{3.42}$$

Für die numerische Praxis ist diese Bedingung ungünstig ((3.10) bzw. (3.11) wird nicht erfüllt). Wir fordern daher

$$
\begin{aligned}
t_{p(j)} - t_{p(j)-1} &\geq \epsilon \\
t_{p(j)+1} - t_{p(j)} &\geq \epsilon \qquad j = 1, \ldots, l
\end{aligned}
\tag{3.43}
$$

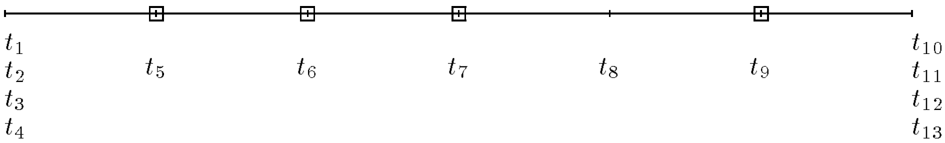
mit einer vorgegebenen Konstanten $\epsilon > 0$. Die Konstante ϵ sollte in Abhängigkeit von der mittleren absoluten Distanz zweier benachbarter Knoten gewählt werden und wird daher als *absolutes* Distanzmaß bezeichnet.

Die Bedingung (3.43) kann äquivalent in Matrixform angegeben werden

$$
\mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h} \qquad \text{mit } \mathbf{C} \in \mathbb{R}^{ncstr,l}, \tilde{\mathbf{t}} \in \mathbb{R}^{l}, \mathbf{h} \in \mathbb{R}^{ncstr}. \tag{3.44}
$$

Die Anzahl der linear unabhängigen Nebenbedingungen sei *ncstr*. Die Matrix $\mathbf{C}$ und der Vektor $\mathbf{h}$ hängen von der *Lagebeziehung* der Knoten ab und können leicht algorithmisch erzeugt werden. Zur Illustration möge ein kleines Beispiel dienen:

Beispiel 1
$k = 4$, $n = 9$, $l = 4$, $p(1) = 5$, $p(2) = 6$, $p(3) = 7$, $p(4) = 9$



$$
\begin{bmatrix}
1 & & & \\
-1 & 1 & & \\
 & -1 & 1 & \\
 & & -1 & \\
 & & & 1 \\
 & & & -1
\end{bmatrix}
\begin{bmatrix} t_5 \\ t_6 \\ t_7 \\ t_9 \end{bmatrix}
\geq
\begin{bmatrix} t_4 + \epsilon \\ \epsilon \\ \epsilon \\ -t_8 + \epsilon \\ t_8 + \epsilon \\ -t_{10} + \epsilon \end{bmatrix}
$$

Bemerkung 8

(i) *Es gilt* $l + 1 \leq ncstr \leq 2l$.

(ii) *Die Matrix* $\mathbf{C}$ *enthält maximal zwei Nichtnullelemente in jeder Zeile.*

(iii) *Mit der Wahl eines beliebig kleinen, aber positiven* ϵ *ist – zumindest theoretisch in rundungsfehlerfreier Arithmetik – die maximale Glattheit des Splines an den Stellen* $t_{p(j)}$, $j = 1, \ldots, l$, *sichergestellt, da die freien Knoten dann einfache Knoten sind.*

3.2.2 Wahl eines relativen Distanzmaßes

Bisher betrachteten wir die Anordnungsbedingung (3.42) bzw. um das Zusammenfallen von Knoten zu verhindern die Bedingung (3.43). Der Nachteil der Bedingung (3.43) ist, daß ϵ eine *absolute* minimale Distanz zwischen den Knoten ist und deshalb

- schwierig zu wählen ist,
- bei sehr ungleichmäßig verteilten Knoten ungeeignet ist.

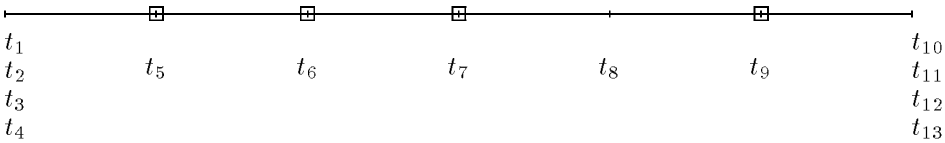
Eine bessere Bedingung stellt die folgende *relative* Bedingung dar, welche auch von de Boor/Rice [dBR68] verwendet wird. Sie verlangen, daß

$$t_{p(j)-1} + \epsilon\left(t_{p(j)+1} - t_{p(j)-1}\right) \le t_{p(j)} \le t_{p(j)+1} - \epsilon\left(t_{p(j)+1} - t_{p(j)-1}\right), d.h.$$

$$\begin{aligned} t_{p(j)} - t_{p(j)-1} &\ge \epsilon\left(t_{p(j)+1} - t_{p(j)-1}\right) \\ t_{p(j)+1} - t_{p(j)} &\ge \epsilon\left(t_{p(j)+1} - t_{p(j)-1}\right) \qquad j = 1, \ldots, l. \end{aligned} \tag{3.45}$$

Das *relative* Distanzmaß ϵ wird von de Boor/Rice mit $\epsilon = 0.0625$ gewählt. Auch hier soll zur Illustration ein Beispiel dienen:

Beispiel 2
$k = 4$, $n = 9$, $l = 4$, $p(1) = 5$, $p(2) = 6$, $p(3) = 7$, $p(4) = 9$



$$\begin{bmatrix} 1 & -\epsilon & & \\ -1 & 1-\epsilon & & \\ \epsilon - 1 & 1 & -\epsilon & \\ \epsilon & -1 & 1-\epsilon & \\ & \epsilon - 1 & 1 & \\ & \epsilon & -1 & \\ & & & 1 \\ & & & -1 \end{bmatrix} \begin{bmatrix} t_5 \\ t_6 \\ t_7 \\ t_9 \end{bmatrix} \ge \begin{bmatrix} (1-\epsilon)\, t_4 \\ -\epsilon\, t_4 \\ 0 \\ 0 \\ \epsilon\, t_8 \\ (\epsilon - 1)\, t_8 \\ (1-\epsilon)\, t_8 + \epsilon\, t_{10} \\ -\epsilon\, t_8 + (\epsilon - 1)\, t_{10} \end{bmatrix}$$

Formal erhalten wir die gleiche Matrixformulierung der Nebenbedingungen:

$$\mathbf{C}\tilde{\mathbf{t}} \ge \mathbf{h} \qquad \text{mit } \mathbf{C} \in \mathbb{R}^{2l,l}, \tilde{\mathbf{t}} \in \mathbb{R}^{l}, \mathbf{h} \in \mathbb{R}^{2l}.$$

Bemerkung 9

(i) Man beachte, daß $ncstr = 2l$ und unabhängig von der Topologie der freien Knoten ist.

(ii) Die Matrix $\mathbf{C} \in \mathbb{R}^{2l,l}$ ist wiederum schwachbesetzt, kann jetzt jedoch maximal drei Nichtnullelemente je Zeile enthalten und ist abhängig von ϵ.

3.3 Nebenbedingungen an Ableitungen des Splines

Eine in der Einleitung formulierte Forderung an die Funktion s war die Formerhaltung. Zu diesem Zweck drücken wir jetzt Nebenbedingungen an Ableitungen des Splines mittels der Splineknoten und der Splinekoeffizienten aus.

Die Knotenfolge $\mathbf{t}$ erfülle die Bedingung (3.11) und es gelte $p \in \{0, \ldots, q\}$. Dann existiert $s^{(p)}$ und besitzt die Matrixdarstellung

$$\begin{aligned} s^{(p)}(x) &= \boldsymbol{\beta}_p(x, \mathbf{t})^T \mathbf{D}_p(\mathbf{t})\boldsymbol{\alpha} \\ &= \boldsymbol{\beta}_p(x, \tilde{\mathbf{t}})^T \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha}. \end{aligned} \tag{3.46}$$

Damit sind wir in der Lage die Nebenbedingung, welche den „shape“ des Splines charakterisiert, zu formulieren. Wir betrachten zunächst die Positivitätsforderung an $s^{(p)}$ im ganzen Intervall $[a, b]$.

3.3.1 Positivitätsforderung an die Ableitung des Splines

$$
\begin{aligned}
s^{(p)}(x) &\geq 0 \\
\boldsymbol{\beta}_p^T(x, \tilde{\mathbf{t}}) \mathbf{D}_p(\tilde{\mathbf{t}}) \boldsymbol{\alpha} &\geq 0 \qquad \forall x \in [t_k, t_{n+1}]
\end{aligned} \tag{3.47}
$$

Die Nebenbedingung (3.47) an die p-te Ableitung des Splines ergibt zusammen mit der Anordnungsbedingung (3.43) bzw. (3.45) für die freien Knoten und dem zu minimierenden Funktional (3.40) bzw. (3.41) ein semiinfinites Optimierungsproblem.

Zumindest für kleine $k-p$ lassen sich – in der Regel nichtlineare – notwendige und hinreichende Bedingungen für die Positivität der Polynome der Ordnung $k-p$ in den Teilintervallen angeben (siehe [SH88]). Damit läßt sich in diesen Fällen das semiinfinite in ein finites Optimierungsproblem überführen.

Um die Lösung des semiinfiniten Optimierungsproblems zu vermeiden, verwenden wir statt der notwendigen und hinreichenden Bedingung (3.47) die – wegen der Nichtnegativität der B-Splines $\boldsymbol{\beta}_p(x, \tilde{\mathbf{t}})$ – hinreichende Bedingung $\boldsymbol{\alpha}^{(p)} = \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \geq \mathbf{0}$. Wir erhalten daher die für die Positivität hinreichende Nebenbedingung

$$
\mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \geq 0. \tag{3.48}
$$

Bemerkung 10
Wir betrachten jetzt eine feste Knotenfolge $\mathbf{t}$. *Da das betrachtete Funktional (3.40) bzw. (3.41) bei Erfülltsein der Whitney-Schoenberg-Bedingung (2.14) bzw. unter der Voraussetzung* $m \geq r$ *und* $\mu > 0$ *streng konvex ist, sind sowohl das semiinfinite Problem (3.47) als auch das finite Problem (3.48) eindeutig lösbar. Die Optimallösung* $\boldsymbol{\alpha}^f$ *des finiten Problems ist wegen* $B_{j,k-p}(x) \geq 0$ *zulässig für das semiinfinite Problem. Falls* $k - p \leq 2$, *so stimmen die Lösungen der beiden Probleme überein,* $\boldsymbol{\alpha}^f = \boldsymbol{\alpha}^{if}$. *Im allgemeinen Fall kann man feststellen, daß* $\boldsymbol{\alpha}^{if}$ *durch* $\boldsymbol{\alpha}^f$ *desto besser approximiert wird, je feiner die Knotenfolge* $\mathbf{t}$ *ist.*

3.3.2 Einfache Schranken für die Ableitung des Splines

Oftmals wird neben der unteren Schranke Null an $s^{(p)}$ gleichzeitig eine obere Schranke vorgegeben, z.B. um eine allzu große Krümmung des Splines zu verhindern, oder es werden auf verschiedenen Intervallen verschiedene „shape constraints“ vorgegeben, z.B. konvex-konkaver Datenabgleich (siehe [SH88]).

Allgemeiner fordern wir daher nun

$$
l_i^{(p)} \leq s^{(p)}(x) \leq u_i^{(p)} \qquad \forall x \in [t_i, t_{i+1}), \quad i = k, \ldots, n \tag{3.49}
$$

mit $2(n-k+1)$ Konstanten

$$
\mathbf{l} := \left(l_k^{(p)}, \ldots, l_n^{(p)}\right)^T \in \mathbb{R}^{n-k+1} \tag{3.50}
$$

$$
\mathbf{u} := \left(u_k^{(p)}, \ldots, u_n^{(p)}\right)^T \in \mathbb{R}^{n-k+1}. \tag{3.51}
$$

Wegen der Eigenschaften von B-Splines gilt

$$
\min\left\{\alpha_j^{(p)} \,:\, j \in K_i\right\} \leq s^{(p)}(x) = \sum_{j \in K_i} B_{j,k-p}(x)\alpha_j^{(p)} \leq \max\left\{\alpha_j^{(p)} \,:\, j \in K_i\right\}
$$

für $x \in [t_i, t_{i+1})$ und $K_i := \{i-k+p+1, \ldots, i\}$. Eine hinreichende Bedingung für (3.49) ist deshalb

$$(3.52)\quad l_i^{(p)} \leq \min\left\{\alpha_j^{(p)}\ :\ j \in K_i\right\} \text{ und } \max\left\{\alpha_j^{(p)}\ :\ j \in K_i\right\} \leq u_i^{(p)} \quad i = k, \ldots, n.$$

Dies können wir äquivalent folgendermaßen formulieren

$$(3.53)\qquad L_j^{(p)} \leq \alpha_j^{(p)} \leq U_j^{(p)} \qquad j = p+1, \ldots, n$$

mit den $2(n-p)$ Konstanten

$$(3.54)\qquad L_j^{(p)} := \max\left\{l_i^{(p)}\ :\ i \in W_j\right\}$$

$$(3.55)\qquad U_j^{(p)} := \min\left\{u_i^{(p)}\ :\ i \in W_j\right\}$$

und der Indexmenge

$$(3.56)\qquad W_j := \Big\{\max\{j,k\}, \ldots, \min\{j+k-p-1, n\}\Big\}$$

bzw.

$$(3.57)\qquad \mathbf{L} \leq \boldsymbol{\alpha}^{(p)} \leq \mathbf{U}$$

$$(3.58)\qquad \mathbf{L} := \left(L_{p+1}^{(p)}, \ldots, L_n^{(p)}\right)^T \in \mathbb{R}^{n-p}$$

$$(3.59)\qquad \mathbf{U} := \left(U_{p+1}^{(p)}, \ldots, U_n^{(p)}\right)^T \in \mathbb{R}^{n-p}.$$

Eine hinreichende Bedingung für das Erfülltsein von (3.49) ist also

$$(3.60)\qquad \mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}.$$

Bemerkung 11

(i) *In den weiteren Ausführungen seien $-\infty$ und $+\infty$ als untere bzw. obere Schranken zugelassen. Damit ergibt sich die Positivtätsforderung an $s^{(p)}$ als Spezialfall der einfachen Schranken. Die verwendeten Algorithmen sind in der Lage, diese Fälle ebenfalls zu behandeln.*

(ii) *Noch allgemeinere Nebenbedingungen an Ableitungen – etwa die simultane Forderung nach Konvexität und Monotonie – führen auf ähnlich strukturierte Nebenbedingungen (siehe [Kun95]). Um die in dieser Arbeit entwickelte Technik anwenden zu können, sind von diesen verallgemeinerten Nebenbedingungen gegebenenfalls in einem „preprocessing step“ redundante Nebenbedingungen zu entfernen.*

(iii) *Für feste zulässige Knotenfolgen $\mathbf{t}$, d.h. $\mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}$, ist der zulässige Bereich des Optimierungsproblems genau dann nicht leer, falls $\mathbf{L} \leq \mathbf{U}$.*

Definition 5 (Konsistenz, strikte Konsistenz)
Die Nebenbedingungen

$$l_i^{(p)} \leq s^{(p)}(x) \leq u_i^{(p)} \qquad \forall x \in [t_i, t_{i+1}), \quad i = k, \ldots, n$$

heißen konsistent, *falls*

$$(3.61)\qquad \begin{array}{c} L_j^{(p)} \leq U_j^{(p)} \qquad j = p+1, \ldots, n \\ \mathbf{L} \leq \mathbf{U}. \end{array}$$

Sie heißen strikt konsistent, *falls*

$$L_j^{(p)} < U_j^{(p)} \qquad j = p+1, \ldots, n \tag{3.62}$$
$$\mathbf{L} < \mathbf{U}.$$

Beispiel 3
$k = 4$, $n = 11$, $p = 2$

$l_4 = -10$	$u_4 = 100$	$W_3 = \{4\}$	$L_3 = -10$	$U_3 = 100$
$l_5 = -8$	$u_5 = 500$	$W_4 = \{4,5\}$	$L_4 = -8$	$U_4 = 100$
$l_6 = -8$	$u_6 = +\infty$	$W_5 = \{5,6\}$	$L_5 = -8$	$U_5 = 500$
$l_7 = -20$	$u_7 = 5$	$W_6 = \{6,7\}$	$L_6 = -8$	$U_6 = 5$
$l_8 = -\infty$	$u_8 = -1$	$W_7 = \{7,8\}$	$L_7 = -20$	$U_7 = -1$
$l_9 = -\infty$	$u_9 = -1$	$W_8 = \{8,9\}$	$L_8 = -\infty$	$U_8 = -1$
$l_{10} = 0.5$	$u_{10} = 80$	$W_9 = \{9,10\}$	$L_9 = 0.5$	$U_9 = -1$
$l_{11} = -\infty$	$u_{11} = +\infty$	$W_{10} = \{10,11\}$	$L_{10} = 0.5$	$U_{10} = 80$
		$W_{11} = \{11\}$	$L_{11} = -\infty$	$U_{11} = +\infty$

$$0.5 = L_9 > U_9 = -1,$$

d.h. die Konsistenzbedingung $\mathbf{L} \le \mathbf{U}$ *ist nicht erfüllt, obwohl* $\mathbf{l} < \mathbf{u}$ *gilt.*

Beispiel 4
$k = 4$, $n = 9$, $p = 1$ *monoton wachsend im Intervall* $[t_4, t_8)$
monoton fallend im Intervall $[t_9, t_{10})$

$$\begin{aligned}
\mathbf{l} &= (0, 0, 0, 0, -\infty, -\infty)^T \\
\mathbf{u} &= (+\infty, +\infty, +\infty, +\infty, +\infty, 0)^T \\
\mathbf{L} &= (0, 0, 0, 0, 0, 0, -\infty, -\infty)^T \\
\mathbf{U} &= (+\infty, +\infty, +\infty, +\infty, +\infty, 0, 0, 0)^T
\end{aligned}$$

d.h. die Nebenbedingungen sind konsistent, aber nicht strikt konsistent.

Beispiel 5
$k = 4$, $n = 9$, $p = 2$ *konvex im Intervall* $[t_4, t_8)$
konkav im Intervall $[t_9, t_{10})$

$$\begin{aligned}
\mathbf{l} &= (0, 0, 0, 0, -\infty, -\infty)^T \\
\mathbf{u} &= (+\infty, +\infty, +\infty, +\infty, +\infty, 0)^T \\
\mathbf{L} &= (0, 0, 0, 0, 0, -\infty, -\infty)^T \\
\mathbf{U} &= (+\infty, +\infty, +\infty, +\infty, +\infty, 0, 0)^T
\end{aligned}$$

d.h. die Nebenbedingungen sind strikt konsistent und damit auch konsistent.

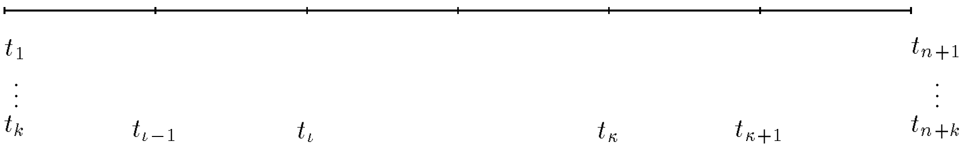
3.3.3 Strikte Konsistenz

In den späteren Anwendungen wird die strikte Konsistenz der Nebenbedingungen benötigt. Im Falle nur einseitiger Schranken an $s^{(p)}$ auf einem Teilintervall ist diese

Bedingung trivialerweise erfüllt. Wir untersuchen jetzt den wichtigen Fall von einseitigen Schranken an $s^{(p)}$ auf verschiedenen Teilintervallen. Wir beschränken uns o.B.d.A. auf zwei Teilintervalle mit Schranken an Ableitungen. Wir betrachten die Nebenbedingung

$$\begin{aligned} s^{(p)}(x) &\geq 0 \qquad \text{für alle} \quad x \in [a, t_\iota) \\ s^{(p)}(x) &\leq 0 \qquad \text{für alle} \quad x \in [t_\kappa, b) \end{aligned}$$

mit $\iota, \kappa \in \{k+1, \ldots, n\}$ und $\iota \leq \kappa$. Die Knoten t_ι und t_κ seien o.B.d.A. hinreichend weit im Inneren des Intervalls $[t_k, t_{n+1}]$.



Dann gilt:

$$\begin{aligned} l_i^{(p)} &= 0 \\ u_i^{(p)} &= +\infty \qquad i = k, \ldots, \iota - 1 \\[1ex] l_i^{(p)} &= -\infty \\ u_i^{(p)} &= +\infty \qquad i = \iota, \ldots, \kappa - 1 \\[1ex] l_i^{(p)} &= -\infty \\ u_i^{(p)} &= 0 \qquad i = \kappa, \ldots, n. \end{aligned}$$

Für die Schranken $L_j^{(p)}$ und $U_j^{(p)}$ für die Splinekoeffizienten $\alpha_j^{(p)}$ $(j = p+1, \ldots, n)$ erhalten wir

$$\begin{aligned} L_j^{(p)} &= 0 \\ U_j^{(p)} &= +\infty \qquad j = p+1, \ldots, \iota - k + p \\[1ex] L_j^{(p)} &= -\infty \\ U_j^{(p)} &= 0 \qquad j = \kappa, \ldots, n. \end{aligned}$$

Die obigen Schranken werden allein durch die Vorgaben $s^{(p)} \geq 0$ auf $[a, t_\iota)$ bzw. $s^{(p)} \leq 0$ auf $[t_\kappa, b]$ bestimmt. Für die Schranken $L_j^{(p)}, U_j^{(p)}$ $(j = \iota - k + p + 1, \ldots, \kappa - 1)$ beeinflussen sich diese Nebenbedingungen gegenseitig.

a) Untere Schranken

$j = \iota - 1$ ist der letzte Index, für welchen gilt $\iota - 1 \in W_j$. Es gilt

$$\begin{aligned} W_{\iota-1} &= \{\iota - 1, \ldots, \iota + k - p - 2\} \\ L_{\iota-1}^{(p)} &= 0 \end{aligned}$$

sowie

$$\begin{aligned} W_\iota &= \{\iota, \ldots, \iota + k - p - 1\} \\ L_{\iota-1}^{(p)} &= -\infty. \end{aligned}$$

b) Obere Schranken

$j = \kappa - k + p + 1$ ist der erste Index, für welchen gilt $\kappa \in W_j$. Es gilt

$$\begin{aligned} W_{\kappa-k+p} &= \{\kappa - k + p, \ldots, \kappa - 1\} \\ U^{(p)}_{\kappa-k+p} &= +\infty \end{aligned}$$

sowie

$$\begin{aligned} W_{\kappa-k+p+1} &= \{\kappa - k + p + 1, \ldots, \kappa\} \\ U^{(p)}_{\kappa-k+p+1} &= 0. \end{aligned}$$

Die Nebenbedingungen sind genau dann strikt konsistent, wenn $\iota - 1 < \kappa - k + p + 1$, d.h. $\kappa - \iota \geq k - p - 1$.

Lemma 4 (Strikte Konsistenz der Nebenbedingungen)
Sei $\kappa - \iota \geq k - p - 1$ und $\iota \leq \kappa$. Für die Nebenbedingung

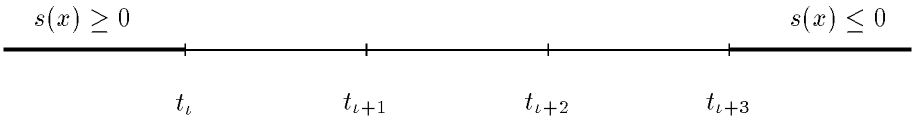
$$\begin{array}{lll} s^{(p)}(x) \geq 0 & \textit{für alle} & x \in [a, t_\iota) \\ s^{(p)}(x) \leq 0 & \textit{für alle} & x \in [t_\kappa, b) \end{array}$$

gilt dann

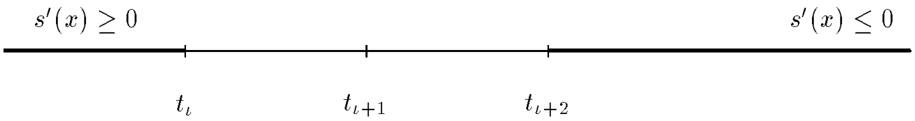
$$\begin{array}{ll} L^{(p)}_j = 0 & j = p+1, \ldots, \iota - 1 \\ L^{(p)}_j = -\infty & j = \iota, \ldots, n \\ & \\ U^{(p)}_j = +\infty & j = p+1, \ldots, \kappa - k + p \\ U^{(p)}_j = 0 & j = \kappa - k + p + 1, \ldots, n. \end{array}$$

Die strikte Konsistenzbedingung $\mathbf{L} < \mathbf{U}$ *ist erfüllt.*

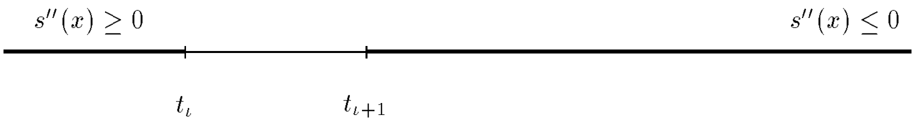
Beispiel 6 (Strikte Konsistenz für kubische Splines)
$p = 0$: $\kappa \geq \iota + 3$



$p = 1$: $\kappa \geq \iota + 2$



$p = 2$: $\kappa \geq \iota + 1$



$p = 3$: $\kappa \geq \iota$

$s'''(x) \geq 0$ $\qquad\qquad\qquad\qquad\qquad\qquad$ $s'''(x) \leq 0$

t_ι

3.4 Formulierung der Nebenbedingungen

3.4.1 Nebenbedingungen an freie Knoten

Sowohl für das absolute als auch für das relative Distanzmaß ϵ erhalten wir lineare Ungleichheitsnebenbedingungen für die freien Knoten in der Form

$$\mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h} \tag{3.63}$$

Die Matrix $\mathbf{C} \in \mathbb{R}^{ncstr,l}$ ist konstant und schwach besetzt. Der Vektor $\mathbf{h} \in \mathbb{R}^{ncstr}$ ist von $\mathbf{t}\backslash\{\tilde{\mathbf{t}}\}$ abhängig. Für die Anzahl $ncstr$ linear unabhängiger Nebenbedingungen gilt $l + 1 \leq ncstr \leq 2l$.

3.4.2 Nebenbedingungen an Ableitungen des Splines

Indem wir formal $-\infty$ und $+\infty$ als untere bzw. obere Schranke zulassen, können wir die Nebenbedingungen an Ableitungen einheitlich in der folgenden Form schreiben:

$$\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U} \tag{3.64}$$

3.5 Formulierung des kompletten Problems

Nun sind wir in der Lage, das komplette Optimierungsproblem zu formulieren. Wir definieren das *vollständige restringierte Approximationsproblem* (full constrained approximation problem, **FCAP**).
Sei $t_j < t_{j+k-q}$ $(j = q+1, \ldots, n)$ und $p \in \{0, \ldots, q\}$.

$$\frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \longrightarrow \min_{\tilde{\mathbf{t}},\boldsymbol{\alpha}} \tag{3.65}$$

bei

$$\mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h} \tag{3.66}$$

und

$$\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}. \tag{3.67}$$

Wir definieren das *vollständige restringierte Glättungsproblem* (full constrained smoothing problem, **FCSP**). Sei $t_j < t_{j+k-q}$ $(j = q+1, \ldots, n)$ und $p, q \in \{0, \ldots, q\}$.

$$\frac{1}{2}\left\|\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}\boldsymbol{\alpha}\right\|^2 \longrightarrow \min_{\tilde{\mathbf{t}},\boldsymbol{\alpha}} \tag{3.68}$$

bei

$$\mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h} \tag{3.69}$$

und

$$\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}. \tag{3.70}$$

Die auftretenden Vektoren, Matrizen und Matrixfunktionen haben die folgenden Dimensionen $\mathbf{y} \in \mathbb{R}^m$; $\boldsymbol{\alpha} \in \mathbb{R}^n$; $\tilde{\mathbf{t}} \in \mathbb{R}^l$; $\mathbf{h} \in \mathbb{R}^{ncstr}$; $\mathbf{U}, \mathbf{L} \in \mathbb{R}^{n-p}$; $\mathbf{B}(.) : \tilde{\mathbf{t}} \in \mathbb{R}^l \to \mathbf{B}(\tilde{\mathbf{t}}) \in \mathbb{R}^{m,n}$; $\mathbf{S}_r(.) : \tilde{\mathbf{t}} \in \mathbb{R}^l \to \mathbf{S}_r(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-r,n}$; $\mathbf{D}_p(.) : \tilde{\mathbf{t}} \in \mathbb{R}^l \to \mathbf{D}_p(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-p,n}$.

4 Die Struktur der Quadratmittelprobleme

Bei den zu untersuchenden Problemen handelt es sich um nichtlineare Quadratmittelprobleme in den Variablen $\boldsymbol{\alpha}$ und $\tilde{\mathbf{t}}$. Die Variable $\boldsymbol{\alpha}$ tritt jedoch in diesen Problemen stets nur linear auf. Solche Probleme werden nach Parks [Par85] als *restringierte semilineare Quadratmittelprobleme* (*constrained semilinear least squares problems, CSLS*) bezeichnet. Die allgemeine Form restringierter semilinearer Quadratmittelprobleme findet man in [Par85]. Wir werden uns aus Gründen der Vereinfachung der Notation auf den zu untersuchenden Spezialfall beschränken. Die Charakteristika solcher Probleme kommen jedoch schon hier zum Ausdruck.

Die Herangehensweise zur Lösung solcher spezieller Probleme besteht darin, nicht das *Ausgangsproblem* direkt, sondern ein *reduziertes Problem* in den nichtlinearen Variablen $\tilde{\mathbf{t}}$ zu lösen. Bei der Definition des reduzierten Problems spielt ein *Subproblem (A)* in den linearen Variablen $\boldsymbol{\alpha}$ zu festen nichtlinearen Variablen $\tilde{\mathbf{t}}$ eine wesentliche Rolle.

In den nächsten Abschnitten werden wir die Frage behandeln,

- wann der Übergang vom Ausgangsproblem zum reduzierten Problem zulässig ist,
- wie Kenntnisse über die Struktur von Subproblem (A) bei der effektiven Lösung des reduzierten Problems angewendet werden können.

Die Darstellung in den Kapiteln 4 und 5 folgt weitgehend dem Vorgehen in [Par85]. Da die Ergebnisse aus [Par85] nicht allgemein bekannt zu sein scheinen und die dort eingeführten Bezeichnungen im weiteren noch intensiv verwendet werden, geben wir die Hauptergebnisse dieser Arbeit hier an.

4.1 Ausgangsproblem, reduziertes Problem, Subproblem (A)

In den folgenden zwei Kapiteln lösen wir uns von dem speziellen Problem der Spline-Approximation mit freien Knoten und betrachten allgemeine Optimierungsprobleme mit einer speziellen Struktur. Im folgenden seien deshalb $\mathbf{y} \in \mathbb{R}^m$; $\mathbf{L}, \mathbf{U} \in \mathbb{R}^{n-p}$ und $\mathbf{h} \in \mathbb{R}^{ncstr}$ beliebige Vektoren, $\mathbf{C} \in \mathbb{R}^{ncstr,l}$ sei eine beliebige Matrix und

$$\mathbf{B}(.) : \tilde{\mathbf{t}} \in \mathbb{R}^l \longrightarrow \mathbf{B}(\tilde{\mathbf{t}}) \in \mathbb{R}^{m,n}$$

sowie

$$\mathbf{D}_p(.) : \tilde{\mathbf{t}} \in \mathbb{R}^l \longrightarrow \mathbf{D}_p(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-p,n}$$

seien beliebige, glatte Matrixfunktionen. Wir definieren die folgenden Optimierungsprobleme:

Das Ausgangsproblem

Sei $\mathfrak{F}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}$.

$$\text{(4.1)} \qquad \underset{\boldsymbol{\alpha}\in\mathbb{R}^n, \tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \ \mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 = \frac{1}{2}\mathfrak{F}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})^T \mathfrak{F}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$$

bei

$$\text{(4.2)} \qquad \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}$$

und

$$(4.3) \qquad \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \geq \mathbf{0}$$

Das reduzierte Problem

$$(4.4) \qquad \underset{\tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \quad f(\tilde{\mathbf{t}}) := \frac{1}{2}\mathbf{F}(\tilde{\mathbf{t}})^T \mathbf{F}(\tilde{\mathbf{t}})$$

bei

$$(4.5) \qquad \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0},$$

wobei

$$(4.6) \qquad \mathbf{F}(\tilde{\mathbf{t}}) := \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$$

und $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ löst das folgende Subproblem (A):

Subproblem (A)

$$(4.7) \qquad \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimiere}} \quad \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2$$

bei

$$(4.8) \qquad \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \geq \mathbf{0}.$$

4.2 Erste Bemerkungen zur Lösung des Problems

Eine erste, naheliegende Möglichkeit zur Lösung des Ausgangsproblems besteht in der *Methode der alternierenden Variablen.*

Algorithmus 1 (Methode der alternierenden Variablen)

S1: Wähle einen zulässigen Startpunkt $\tilde{\mathbf{t}}^0$, *d.h.* $\mathbf{C}\tilde{\mathbf{t}}^0 - \mathbf{h} \geq \mathbf{0}$

S2: Berechne die zulässigen Koeffizienten $\boldsymbol{\alpha}^0$

S3: Minimiere abwechselnd das Funktional $\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$ *bezüglich der nichtlinearen Variablen* $\tilde{\mathbf{t}}$ *bzw. bezüglich der linearen Variablen* $\boldsymbol{\alpha}$, *wobei jeweils die Werte der Variablen der anderen Variablengruppe konstant gehalten werden.*

Die Nachteile dieser Methode sind

- Es ist in jedem Fall nur lineare Konvergenz zu erwarten, vgl. die Konvergenzaussagen im unrestringierten Fall (Algorithm III (NIPALS) in [RW80]).

- Für festes $\boldsymbol{\alpha}$ sind die Nebenbedingungen nichtlinear in $\tilde{\mathbf{t}}$.

Falls die Voraussetzungen für den Übergang vom Ausgangsproblem auf das reduzierte Problem erfüllt sind, besteht eine erste Möglichkeit zur Lösung des reduzierten Problems in folgendem *naiven* Algorithmus:

Algorithmus 2 (Naiver Algorithmus für das reduzierte Problem)

S1: Wähle einen zulässigen Startpunkt $\tilde{\mathbf{t}}^0$, *d.h.* $\mathbf{C}\tilde{\mathbf{t}}^0 - \mathbf{h} \geq \mathbf{0}$

S2: Berechne die zulässigen Variablen $\boldsymbol{\alpha}^0 = \boldsymbol{\alpha}(\tilde{\mathbf{t}}^0)$, *d.h.* $\boldsymbol{\alpha}(\tilde{\mathbf{t}}^0)$ *löst Subproblem (A) bei festen* $\tilde{\mathbf{t}}^0$

S3: Minimiere das reduzierte Funktional $f(\tilde{\mathbf{t}})$ *mit einem ableitungsfreien verallgemeinerten Gauß-Newton-Verfahren*

Jedesmal wenn ein Funktionswert $\mathbf{F}(\tilde{\mathbf{t}})$ *benötigt wird, löse das Subproblem (A).*

Um die Äquivalenz von Ausgangs- und reduziertem Problem zu zeigen, benötigen wir Aussagen über Ableitungen des reduzierten Funktionals. Im Gegensatz zu Ableitungsinformationen des Ausgangsproblems sind diese nicht unmittelbar verfügbar, da das reduzierte Funktional selbst nur implizit durch das Subproblem (A) gegeben ist. Wir betrachten daher zunächst das Subproblem (A).

4.3 Qualitative Analyse von Subproblem (A)

Zu gegebenen, festen Parametern $\tilde{\mathbf{t}}$ ist das restringierte Quadratmittelproblem

$$P(\tilde{\mathbf{t}}): \qquad \begin{array}{ll} \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimiere}} & \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \\ \text{bei} & \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \geq \mathbf{0} \end{array}$$

zu lösen. Zu untersuchen ist die Abhängigkeit der Lösung $\boldsymbol{\alpha}$ von $P(\tilde{\mathbf{t}})$ gegenüber kleinen Störungen $\delta\tilde{\mathbf{t}}$ der Parameter $\tilde{\mathbf{t}}$, insbesondere stetige bzw. differenzierbare Abhängigkeit.

Theorem 3 (Hauptsatz der Sensitivitätstheorie)

(i) Die das Problem $P(\tilde{\mathbf{t}})$ *definierenden Funktionen seien zweimal stetig differenzierbar bzgl.* $\boldsymbol{\alpha}$.

Die Gradienten dieser Funktionen bzgl. $\boldsymbol{\alpha}$ *und die Nebenbedingungen seien einmal stetig differenzierbar bzgl.* $\tilde{\mathbf{t}}$ *in einer Umgebung von* $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}})$.

(ii) Die hinreichenden Optimalitätsbedingungen zweiter Ordnung für ein lokales Minimum von $P(\tilde{\mathbf{t}})$ *gelten an der Stelle* $\boldsymbol{\alpha}^*$ *mit zugehörigen Lagrange-Parametern* $\mathbf{u}^*$.

(iii) Die Gradienten für alle aktiven Nebenbedingungen an der Stelle $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}})$ *seien linear unabhängig.*

(iv) Strikte Komplementarität gelte.

Dann gilt:

1. $\boldsymbol{\alpha}^*$ *ist eine isolierte lokale Minimumstelle von* $P(\tilde{\mathbf{t}})$. *Die zugehörigen Lagrange-Parameter* $\mathbf{u}^*$ *sind eindeutig.*

2. *Für alle* $\delta\tilde{\mathbf{t}}$ *in einer Umgebung von* $\mathbf{0}$ *existiert eine eindeutige, einmal stetig differenzierbare Funktion* $\mathbf{y}(\tilde{\mathbf{t}} + \delta\tilde{\mathbf{t}}) = \left(\boldsymbol{\alpha}(\tilde{\mathbf{t}} + \delta\tilde{\mathbf{t}}), \mathbf{u}(\tilde{\mathbf{t}} + \delta\tilde{\mathbf{t}})\right)^T$, *welche die hinreichenden Optimalitätsbedingungen zweiter Ordnung für ein lokales Minimum von* $P(\tilde{\mathbf{t}} + \delta\tilde{\mathbf{t}})$ *erfüllt, so daß* $\mathbf{y}(\tilde{\mathbf{t}}) = (\boldsymbol{\alpha}^*, \mathbf{u}^*)$. *Also ist* $\boldsymbol{\alpha}(\tilde{\mathbf{t}} + \delta\tilde{\mathbf{t}})$ *eine*

lokal eindeutige lokale Minimumstelle von $P(\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}})$ mit den zugehörigen eindeutigen Lagrange-Parametern $\mathbf{u}(\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}})$.

3. *Für $\delta\tilde{\mathbf{t}}$ nahe $\mathbf{0}$ ändert sich die Menge der aktiven Restriktionen nicht mehr, strikte Komplementarität gilt und die Gradienten der bindenden Restriktionen an der Stelle $\boldsymbol{\alpha}(\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}})$ sind linear unabhängig.*

Beweis. Satz über implizite Funktionen auf System der Kuhn-Tucker-Bedingungen anwenden ∎

Bemerkung 12

(i) *Der obige Satz gilt allgemein für nichtlineare Optimierungsprobleme $P(\tilde{\mathbf{t}})$ der Form*

$$\underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimiere}}\ f(\boldsymbol{\alpha};\tilde{\mathbf{t}}) \qquad \text{bei} \qquad \mathbf{g}(\boldsymbol{\alpha};\tilde{\mathbf{t}}) \geq \mathbf{0} \qquad \text{und} \qquad \mathbf{h}(\boldsymbol{\alpha};\tilde{\mathbf{t}}) = \mathbf{0}.$$

Er liefert Sensitivitätsaussagen erster Ordnung für ein lokales Minimum, welches die Optimalitätsbedingungen zweiter Ordnung erfüllt, und scheint in dieser Form zuerst von Fiacco [Fia76] angegeben worden zu sein.

(ii) *Unter schwächeren Voraussetzungen (nichtstrikte Komplementarität) kann man zeigen, daß die Richtungsableitungen der Lösung und der Lagrange-Parameter nach den Parametern $\tilde{\mathbf{t}}$ existieren (siehe [BS74]). In [Fia83] wird ein Überblick über weitere Sensitivitätsaussagen der nichtlinearen Optimierung und ihre Anwendung gegeben.*

Wir können also schlußfolgern, daß unter obigen Bedingungen die Lösung $\boldsymbol{\alpha}$ von Subproblem (A) bei kleinen Störungen $\delta\tilde{\mathbf{t}}$ der Parameter $\tilde{\mathbf{t}}$ eine eindeutige, stetig differenzierbare Funktion der Störung ist, deren Wert an der Stelle $\delta\tilde{\mathbf{t}} = \mathbf{0}$ mit der Lösung des ungestörten Problems übereinstimmt.

Nach dieser qualitativen Untersuchung von Subproblem (A) gegenüber Störungen des Parameters $\tilde{\mathbf{t}}$ kommen wir nun zur quantitativen Untersuchung der Abhängigkeit. Das Ziel des nächsten Abschnittes ist die Gewinnung von Ausdrücken für $d\boldsymbol{\alpha}/d\tilde{\mathbf{t}}$ und $d\mathbf{u}/d\tilde{\mathbf{t}}$.

4.4 Quantitative Analyse von Subproblem (A)

Für die Lagrange-Funktion l zu Subproblem (A) erhalten wir

$$l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 - \mathbf{u}^T\left\{\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha} - \begin{pmatrix}\mathbf{L}\\ -\mathbf{U}\end{pmatrix}\right\}. \tag{4.9}$$

Wir vereinbaren folgende Bezeichnungen:

$\boldsymbol{\partial}$	Operator der Fréchet-Ableitung bzgl. $\tilde{\mathbf{t}}$
$\mathbf{g}$	$\mathbf{g} = \mathbf{g}(\boldsymbol{\alpha};\tilde{\mathbf{t}}) := \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \in \mathbb{R}^{2(n-p)}$ Vektor der Nebenbedingungen
$\mathcal{I}$	$\mathcal{I} := \big\{ i \in \{1, \ldots, 2(n-p)\} \,\vert\, g_i(\boldsymbol{\alpha};\tilde{\mathbf{t}}) = 0 \big\}$ Indexmenge der aktiven Nebenbedingungen
$nactiv$	$nactiv := \#\mathcal{I}$ Anzahl der aktiven Nebenbedingungen
$\bar{\mathbf{g}}$	$\bar{\mathbf{g}} := \{g_i \,\vert\, i \in \mathcal{I}\} \in \mathbb{R}^{nactiv}$ Vektor der aktiven Nebenbedingungen
$\mathbf{u}$	$\mathbf{u} = (u_1, \ldots, u_{2(n-p)})^T \in \mathbb{R}_+^{2(n-p)}$ Lagrange-Parameter
$\bar{\mathbf{u}}$	$\bar{\mathbf{u}} := \{u_i \,\vert\, i \in \mathcal{I}\} \in \mathbb{R}_+^{nactiv}$ Lagrange-Parameter zu aktiven Nebenbedingungen
$\boldsymbol{\xi}$	$\boldsymbol{\xi} := -\begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \in \mathbb{R}^{2(n-p)}$
$\bar{\boldsymbol{\xi}}$	$\bar{\boldsymbol{\xi}} := \{\xi_i \,\vert\, i \in \mathcal{I}\} \in \mathbb{R}^{nactiv}$
$\mathbf{R}$	$\mathbf{R} := -\left(\nabla_{\boldsymbol{\alpha}} \mathbf{g}\right)^T = -\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \in \mathbb{R}^{2(n-p),n}$
$\bar{\mathbf{R}}$	$\bar{\mathbf{R}} := -\left(\nabla_{\boldsymbol{\alpha}} \bar{\mathbf{g}}\right)^T = -\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i \in \mathcal{I}} \in \mathbb{R}^{nactiv,n}$
$\boldsymbol{\Gamma}$	$\boldsymbol{\Gamma} := -\left(\nabla_{\tilde{\mathbf{t}}} \mathbf{g}\right)^T = -\left(\nabla_{\tilde{\mathbf{t}}} \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}\right)^T \in \mathbb{R}^{2(n-p),l}$
$\bar{\boldsymbol{\Gamma}}$	$\bar{\boldsymbol{\Gamma}} := -\left(\nabla_{\tilde{\mathbf{t}}} \bar{\mathbf{g}}\right)^T = -\left(\nabla_{\tilde{\mathbf{t}}} \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i \in \mathcal{I}} \boldsymbol{\alpha}\right)^T \in \mathbb{R}^{nactiv,l}$

Parks zeigt, daß die Ableitungen der Lagrange-Multiplikatoren zu inaktiven Nebenbedingungen bei Störungen $\delta\tilde{\mathbf{t}}$ der Parameter $\tilde{\mathbf{t}}$ alle Null sind. Sie erhält folgendes System, welches die Abhängigkeit der Lösung $\boldsymbol{\alpha}$ und der Lagrange-Parameter $\bar{\mathbf{u}}$ von den Parametern $\tilde{\mathbf{t}}$ beschreibt:

$$\begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} l & \bar{\mathbf{R}}^T \\ \bar{\mathbf{R}} & \mathbf{0} \end{bmatrix} \begin{pmatrix} \boldsymbol{\partial}\boldsymbol{\alpha} \\ \boldsymbol{\partial}\bar{\mathbf{u}} \end{pmatrix} = -\begin{pmatrix} \nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l \\ \bar{\boldsymbol{\Gamma}} \end{pmatrix} \tag{4.10}$$

Nach der Regularitätsannahme und dem Ergebnis des Sensitivitätstheorems hat die Matrix $\bar{\mathbf{R}}$ vollen Zeilenrang *nactiv*. Wegen der hinreichenden Optimalitätsbedingungen zweiter Ordnung ist daher die Matrix $\begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} l & \bar{\mathbf{R}}^T \\ \bar{\mathbf{R}} & \mathbf{0} \end{bmatrix}$ regulär. Sei

$\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ eine Basis für den Nullraum von $\bar{\mathbf{R}}$. Dann gilt

$$\begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} l & \bar{\mathbf{R}}^T \\ \bar{\mathbf{R}} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^T & \mathbf{W}_{22} \end{bmatrix} \tag{4.11}$$

mit

$$\mathbf{W}_{11} = \mathbf{N}\left[\mathbf{N}^T\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\mathbf{N}\right]^{-1}\mathbf{N}^T \in \mathbb{R}^{n,n} \tag{4.12}$$

$$\mathbf{W}_{12} = \left[\mathbf{I} - \mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\right]\bar{\mathbf{R}}^+ \in \mathbb{R}^{n,nactiv} \tag{4.13}$$

$$\mathbf{W}_{22} = -\mathbf{W}_{12}^T\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\mathbf{W}_{12} \in \mathbb{R}^{nactiv,nactiv}. \tag{4.14}$$

Bemerkung 13

(i) Die Matrizen $\mathbf{W}_{11}$ und $\mathbf{W}_{22}$ sind symmetrisch.

(ii) $\bar{\mathbf{R}}^+$ sei die Moore-Penrose-Inverse von $\bar{\mathbf{R}}$. Sie kann durch eine beliebige verallgemeinerte Inverse $\bar{\mathbf{R}}^g$ ersetzt werden, d.h. $\bar{\mathbf{R}}\bar{\mathbf{R}}^g\bar{\mathbf{R}} = \bar{\mathbf{R}}$.

(iii) Da jede Spalte von $\mathbf{N}$ orthogonal zu den Gradienten der aktiven Nebenbedingungen ist, besagt die erfüllte Optimalitätsbedingung zweiter Ordnung, daß $\mathbf{N}^T\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\mathbf{N}$ positiv definit ist.

Damit erhalten wir das Hauptergebnis dieses Abschnittes

$$\begin{aligned} \partial\boldsymbol{\alpha} &= -\left\{\mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l\right) + \mathbf{W}_{12}\bar{\boldsymbol{\Gamma}}\right\} \\ \partial\bar{\mathbf{u}} &= -\left\{\mathbf{W}_{12}^T\left(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l\right) + \mathbf{W}_{22}\bar{\boldsymbol{\Gamma}}\right\} \end{aligned} \tag{4.15}$$

4.5 Analyse des reduzierten Problems

Nach der quantitativen Analyse von Subproblem (A) kommen wir nun zur Untersuchung des reduzierten Problems. Im folgenden zitieren wir die Ergebnisse aus [Par85]. Für jedes betrachtete $\tilde{\mathbf{t}}$ gelten dabei die Voraussetzungen des Hauptsatzes der Sensitivitätstheorie, Theorem 3.

Lemma 5 (Gradient des reduzierten Funktionals)

$$\nabla_{\tilde{\mathbf{t}}} f = \nabla_{\tilde{\mathbf{t}}} \mathfrak{f}(\boldsymbol{\alpha}(\tilde{\mathbf{t}}), \tilde{\mathbf{t}}) + \boldsymbol{\Gamma}^T\mathbf{u}$$

$\boldsymbol{\alpha}$ wird an der Stelle $\tilde{\mathbf{t}}$ berechnet.

Lemma 6 (Hesse-Matrix des reduzierten Funktionals)

$$\begin{aligned} \nabla^2_{\tilde{\mathbf{t}}} f &= \nabla^2_{\tilde{\mathbf{t}}} l - \left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l \,|\, \bar{\boldsymbol{\Gamma}}^T\right) \begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} l & \bar{\mathbf{R}}^T \\ \bar{\mathbf{R}} & \mathbf{0} \end{bmatrix}^{-1} \begin{pmatrix} \nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l \\ \bar{\boldsymbol{\Gamma}} \end{pmatrix} \\ &= \left(\partial\boldsymbol{\alpha}^T \,|\, \mathbf{I}\right) \begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} l & \nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l \\ \nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l & \nabla^2_{\tilde{\mathbf{t}}} l \end{bmatrix} \begin{pmatrix} \partial\boldsymbol{\alpha} \\ \mathbf{I} \end{pmatrix} \end{aligned}$$

Die Größen auf der rechten Seite werden an der Stelle $(\boldsymbol{\alpha}(\tilde{\mathbf{t}}), \tilde{\mathbf{t}})$ berechnet.

4.6 Die Beziehung zwischen Ausgangs- und reduziertem Problem

Das folgende Theorem zeigt, unter welchen Voraussetzungen der Übergang vom Ausgangsproblem zum reduzierten Problem berechtigt ist, und in welcher Beziehung die Lösungen der Probleme stehen. Es ist von ähnlicher grundlegender Bedeutung wie das entsprechende Theorem für separable Quadratmittelprobleme (siehe Theorem 2.1, [GP73]) und ist eine direkte Verallgemeinerung dieses Theorems.

Theorem 4 (Äquivalenz von Ausgangs- und reduziertem Problem)

1. *Die Funktion* $\mathfrak{f} = \mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$ *sei zweimal stetig differenzierbar bzgl.* $\boldsymbol{\alpha}$*, der Gradient* $\nabla_{\boldsymbol{\alpha}}\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$ *sei stetig differenzierbar bzgl.* $\tilde{\mathbf{t}}$.

2. *Jede Nebenbedingung sei stetig differenzierbar bzgl. ihrer Argumente.*

3. *Für jedes* $\tilde{\mathbf{t}}$ *habe das Subproblem (A) eine Lösung* $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$*, so daß*

   (a) *die hinreichenden Optimalitätsbedingungen zweiter Ordnung für ein lokales Minimum von Subproblem (A) an der Stelle* $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ *gelten,*

   (b) *die Gradienten bzgl.* $\boldsymbol{\alpha}$ *für die aktiven Nebenbedingungen von Subproblem (A) an der Stelle* $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ *linear unabhängig sind,*

   (c) *strikte Komplementarität für Subproblem (A) an der Stelle* $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ *gilt.*

Dann gilt für das Ausgangsproblem und das reduzierte Problem

(i) *Sei* $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ *eine globale Minimumstelle des Ausgangsproblems. Dann erfüllt* $\boldsymbol{\alpha}^*$ *die notwendigen Optimalitätsbedingungen erster Ordnung für das Subproblem (A),* $\tilde{\mathbf{t}}^*$ *ist globale Minimumstelle des reduzierten Problems und es gilt*

$$f(\tilde{\mathbf{t}}^*) = \mathfrak{f}(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*).$$

Wenn es ein eindeutiges $\boldsymbol{\alpha}^*$ *unter allen Paaren* $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$*, welche* $\mathfrak{f}$ *minimieren und denselben Minimalwert ergeben, gibt, so gilt*

$$\boldsymbol{\alpha}^* = \boldsymbol{\alpha}\left(\tilde{\mathbf{t}}^*\right).$$

(ii) *Wenn* $\tilde{\mathbf{t}}^*$ *die notwendigen Optimalitätsbedingungen erster Ordnung für das reduzierte Problem erfüllt, so erfüllt* $\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right)$ *die notwendigen Optimalitätsbedingungen erster Ordnung für das Ausgangsproblem.*

Bemerkung 14
Dieses Theorem gilt allgemein für nichtlineare Optimierungsprobleme der folgenden Form (für Details siehe [Par85]):

Ausgangsproblem

$$\underset{\boldsymbol{\alpha}\in\mathbb{R}^n, \tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \ \mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$$

bei

$$\begin{aligned} g_i(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) &\geq 0 && i = 1, \ldots, p_1 \\ h_i(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) &= 0 && i = 1, \ldots, p_2 \\ c_i(\boldsymbol{\alpha}) &\geq 0 && i = 1, \ldots, p_3 \\ d_i(\boldsymbol{\alpha}) &= 0 && i = 1, \ldots, p_4 \\ r_i(\tilde{\mathbf{t}}) &\geq 0 && i = 1, \ldots, p_5 \\ s_i(\tilde{\mathbf{t}}) &= 0 && i = 1, \ldots, p_6 \end{aligned}$$

Reduziertes Problem

$$\underset{\tilde{\mathbf{t}}\in\mathbb{R}^{l}}{minimiere}\;\; f(\tilde{\mathbf{t}}) := \mathfrak{f}(\boldsymbol{\alpha}(\tilde{\mathbf{t}}), \tilde{\mathbf{t}})$$

bei

$$\begin{aligned} r_i(\tilde{\mathbf{t}}) &\geq 0 \qquad & i &= 1, \ldots, p_5 \\ s_i(\tilde{\mathbf{t}}) &= 0 \qquad & i &= 1, \ldots, p_6 \end{aligned}$$

wobei $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ *das folgende Subproblem (A) löst:*

Subproblem (A)

$$\underset{\boldsymbol{\alpha}\in\mathbb{R}^{n}}{minimiere}\;\; \mathfrak{f}(\boldsymbol{\alpha}; \tilde{\mathbf{t}})$$

bei

$$\begin{aligned} g_i(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) &\geq 0 \qquad & i &= 1, \ldots, p_1 \\ h_i(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) &= 0 \qquad & i &= 1, \ldots, p_2 \\ c_i(\boldsymbol{\alpha}) &\geq 0 \qquad & i &= 1, \ldots, p_3 \\ d_i(\boldsymbol{\alpha}) &= 0 \qquad & i &= 1, \ldots, p_4 \end{aligned}$$

5 Restringierte semilineare Quadratmittelprobleme

In diesem Kapitel werden wir die Ergebnisse aus Kapitel 4 für allgemeine nichtlineare, reduzible Optimierungsprobleme auf restringierte semilineare Quadratmittelprobleme der Form (4.1) – (4.8) anwenden. Man beachte, daß (4.1) – (4.8) nicht die allgemeine Form eines CSLS darstellt, sondern nur den hier benötigten Spezialfall. Das Ausgangsproblem eines allgemeinen CSLS hat die Gestalt

$$\underset{\boldsymbol{\alpha}\in\mathbb{R}^n,\tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \ \mathfrak{f}(\boldsymbol{\alpha},\tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2$$

bei

$$\begin{aligned} \mathbf{G}(\tilde{\mathbf{t}})\boldsymbol{\alpha}-\boldsymbol{\beta}(\tilde{\mathbf{t}}) &\geq \mathbf{0} \\ \mathbf{H}(\tilde{\mathbf{t}})\boldsymbol{\alpha}-\boldsymbol{\delta}(\tilde{\mathbf{t}}) &= \mathbf{0} \\ \mathbf{C}\boldsymbol{\alpha}-\boldsymbol{\rho} &\geq \mathbf{0} \\ \mathbf{D}\boldsymbol{\alpha}-\boldsymbol{\sigma} &= \mathbf{0} \\ \mathbf{r}(\tilde{\mathbf{t}}) &\geq \mathbf{0} \\ \mathbf{s}(\tilde{\mathbf{t}}) &= \mathbf{0} \end{aligned}$$

Reduziertes Problem und Subproblem (A) werden dann gemäß Bemerkung 14 gebildet.
Zusätzlich zu den Bezeichnungen aus Abschnitt 4.4 vereinbaren wir folgende weitere Bezeichnungen:

N $\mathbf{N}\in\mathbb{R}^{n,n-nactiv}$
Basis des Nullraumes von $\bar{\mathbf{R}}$

$\mathfrak{J}$ $\mathfrak{J} := [\mathfrak{J}_{\boldsymbol{\alpha}}\,\mathfrak{J}_{\tilde{\mathbf{t}}}]\in\mathbb{R}^{m,n+l}$ mit $\mathfrak{J}_{\boldsymbol{\alpha}} := -\mathbf{B}\in\mathbb{R}^{m,n}$ und $\mathfrak{J}_{\tilde{\mathbf{t}}} := -\partial\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\in\mathbb{R}^{m,l}$
Jacobi-Matrix von $\mathfrak{F} := \mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}$

J $\mathbf{J}\in\mathbb{R}^{m,l}$
Jacobi-Matrix von $\mathbf{F} := \mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$

K $\mathbf{K} := \left(-\partial\mathbf{B}(\tilde{\mathbf{t}})^T\right)\left(\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right) - \left(\partial\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}^T\right)\mathbf{u}\in\mathbb{R}^{n,l}$

$\mathbf{S}_{\tilde{\mathbf{t}}}$ $\mathbf{S}_{\tilde{\mathbf{t}}} := -\mathbf{u}^T\partial^2\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}+\left(-\partial^2\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right)^T\left(\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right)\in\mathbb{R}^{l,l}$

Y $\mathbf{Y} := \left[(\mathbf{BN})^+\right]^T\mathbf{N}^T\mathbf{K}\in\mathbb{R}^{m,l}$

Z $\mathbf{Z} := \mathbf{B}\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}\in\mathbb{R}^{m,l}$

Die Lagrange-Funktion l von Subproblem (A) ist

$$l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2-\mathbf{u}^T\left\{\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}-\begin{pmatrix}\mathbf{L}\\ -\mathbf{U}\end{pmatrix}\right\}. \tag{5.1}$$

Dann gilt

$$(5.2)\quad \nabla_{\boldsymbol{\alpha}} l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) = \left(-\mathbf{B}(\tilde{\mathbf{t}})\right)^T \left(\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right) - \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}^T \mathbf{u} \in \mathbb{R}^n$$

$$(5.3)\quad \nabla_{\tilde{\mathbf{t}}} l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) = \left(-\partial\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right)^T \left(\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right) - \left(\partial \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}\right)^T \mathbf{u} \in \mathbb{R}^l.$$

Für die Komponenten der Hesse-Matrix von l erhalten wir

$$(5.4)\quad \nabla^2_{\boldsymbol{\alpha}} l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) = \mathbf{B}(\tilde{\mathbf{t}})^T\mathbf{B}(\tilde{\mathbf{t}}) \in \mathbb{R}^{n,n}$$

$$(5.5)\quad \begin{aligned} \nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) &= \left(-\mathbf{B}(\tilde{\mathbf{t}})\right)^T \left(-\partial\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right) + \mathbf{K} \\ &= -\mathbf{B}(\tilde{\mathbf{t}})^T \mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{K} \in \mathbb{R}^{n,l} \end{aligned}$$

$$(5.6)\quad \begin{aligned} \nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) &= \left(-\partial\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right)^T \left(-\mathbf{B}(\tilde{\mathbf{t}})\right) + \mathbf{K}^T \\ &= -\mathfrak{J}_{\tilde{\mathbf{t}}}^T \mathbf{B}(\tilde{\mathbf{t}}) + \mathbf{K}^T \in \mathbb{R}^{l,n} \end{aligned}$$

$$(5.7)\quad \begin{aligned} \nabla^2_{\tilde{\mathbf{t}}} l(\boldsymbol{\alpha},\mathbf{u};\tilde{\mathbf{t}}) &= \left(-\partial\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right)^T \left(-\partial\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right) + \mathbf{S}_{\tilde{\mathbf{t}}} \\ &= \mathfrak{J}_{\tilde{\mathbf{t}}}^T \mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{S}_{\tilde{\mathbf{t}}} \in \mathbb{R}^{l,l} \end{aligned}$$

mit

$$(5.8)\quad \mathbf{K} := \left(-\partial\mathbf{B}(\tilde{\mathbf{t}})^T\right)\left(\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right) - \left(\partial \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}^T\right) \mathbf{u} \in \mathbb{R}^{n,l}$$

und

$$(5.9)\quad \mathbf{S}_{\tilde{\mathbf{t}}} := -\mathbf{u}^T \partial^2 \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} + \left(-\partial^2\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right)^T \left(\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right) \in \mathbb{R}^{l,l}.$$

Bemerkung 15

(i) Der Term $\left(-\partial^2\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right)^T \left(\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right)$ *in* $\mathbf{S}_{\tilde{\mathbf{t}}}$ *fehlt in [Par85].*

(ii) Nur der Term $\mathbf{S}_{\tilde{\mathbf{t}}}$ *enthält zweite Ableitungen der im Problem auftretenden Matrizen. Er entspricht dem Term* $\mathbf{S} = \mathbf{F} \circ \mathbf{F}''$ *bei* $\nabla^2 \frac{1}{2}\|\mathbf{F}(\mathbf{x})\|^2$.

5.1 Der Gradient des reduzierten Funktionals

Satz 7 (Gradient des reduzierten Funktionals)

$$\nabla_{\tilde{\mathbf{t}}} f = \mathfrak{J}_{\tilde{\mathbf{t}}}^T \left(\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\right) + \mathbf{u}^T\boldsymbol{\Gamma}$$

5.2 Die Hesse-Matrix des reduzierten Funktionals

In Lemma 6 wurde gezeigt, daß

$$\begin{aligned} \nabla^2_{\tilde{\mathbf{t}}} f &= \nabla^2_{\tilde{\mathbf{t}}} l - \left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l \,|\, \bar{\boldsymbol{\Gamma}}^T\right) \begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} l & \bar{\mathbf{R}}^T \\ \bar{\mathbf{R}} & \mathbf{0} \end{bmatrix}^{-1} \begin{pmatrix} \nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l \\ \bar{\boldsymbol{\Gamma}} \end{pmatrix} \\ &= \nabla^2_{\tilde{\mathbf{t}}} l - \left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l \,|\, \bar{\boldsymbol{\Gamma}}^T\right) \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^T & \mathbf{W}_{22} \end{bmatrix} \begin{pmatrix} \nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l \\ \bar{\boldsymbol{\Gamma}} \end{pmatrix} \end{aligned}$$

mit

$$\mathbf{W}_{11} = \mathbf{N}\left[\mathbf{N}^T\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\mathbf{N}\right]^{-1}\mathbf{N}^T$$
$$\mathbf{W}_{12} = \left[\mathbf{I} - \mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\right]\bar{\mathbf{R}}^+$$
$$\mathbf{W}_{22} = -\mathbf{W}_{12}^T\left(\nabla^2_{\boldsymbol{\alpha}} l\right)\mathbf{W}_{12}.$$

Indem wir (5.4) – (5.7) in diesen Ausdruck einsetzen, erhalten wir

$$\nabla^2_{\tilde{\mathbf{t}}} f = \mathfrak{J}^T_{\tilde{\mathbf{t}}}\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{S}_{\tilde{\mathbf{t}}} - \left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l \,|\, \bar{\mathbf{\Gamma}}^T\right)\begin{bmatrix} \mathbf{W}_{11}\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l + \mathbf{W}_{12}\bar{\mathbf{\Gamma}} \\ \mathbf{W}_{12}^T\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l + \mathbf{W}_{22}\bar{\mathbf{\Gamma}} \end{bmatrix} \tag{5.10}$$

$$\begin{aligned} &= \mathfrak{J}^T_{\tilde{\mathbf{t}}}\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{S}_{\tilde{\mathbf{t}}} - \left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l\right)\mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l\right) - \left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l\right)\mathbf{W}_{12}\bar{\mathbf{\Gamma}} - \bar{\mathbf{\Gamma}}^T\mathbf{W}_{12}^T\left(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l\right) \\ &\quad - \bar{\mathbf{\Gamma}}^T\mathbf{W}_{22}\bar{\mathbf{\Gamma}}. \end{aligned} \tag{5.11}$$

Ich untersuche nun die Summanden von (5.11) einzeln:

$$\begin{aligned} \left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l\right)\mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l\right) &= \left(-\mathfrak{J}^T_{\tilde{\mathbf{t}}}\mathbf{B} + \mathbf{K}^T\right)\mathbf{W}_{11}\left(-\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{K}\right) \\ &= \mathfrak{J}^T_{\tilde{\mathbf{t}}}\mathbf{B}\mathbf{W}_{11}\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{K}^T\mathbf{W}_{11}\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathfrak{J}^T_{\tilde{\mathbf{t}}}\mathbf{B}\mathbf{W}_{11}\mathbf{K} + \mathbf{K}^T\mathbf{W}_{11}\mathbf{K} \end{aligned}$$

Allgemein gilt $\mathbf{P}_A = \mathbf{A}\mathbf{A}^+ = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$, falls $\mathbf{A} \in \mathbb{R}^{m,n}$ Vollrang n hat. Wir erhalten daher

$$\mathbf{B}\mathbf{W}_{11}\mathbf{B}^T = \mathbf{B}\mathbf{N}(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{B}^T = \mathbf{P}_{BN}.$$

Sei $\mathbf{Y} := \left[(\mathbf{B}\mathbf{N})^+\right]^T\mathbf{N}^T\mathbf{K} \in \mathbb{R}^{m,l}$.
Dann gilt

$$\begin{aligned} \mathbf{B}\mathbf{W}_{11}\mathbf{K} &= \mathbf{B}\mathbf{N}(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{K} \\ &= \mathbf{B}\mathbf{N}(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N})^{-T}\mathbf{N}^T\mathbf{K} \\ &= \left[(\mathbf{B}\mathbf{N})^+\right]^T\mathbf{N}^T\mathbf{K} \\ &= \mathbf{Y}. \end{aligned}$$

Weiterhin gilt

$$\begin{aligned} \mathbf{Y}^T\mathbf{Y} &= \mathbf{K}^T\mathbf{N}(\mathbf{B}\mathbf{N})^+\left((\mathbf{B}\mathbf{N})^+\right)^T\mathbf{N}^T\mathbf{K} \\ &= \mathbf{K}^T\mathbf{N}(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{B}^T\left(\mathbf{B}\mathbf{N}(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N})^{-T}\right)\mathbf{N}^T\mathbf{K} \\ &= \mathbf{K}^T\mathbf{N}(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{K} \\ &= \mathbf{K}^T\mathbf{W}_{11}\mathbf{K}. \end{aligned}$$

Wegen der Symmetrie von $\mathbf{W}_{11}$ erhalten wir

$$\mathbf{K}^T\mathbf{W}_{11}\mathbf{B}^T = \mathbf{K}^T\mathbf{W}_{11}^T\mathbf{B}^T = (\mathbf{B}\mathbf{W}_{11}\mathbf{K})^T = \mathbf{Y}^T.$$

Zusammenfassend ergibt sich für den ersten Summanden von (5.11)

$$\left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l\right)\mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}} l\right) = \mathfrak{J}^T_{\tilde{\mathbf{t}}}\mathbf{P}_{BN}\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{Y}^T\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathfrak{J}^T_{\tilde{\mathbf{t}}}\mathbf{Y} + \mathbf{Y}^T\mathbf{Y}. \tag{5.12}$$

Kommen wir nun zur Berechnung des zweiten Summanden

$$\left(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}} l\right)\mathbf{W}_{12}\bar{\mathbf{\Gamma}} = \left(-\mathfrak{J}^T_{\tilde{\mathbf{t}}}\mathbf{B} + \mathbf{K}^T\right)\mathbf{W}_{12}\bar{\mathbf{\Gamma}}.$$

Mit $\mathbf{Z} := \mathbf{B}\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}$ ergibt sich für $\mathbf{W}_{12}$

$$\mathbf{W}_{12} = \left\{\mathbf{I} - \mathbf{N}(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{B}^T\mathbf{B}\right\}\bar{\mathbf{R}}^+$$

und damit

$$\begin{aligned}\mathbf{B}\mathbf{W}_{12}\bar{\mathbf{\Gamma}} &= \left\{\mathbf{B} - \mathbf{B}\mathbf{N}(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N})^{-1}\mathbf{N}^T\mathbf{B}^T\mathbf{B}\right\}\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}\\ &= (\mathbf{I} - \mathbf{P}_{BN})\,\mathbf{B}\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}\\ &= \mathbf{P}_{BN}^{\perp}\mathbf{Z}.\end{aligned}$$

Analog erhalten wir

$$\begin{aligned}\mathbf{K}^T\mathbf{W}_{12}\bar{\mathbf{\Gamma}} &= \mathbf{K}^T\left(\mathbf{I} - \mathbf{W}_{11}\mathbf{B}^T\mathbf{B}\right)\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}\\ &= \mathbf{K}^T\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}} - \mathbf{K}^T\mathbf{W}_{11}\mathbf{B}^T\mathbf{Z}\\ &= \mathbf{K}^T\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}} - \mathbf{Y}^T\mathbf{Z}\end{aligned}$$

und schließlich

$$(\nabla^2_{\tilde{\mathbf{t}}\boldsymbol{\alpha}}l)\,\mathbf{W}_{12}\bar{\mathbf{\Gamma}} = -\mathfrak{J}_{\tilde{\mathbf{t}}}^T\mathbf{P}_{BN}^{\perp}\mathbf{Z} - \mathbf{Y}^T\mathbf{Z} + \mathbf{K}^T\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}. \tag{5.13}$$

Jetzt betrachte ich den Term

$$\begin{aligned}\bar{\mathbf{\Gamma}}^T\mathbf{W}_{12}^T(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}}l) &= \bar{\mathbf{\Gamma}}^T\left\{\left[\mathbf{I} - \mathbf{W}_{11}\mathbf{B}^T\mathbf{B}\right]\bar{\mathbf{R}}^+\right\}^T\left(-\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{K}\right)\\ &= \left\{\left(-\mathfrak{J}_{\tilde{\mathbf{t}}}^T\mathbf{B} + \mathbf{K}^T\right)\left[\mathbf{I} - \mathbf{W}_{11}\mathbf{B}^T\mathbf{B}\right]\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}\right\}^T\\ &= \left\{-\mathfrak{J}_{\tilde{\mathbf{t}}}^T\mathbf{P}_{BN}^{\perp}\mathbf{Z} - \mathbf{Y}^T\mathbf{Z} + \mathbf{K}^T\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}}\right\}^T,\end{aligned}$$

also

$$\bar{\mathbf{\Gamma}}^T\mathbf{W}_{12}^T(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}}l) = -\mathbf{Z}^T\mathbf{P}_{BN}^{\perp}\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{Z}^T\mathbf{Y} + \bar{\mathbf{\Gamma}}^T\left(\bar{\mathbf{R}}^+\right)^T\mathbf{K}. \tag{5.14}$$

Für den letzten Summanden erhält man schließlich

$$\begin{aligned}\bar{\mathbf{\Gamma}}^T\mathbf{W}_{22}\bar{\mathbf{\Gamma}} &= -\bar{\mathbf{\Gamma}}^T\mathbf{W}_{12}^T\mathbf{B}^T\mathbf{B}\mathbf{W}_{12}\bar{\mathbf{\Gamma}}\\ &= -\mathbf{Z}^T\left(\mathbf{P}_{BN}^{\perp}\right)^T\mathbf{P}_{BN}^{\perp}\mathbf{Z}\end{aligned}$$

$$\bar{\mathbf{\Gamma}}^T\mathbf{W}_{22}\bar{\mathbf{\Gamma}} = -\mathbf{Z}^T\mathbf{P}_{BN}^{\perp}\mathbf{Z}. \tag{5.15}$$

Setzt man (5.12)–(5.15) in (5.11) ein, so ergibt sich

Satz 8 (Hesse-Matrix des reduzierten Funktionals)

$$\begin{aligned}\nabla^2_{\tilde{\mathbf{t}}}f = {} & \mathfrak{J}_{\tilde{\mathbf{t}}}^T\mathbf{P}_{BN}^{\perp}\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathfrak{J}_{\tilde{\mathbf{t}}}^T\mathbf{Y} + \mathbf{Y}^T\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{Y}^T\mathbf{Y}\\ & + \mathfrak{J}_{\tilde{\mathbf{t}}}^T\mathbf{P}_{BN}^{\perp}\mathbf{Z} + \mathbf{Z}^T\mathbf{P}_{BN}^{\perp}\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{Y}^T\mathbf{Z} + \mathbf{Z}^T\mathbf{Y}\\ & - \mathbf{K}^T\bar{\mathbf{R}}^+\bar{\mathbf{\Gamma}} - \bar{\mathbf{\Gamma}}^T\left(\bar{\mathbf{R}}^+\right)^T\mathbf{K} + \mathbf{Z}^T\mathbf{P}_{BN}^{\perp}\mathbf{Z} + \mathbf{S}_{\tilde{\mathbf{t}}}\end{aligned}$$

Das reduzierte Problem ist ein nichtlineares Quadratmittelproblem in $\tilde{\mathbf{t}}$ mit linearen Ungleichheitsnebenbedingungen. Dieses Problem wollen wir mit einem verallgemeinerten Gauß-Newton-Verfahren lösen. Dazu benötigen wir die Residiumsfunktion $\mathbf{F}$ und die Jacobi-Matrix $\mathbf{J}$ des reduzierten Funktionals.

5.3 Die Residiumsfunktion des reduzierten Funktionals

Nach Definition ist $\mathbf{F}(\tilde{\mathbf{t}}) = \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$, wobei $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ das Subproblem (A) löst.

Satz 9 (Residiumsfunktion des reduzierten Funktionals)

$$\mathbf{F}(\tilde{\mathbf{t}}) = \mathbf{P}_{BN}^{\perp}\left(\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\bar{\mathbf{R}}^+(\tilde{\mathbf{t}})\bar{\boldsymbol{\xi}}(\tilde{\mathbf{t}})\right)$$

Beweis. siehe [Par85, Lemma 6.1], **Beachte:** Dort ist $\mathbf{F}(\tilde{\mathbf{t}}) = \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}}) - \mathbf{y}$! ∎

5.4 Die Jacobi-Matrix des reduzierten Funktionals

Es gilt $\mathbf{F}(\tilde{\mathbf{t}}) = \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ und damit

$$\begin{aligned}\mathbf{J}(\tilde{\mathbf{t}}) &= \partial\mathbf{y} - \partial\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}}) - \mathbf{B}(\tilde{\mathbf{t}})\partial\boldsymbol{\alpha} \\ &= \mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{B}\partial\boldsymbol{\alpha}.\end{aligned}$$

Nun gilt aber nach (4.15)

$$\begin{aligned}\partial\boldsymbol{\alpha} &= -\mathbf{W}_{11}\left(\nabla^2_{\boldsymbol{\alpha}\tilde{\mathbf{t}}}l\right) - \mathbf{W}_{12}\bar{\boldsymbol{\Gamma}} \\ &= -\mathbf{W}_{11}\left(-\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{K}\right) - \left(\mathbf{I} - \mathbf{W}_{11}\mathbf{B}^T\mathbf{B}\right)\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}} \\ &= \mathbf{W}_{11}\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{W}_{11}\mathbf{K} - \left(\mathbf{I} - \mathbf{W}_{11}\mathbf{B}^T\mathbf{B}\right)\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}.\end{aligned}$$

Für $\mathbf{B}\partial\boldsymbol{\alpha}$ erhalten wir

$$\begin{aligned}\mathbf{B}\partial\boldsymbol{\alpha} &= \mathbf{B}\mathbf{W}_{11}\mathbf{B}^T\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{B}\mathbf{W}_{11}\mathbf{K} - \left(\mathbf{B} - \mathbf{B}\mathbf{W}_{11}\mathbf{B}^T\mathbf{B}\right)\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}} \\ &= \mathbf{P}_{BN}\mathfrak{J}_{\tilde{\mathbf{t}}} - \mathbf{Y} - \mathbf{P}^{\perp}_{BN}\mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}.\end{aligned}$$

Die Jacobi-Matrix $\mathbf{J}(\tilde{\mathbf{t}})$ hat daher die Form

$$\begin{aligned}\mathbf{J}(\tilde{\mathbf{t}}) &= \mathbf{P}^{\perp}_{BN}\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{P}^{\perp}_{BN}\mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}} + \mathbf{Y} \\ &= \mathbf{P}^{\perp}_{BN}\left(\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\right) + \mathbf{Y}.\end{aligned}$$

Zum späteren Vergleich mit dem unrestringierten Fall formen wir $\mathbf{J}(\tilde{\mathbf{t}})$ noch geeignet um. Wegen $\mathbf{Y} = \left((\mathbf{BN})^+\right)^T\mathbf{N}^T\mathbf{K}$ und der Beziehung $\mathbf{P}_{BN}\left((\mathbf{BN})^+\right)^T = \left((\mathbf{BN})^+\right)^T$ gilt $\mathbf{Y} = \mathbf{P}_{BN}\mathbf{Y}$.

Satz 10 (Jacobi-Matrix des reduzierten Funktionals)

$$\mathbf{J}(\tilde{\mathbf{t}}) = \mathbf{P}^{\perp}_{BN}\left(\mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\right) + \mathbf{P}_{BN}\left((\mathbf{BN})^+\right)^T\mathbf{N}^T\mathbf{K}$$

Die Größen auf der rechten Seite sind jeweils an der Stelle $\tilde{\mathbf{t}}$ zu berechnen.

5.5 Struktur der Residiumsfunktion und der Jacobi-Matrix

In den letzten Aussagen kann man mit den bisherigen Bezeichnungen eine Struktur nur schwer erkennen. Wir definieren daher

$$\begin{aligned}\boldsymbol{\psi} &:= \mathfrak{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}} \\ \boldsymbol{\phi} &:= \left((\mathbf{BN})^+\right)^T\mathbf{N}^T\mathbf{K} \\ \mathbf{v} &:= \mathbf{y} - \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\xi}}.\end{aligned}$$

Mit $\mathbf{P} = \mathbf{P}_{BN}$ lauten die letzten Aussagen in den neuen Bezeichnungen

$$\begin{aligned}\mathbf{F} &= \mathbf{P}^{\perp}\mathbf{v} \\ \mathbf{J} &= \mathbf{P}^{\perp}\boldsymbol{\psi} + \mathbf{P}\boldsymbol{\phi}.\end{aligned}$$

Betrachten wir nun erneut den Gradienten des reduzierten Funktionals.

$$\begin{aligned}\mathbf{J}^T\mathbf{F} &= \left(\mathbf{P}^\perp\boldsymbol{\psi} + \mathbf{P}\boldsymbol{\phi}\right)^T \mathbf{P}^\perp\mathbf{v} \\ &= \boldsymbol{\psi}^T\left(\mathbf{P}^\perp\right)^T\mathbf{P}^\perp\mathbf{v} + \boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}^\perp\mathbf{v} \\ &= \boldsymbol{\psi}^T\left(\mathbf{P}^\perp\right)^T\mathbf{P}^\perp\mathbf{v}.\end{aligned}$$

Für die Hesse-Matrix des Gauß-Newton-Modells erhalten wir

$$\begin{aligned}\mathbf{J}^T\mathbf{J} &= \boldsymbol{\psi}^T\left(\mathbf{P}^\perp\right)^T\mathbf{P}^\perp\boldsymbol{\psi} + \boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}^\perp\boldsymbol{\psi} + \boldsymbol{\psi}^T\left(\mathbf{P}^\perp\right)^T\mathbf{P}\boldsymbol{\phi} + \boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi} \\ &= \boldsymbol{\psi}^T\left(\mathbf{P}^\perp\right)^T\mathbf{P}^\perp\boldsymbol{\psi} + \boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi}.\end{aligned}$$

Folgerung 3
Der Term $\mathbf{P}\boldsymbol{\phi}$ der Jacobi-Matrix $\mathbf{J} = \mathbf{P}^\perp\boldsymbol{\psi} + \mathbf{P}\boldsymbol{\phi}$ trägt nicht zum Gradienten $\mathbf{J}^T\mathbf{F}$ bei, der Beitrag zur Hesse-Matrix $\mathbf{J}^T\mathbf{J}$ ist $\boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi}$.

Mit $\mathbf{R} = \mathbf{0}$ und $\boldsymbol{\xi} = \mathbf{0}$ erhält man als Spezialfall die Ergebnisse für den unrestringierten Fall. Es gilt $\mathbf{N} = \mathbf{I}$ und $\boldsymbol{\alpha} = \mathbf{B}^+\mathbf{y}$. Damit ergibt sich

$$\begin{aligned}\mathbf{P} &:= \mathbf{B}\mathbf{B}^+ \\ \boldsymbol{\psi} &:= \mathfrak{J}_{\tilde{\mathbf{t}}} \\ &= -(\partial\mathbf{B})\boldsymbol{\alpha} \\ &= -(\partial\mathbf{B})\mathbf{B}^+\mathbf{y} \\ \boldsymbol{\phi} &:= \left(\mathbf{B}^+\right)^T\mathbf{K} \\ &= \left(\mathbf{B}^+\right)^T\left(-\partial\mathbf{B}^T\right)\left(\mathbf{y} - \mathbf{B}\boldsymbol{\alpha}\right) \\ &= \left(\mathbf{B}^+\right)^T\left(-\partial\mathbf{B}\right)^T\left(\mathbf{I} - \mathbf{B}\mathbf{B}^+\right)\mathbf{y} \\ &= -\left(\mathbf{P}^\perp(\partial\mathbf{B})\mathbf{B}^+\right)^T\mathbf{y} \\ \mathbf{v} &:= \mathbf{y}.\end{aligned}$$

Für die Jacobi-Matrix des reduzierten Funktionals bekommt man nun die Ergebnisse von Golub/Pereyra

$$\begin{aligned}\mathbf{J} &= \mathbf{P}^\perp\boldsymbol{\psi} + \mathbf{P}\boldsymbol{\phi} \\ &= -\mathbf{P}_B^\perp(\partial\mathbf{B})\mathbf{B}^+\mathbf{y} - \mathbf{P}_B\left(\mathbf{P}_B^\perp(\partial\mathbf{B})\mathbf{B}^+\right)^T\mathbf{y} \\ &= -\mathbf{P}_B^\perp(\partial\mathbf{B})\mathbf{B}^+\mathbf{y} - \left(\mathbf{P}_B^\perp(\partial\mathbf{B})\mathbf{B}^+\right)^T\mathbf{y}.\end{aligned}$$

Bemerkung 16
Im unrestringierten Fall wurde erstmals von Kaufman [Kau75] beobachtet, daß der Term $\mathbf{P}_B\boldsymbol{\phi}$ nicht zum Gradienten $\mathbf{J}^T\mathbf{F}$ des reduzierten Funktionals beiträgt, und daß der Beitrag zur Hesse-Matrix $\mathbf{J}^T\mathbf{J}$ von der Größenordnung $\mathcal{O}(\|\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\|^2)$ ist. Im Rahmen der Gauß-Newton-Philosophie kann dieser Term daher vernachlässigt werden. Eine direkte Herleitung dieser Eigenschaft sowie der Nachweis, daß die schnelle Konvergenz für Probleme mit kleinen Residuen erhalten bleibt, findet sich in [RW80] (siehe auch [Sch91]). Parks merkte in [Par85] ebenfalls an, daß sich die Struktur, welche von Kaufman ausgenutzt wurde, auf das allgemeine Problem überträgt. Eine Untersuchung des Einflusses der Approximation findet nicht statt.

Definition 6 (Kaufman-Approximation)
Die Approximation $\mathbf{J}_K \in \mathbb{R}^{m,l}$

$$\begin{aligned}\mathbf{J}_K &:= \mathbf{P}^\perp\boldsymbol{\psi} \\ &= \mathbf{P}^\perp\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\right)\end{aligned}$$

für die Jacobi-Matrix

$$\mathbf{J} = \mathbf{J}^{\perp}\boldsymbol{\psi} + \mathbf{P}\boldsymbol{\phi} = \mathbf{P}_{BN}^{\perp}\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right) + \mathbf{P}_{BN}^{\perp}\left((\mathbf{BN})^{+}\right)^{T}\mathbf{N}^{T}\mathbf{K}$$

des reduzierten Funktionals heißt Kaufman-Approximation.

Wir wollen nun den Einfluß der Verwendung dieser Approximation auf die Güte von lokalen Approximationen des reduzierten Funktionals untersuchen. Zunächst vergleichen wir die Hesse-Matrix $\mathbf{J}^T\mathbf{J}$ des Gauß-Newton-Modells $\mu_{GN}(\tilde{\mathbf{t}} + \mathbf{s}) = \frac{1}{2}\left\|\mathbf{F}(\tilde{\mathbf{t}}) + \mathbf{J}\mathbf{s}\right\|^2$; $\mathbf{J} = \mathbf{F}'(\tilde{\mathbf{t}})$ mit der Hesse-Matrix $\nabla^2_{\tilde{\mathbf{t}}} f$ des reduzierten Funktionals (des Newton-Modells $\mu_N(\tilde{\mathbf{t}} + \mathbf{s}) = \frac{1}{2}\left\|\mathbf{F}(\tilde{\mathbf{t}})\right\|^2 + \mathbf{J}^T\mathbf{F}(\tilde{\mathbf{t}})\mathbf{s} + \frac{1}{2}\mathbf{s}^T\nabla^2_{\tilde{\mathbf{t}}} f(\tilde{\mathbf{t}})\mathbf{s})$.

Lemma 11

$$\begin{aligned}\nabla^2_{\tilde{\mathbf{t}}} f &= \mathbf{J}^T\mathbf{J} - 2\mathbf{Y}^T\mathbf{Y} + \mathbf{Y}^T\mathbf{Z} + \mathbf{Z}^T\mathbf{Y} \\ &\quad - \mathbf{K}^T\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}} - \bar{\mathbf{\Gamma}}^T\left(\bar{\mathbf{R}}^{+}\right)^T\mathbf{K} + \mathbf{S}_{\tilde{\mathbf{t}}}\end{aligned}$$

Verwenden wir die Kaufman-Approximation $\mathbf{J}_K$ der Jacobi-Matrix, so erhalten wir

Lemma 12

$$\begin{aligned}\nabla^2_{\tilde{\mathbf{t}}} f &= \mathbf{J}_K^T\mathbf{J}_K - \mathbf{Y}^T\mathbf{Y} + \mathbf{Y}^T\mathbf{Z} + \mathbf{Z}^T\mathbf{Y} \\ &\quad - \mathbf{K}^T\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}} - \bar{\mathbf{\Gamma}}^T\left(\bar{\mathbf{R}}^{+}\right)^T\mathbf{K} + \mathbf{S}_{\tilde{\mathbf{t}}}\end{aligned}$$

Wir untersuchen jetzt die Größenordnung des Fehlers in der Hesse-Matrix $\mathbf{J}^T\mathbf{J}$ des Gauß-Newton-Modells, wenn wir anstelle von $\mathbf{J}$ mit der Kaufman-Approximation $\mathbf{J}_K$ arbeiten. Nach (3) gilt

$$\mathbf{J}^T\mathbf{J} = \mathbf{J}_K^T\mathbf{J}_K + \boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi}$$

mit

$$\begin{aligned}\boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi} &= \mathbf{Y}^T\mathbf{P}_{BN}^T\mathbf{P}_{BN}\mathbf{Y} \\ &= \mathbf{Y}^T\mathbf{Y} \\ &= \mathbf{K}^T\mathbf{W}_{11}\mathbf{K} \\ &= \mathbf{K}^T\mathbf{N}\left(\mathbf{N}^T\mathbf{B}^T\mathbf{B}\mathbf{N}\right)^{-1}\mathbf{N}^T\mathbf{K}.\end{aligned}$$

Für die Matrix $\mathbf{K}$ gilt

$$\mathbf{K} = \underbrace{\left(-\partial\mathbf{B}^T\right)(\mathbf{y} - \mathbf{B}\boldsymbol{\alpha})}_{\mathcal{O}(\|\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\|)} - \partial\begin{bmatrix}\mathbf{D}_p \\ -\mathbf{D}_p\end{bmatrix}^T\mathbf{u}.$$

Wegen des Auftretens des Terms $\partial\begin{bmatrix}\mathbf{D}_p \\ -\mathbf{D}_p\end{bmatrix}^T\mathbf{u}$ ist der zusätzliche Term $\boldsymbol{\phi}^T\mathbf{P}^T\mathbf{P}\boldsymbol{\phi}$ in der Hesse-Matrix im Gegensatz zum unrestringierten Fall nicht notwendigerweise von der Größenordnung $\mathcal{O}\left(\|\mathbf{y} - \mathbf{B}\boldsymbol{\alpha}\|^2\right)$. Man beachte jedoch, daß bereits beim Übergang vom Newton-Modell zum Gauß-Newton-Modell für das reduzierte Funktional Terme dieser Größenordnung vernachlässigt werden (z.B. $\mathbf{Y}^T\mathbf{Y}$, $\mathbf{K}^T\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}$ usw. siehe Lemma 11).

6 Anwendung auf Spline-Approximation und Spline-Glättung

In diesem Abschnitt wollen wir die beschriebene Reduktionstechnik auf die Quadratmittelapproximation und die Spline-Glättung durch B-Splines mit freien Knoten und Ungleichheitsnebenbedingungen an Ableitungen anwenden. Nachdem wir in diesem Kapitel die Frage nach der Existenz von Lösungen des reduzierten Problems sowie nach der Äquivalenz von Ausgangsproblem und reduziertem Problem sowohl für die Spline-Approximation als auch für die Spline-Glättung untersucht haben, wenden wir uns in den Kapiteln 7 und 8 jweweils der numerischen Lösung im Fall der Spline-Approximation bzw. im Fall der Spline-Glättung widmen.

Offensichtlich bleiben die Aussagen der letzten beiden Kapitel richtig, wenn man die *beliebigen* Vektoren, Matrizen und Matrixfunktionen durch die *speziellen* Bezeichnungen des vollständigen restringierten Approximationsproblems **FCAP** ((3.65), (3.66) und (3.67)) ersetzt. Zu den Aussagen für das vollständige restringierte Glättungsproblem **FCSP** ((3.68), (3.69) und (3.70)) gelangt man, indem man das Paar $\{\mathbf{B}(\tilde{\mathbf{t}}), \mathbf{y}\}$ durch das Paar

$$\left\{ \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}, \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \right\}$$

ersetzt. Wir erhalten daher die Optimierungsprobleme

Das Ausgangsproblem

$$\text{(6.1)} \qquad \underset{\boldsymbol{\alpha}\in\mathbb{R}^n, \tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \ \mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := \frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2$$

bzw.

$$\text{(6.2)} \qquad \underset{\boldsymbol{\alpha}\in\mathbb{R}^n, \tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \ \mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := \frac{1}{2} \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right\|^2$$

bei

$$\text{(6.3)} \qquad \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}$$

und

$$\text{(6.4)} \qquad \mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}.$$

Das reduzierte Problem

$$\text{(6.5)} \qquad \underset{\tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \ f(\tilde{\mathbf{t}}) := \frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}}) \right\|^2$$

bzw.

$$\text{(6.6)} \qquad \underset{\tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \ f(\tilde{\mathbf{t}}) := \frac{1}{2} \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}) \right\|^2$$

bei

$$\text{(6.7)} \qquad \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0},$$

wobei $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ das Subproblem (A) löst.

Subproblem (A)

$$\text{(6.8)} \qquad \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimiere}} \ \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2$$

bzw.

$$\text{(6.9)} \qquad \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimiere}} \ \frac{1}{2}\left\|\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}-\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}\right\|^2$$

bei

$$\text{(6.10)} \qquad \mathbf{L} \le \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \le \mathbf{U}.$$

Das Problem (6.5), (6.7), (6.8), (6.10) bezeichnen wir als *reduziertes restringiertes Approximationsproblem* (reduced constrained approximation problem, **RCAP**), das Problem (6.6), (6.7), (6.9), (6.10) als *reduziertes restringiertes Glättungsproblem* (reduced constrained smoothing problem, **RCSP**).

6.1 Die Voraussetzungen des Hauptsatzes der Sensitivitätstheorie

Das wesentliche Hilfsmittel, um den Übergang vom Ausgangsproblem zum reduzierten Problem zu zeigen, besteht im Hauptsatz der Sensitivitätstheorie (Theorem 3). Wir werden deshalb im folgenden die Voraussetzungen dieses Satzes im einzelnen untersuchen. Ausgangspunkt ist das Problem

$$P(\tilde{\mathbf{t}}): \qquad \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimiere}} \ \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \text{ bei } \begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}-\begin{pmatrix}\mathbf{L}\\ -\mathbf{U}\end{pmatrix}\ge\mathbf{0}$$

bzw.

$P(\tilde{\mathbf{t}})$:

$$\underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimiere}} \ \frac{1}{2}\left\|\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}-\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}\right\|^2 \text{ bei } \begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}-\begin{pmatrix}\mathbf{L}\\ -\mathbf{U}\end{pmatrix}\ge\mathbf{0}.$$

Im folgenden seien die folgenden Voraussetzungen erfüllt:

Voraussetzung 1
Für $0 \le q < k$ mit $q \le n-1$ gelte

$$t_j < t_{j+k-q} \qquad (j = q+1, \ldots, n).$$

Voraussetzung 2
Die freien Knoten $\tilde{\mathbf{t}}$ seien einfache Knoten, d.h.

$$\#t_{p(j)} = 1 \qquad (j = 1, \ldots, l).$$

Es gelte $k \ge 3$.

Bemerkung 17

(i) *Unter der Voraussetzung 1 existieren die Elemente der Matrizen $\mathbf{B}$, $\mathbf{D}_p$ und $\mathbf{S}_r$ mit $0 \le p, r \le q$ und sind stetige Funktionen der Knoten.*

(ii) *Unter der Voraussetzung 2 besitzt der Spline s nach dem Satz von Curry-Schoenberg an den Stellen* $t_{p(j)}$ $(j = 1, \ldots, l)$ *die maximale Glattheit, genauer* $s \in \mathbb{C}^{k-2}$ *in einer Umgebung von* $t_{p(j)}$. *Falls* $k \geq 3$ *ist, so existieren nach Folgerung 1 die Ableitungen der B-Splines nach den freien Knoten für alle* $x \in [a, b]$.

(iii) *Die Knotenvielfachheiten* $\#t_j$ *können sich wegen der Nebenbedingung* $\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}$ *im Laufe des Optimierungsprozesses nicht ändern, d.h. falls eine Knotenfolge* $\mathbf{t}^0$ *(bzw.* $\tilde{\mathbf{t}}^0$*) die Voraussetzungen 1 und 2 erfüllt, so erfüllen alle Knotenfolgen aus dem zulässigen Bereich* $\{\tilde{\mathbf{t}} \in \mathbb{R}^l : \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}\}$ *ebenfalls diese Voraussetzungen.*

6.1.1 Differenzierbarkeit der Problemfunktionen

Es gilt

$$\nabla_{\boldsymbol{\alpha}} \left(\frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2 \right) = -\mathbf{B}(\tilde{\mathbf{t}})^T \left(\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right) \tag{6.11}$$

$$\nabla^2_{\boldsymbol{\alpha}} \left(\frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2 \right) = \mathbf{B}(\tilde{\mathbf{t}})^T \mathbf{B}(\tilde{\mathbf{t}}) \tag{6.12}$$

bzw.

$$\begin{aligned} \nabla_{\boldsymbol{\alpha}} \left(\frac{1}{2} \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right\|^2 \right) & \\ = \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r\tilde{\mathbf{t}}) \end{bmatrix}^T & \left(\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right) \end{aligned} \tag{6.13}$$

$$\nabla^2_{\boldsymbol{\alpha}} \left(\frac{1}{2} \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right\|^2 \right) = - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r\tilde{\mathbf{t}}) \end{bmatrix}^T \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r\tilde{\mathbf{t}}) \end{bmatrix} \tag{6.14}$$

und

$$\nabla_{\boldsymbol{\alpha}} \left(\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \right) = \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}^T \tag{6.15}$$

$$\nabla^2_{\boldsymbol{\alpha}} \left(\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \right) = \mathbf{0}. \tag{6.16}$$

Da die Matrixfunktionen $\mathbf{B}(.)$, $\mathbf{D}_p(.)$ und $\mathbf{S}_r(.)$ wegen Voraussetzung 1 stetig sind, ist die zweimalige stetige Differenzierbarkeit der Problemfunktionen bzgl. $\boldsymbol{\alpha}$ gesichert. Weiterhin gilt für die Gradienten dieser Funktionen bzgl. $\boldsymbol{\alpha}$ und die Nebenbedingungen

$$\nabla_{\tilde{\mathbf{t}}} \nabla_{\boldsymbol{\alpha}} \left(\frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2 \right)_{|\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}}} = \nabla_{\tilde{\mathbf{t}}} \left(-\mathbf{B}(\tilde{\mathbf{t}})^T \left(\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right) \right)_{|\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}}} \tag{6.17}$$

$$(6.18)\quad \nabla_{\tilde{\mathbf{t}}}\nabla_{\boldsymbol{\alpha}}\left(\frac{1}{2}\left\|\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}-\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}\right\|^2\right)_{|\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}}}$$

$$=\nabla_{\tilde{\mathbf{t}}}\left(-\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r\tilde{\mathbf{t}})\end{bmatrix}^T\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}-\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}\right)\right)_{|\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}}}$$

$$(6.19)\quad \nabla_{\tilde{\mathbf{t}}}\nabla_{\boldsymbol{\alpha}}\left(\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}-\begin{pmatrix}\mathbf{L}\\ -\mathbf{U}\end{pmatrix}\right)_{\tilde{\mathbf{t}}}\nabla_{\boldsymbol{\alpha}}=\nabla_{\tilde{\mathbf{t}}}\left(\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}^T\right)_{|\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}}}$$

und

$$(6.20)\quad \nabla_{\tilde{\mathbf{t}}}\left(\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}-\begin{pmatrix}\mathbf{L}\\ -\mathbf{U}\end{pmatrix}\right)_{|\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}}}=\nabla_{\tilde{\mathbf{t}}}\left(\begin{bmatrix}\mathbf{D}_p(\tilde{\mathbf{t}})\\ -\mathbf{D}_p(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}\right)_{|\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}}}$$

Wegen Voraussetzung 2 existiert eine Umgebung $\mathcal{U}$ von $\tilde{\mathbf{t}}$ so, daß für alle $\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}}\in\mathcal{U}$ die Elemente von $\mathbf{B}(\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}})$, $\mathbf{D}_p(\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}})$ und $\mathbf{S}_r(\tilde{\mathbf{t}}+\delta\tilde{\mathbf{t}})$ definiert sind, und als Summen bzw. Produkte von rationalen Funktionen in t_j stetig differenzierbar sind. Damit sind auch die Gradienten der Problemfunktionen bzgl. $\boldsymbol{\alpha}$ und die Nebenbedingungen stetig differenzierbar bzgl. $\tilde{\mathbf{t}}$ in einer Umgebung von $(\boldsymbol{\alpha}^*,\tilde{\mathbf{t}})$ ($\boldsymbol{\alpha}^*$ – lokale Minimumstelle von $P(\tilde{\mathbf{t}})$).

6.1.2 Hinreichende Optimalitätsbedingungen zweiter Ordnung

Voraussetzung 3

(i) Für das Problem der Spline-Approximation sei die Whitney-Schoenberg-Bedingung an der Stelle $\tilde{\mathbf{t}}$ erfüllt.

(ii) Für das Problem der Spline-Glättung gelte $m\geq r$ und $\mu>0$.

Unter der Voraussetzung 3 ist $\mathbf{B}(\tilde{\mathbf{t}})^T\mathbf{B}(\tilde{\mathbf{t}})$ bzw. $\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}})\end{bmatrix}^T\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}})\end{bmatrix}$ positiv definit und das Quadratmittelproblem $P(\tilde{\mathbf{t}})$ hat genau eine Lösung. Die Lösung $\boldsymbol{\alpha}^*$ zu fester Knotenfolge $\tilde{\mathbf{t}}$ erfüllt die hinreichenden Optimalitätsbedingungen zweiter Ordnung an der Stelle $\tilde{\mathbf{t}}$. Wenn die Gradienten der aktiven Nebenbedingungen linear unabhängig sind, existieren eindeutige Lagrange-Parameter $\mathbf{u}^*$.

6.1.3 Lineare Unabhängigkeit der Gradienten der aktiven Restriktionen

Die Gradienten der aktiven Restriktionen bzgl. $\boldsymbol{\alpha}$ sind linear unabhängig, wenn dies für die Zeilen von $\bar{\mathbf{R}}$ gilt.

Wir betrachten zunächst die Nebenbedingung $\mathbf{g}=\mathbf{g}(\boldsymbol{\alpha};\tilde{\mathbf{t}}):=\mathbf{D}_p\boldsymbol{\alpha}-\mathbf{L}\geq\mathbf{0}$. Die Gradienten der aktiven Restriktionen sind in diesem Fall linear unabhängig, wenn die Zeilen von $\mathbf{D}_p$, für welche die Nebenbedingungen aktiv sind, linear unabhängig sind. Das ist auf jeden Fall gesichert, wenn *alle* Zeilen von $\mathbf{D}_p$ linear unabhängig sind, mit anderen Worten $\operatorname{rank}\mathbf{D}_p=n-p$.

Lemma 13

$\operatorname{rank}\mathbf{D}_p=n-p\Longleftrightarrow(\mathbf{D}_p)_{ii}\neq 0\ (i=1,\ldots,n-p)$

Beweis. Klar, da $\mathbf{D}_p$ eine obere Dreiecksmatrix ist. ∎

Lemma 14

$$(\mathbf{D}_p)_{ii} = (-1)^p \prod_{\nu=1}^{p} (k-\nu) \frac{1}{t_{k+i} - t_{i+\nu}} \quad \textit{für} \quad i = 1, \ldots, n-p \quad \textit{und} \quad p \geq 1.$$

Beweis. Beweis durch vollständige Induktion ∎

Unter der Voraussetzung $t_j < t_{j+k-p}$ $(j = p+1, \ldots, n)$ folgt damit

Satz 15
Die Gradienten der aktiven Nebenbedingungen zu $\mathbf{g} = \mathbf{g}(\boldsymbol{\alpha}; \tilde{\mathbf{t}}) := \mathbf{D}_p \boldsymbol{\alpha} - \mathbf{L} \geq \mathbf{0}$ *sind linear unabhängig.*

Betrachten wir nun die ursprünglichen Nebenbedingungen $\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$ bzw.

$$\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \geq \mathbf{0}. \tag{6.21}$$

Wenn $\mathbf{L} < \mathbf{U}$ gilt, so werden die Nebenbedingungen für die i-te Komponente von $\mathbf{g}$ entweder an der Stelle L_i oder an der Stelle U_i aktiv, aber niemals gleichzeitig. Für die Matrix $\bar{\mathbf{R}} := -\left(\nabla_{\boldsymbol{\alpha}} \mathbf{g}_i^T\right)_{i \in \mathcal{I}} = -\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i \in \mathcal{I}}$ bedeutet das, daß niemals die gleichen Zeilen (bis auf den Faktor -1) im oberen und unteren Teil auftreten können. Multiplizieren wir die entsprechenden Zeilen noch mit dem Faktor -1, so enthält $\bar{\mathbf{R}}$ ausschließlich Zeilen von $\mathbf{D}_p$. Also sind die Zeilen von $\bar{\mathbf{R}}$ linear unabhängig. Umgekehrt enthält $\bar{\mathbf{R}}$ zwei (bis auf den Faktor -1) identische Zeilen von $\mathbf{D}_p$, falls $L_i = U_i$, d.h. $\bar{\mathbf{R}}$ hat dann keinen Vollrang.

Satz 16 (Lineare Unabhängigkeit der Gradienten)
Sei $t_j < t_{j+k-p}$ $(j = p+1, \ldots, n)$ *und sei*

$$\mathbf{g} = \mathbf{g}(\boldsymbol{\alpha}; \tilde{\mathbf{t}}) := \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} - \begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix} \geq \mathbf{0}$$

die Nebenbedingung von Problem $P(\tilde{\mathbf{t}})$*. Die Gradienten der aktiven Nebenbedingungen sind linear unabhängig, d.h.*

$$\bar{\mathbf{R}} := -\left(\nabla_{\boldsymbol{\alpha}} \mathbf{g}_i^T\right)_{i \in \mathcal{I}} = -\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i \in \mathcal{I}} \in \mathbb{R}^{nactiv, n}$$

hat Vollrang rank $\bar{\mathbf{R}} = nactiv = \#\mathcal{I}$ *genau dann, wenn die* strikte Konsistenzbedingung

$$\begin{aligned} L_i &< U_i \qquad i = 1, \ldots, n-p \\ \mathbf{L} &< \mathbf{U} \end{aligned}$$

gilt.

Bemerkung 18

(i) *Die strikte Konsistenzbedingung* $\mathbf{L} < \mathbf{U}$ *stellt eine Verschärfung der Konsistenzbedingung* $\mathbf{L} \leq \mathbf{U}$ *dar. Sie bedeutet insbesondere, daß nicht* $s^{(p)}(x) \equiv 0$ *für alle* $x \in [t_i, t_j]$, $j > i$ *bzw.* $\mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \equiv 0$ *gefordert werden darf.*

Beachte! *Die Bedingung* $l_i^{(p)} \leq s^{(p)}(x) \leq u_i^{(p)}$ *für alle* $x \in [t_i, t_{i+1})$, $i = k, \ldots, n$ *mit* $l_i^{(p)} < u_i^{(p)}$ *ist* ***nicht*** *hinreichend für strikte Konsistenz (siehe Beispiele in Abschnitt 3.3.2).*

(ii) *Die Bedingung der strikten Konsistenz ist in diesem Fall mit der Slater-Bedingung, d.h.* $\exists \boldsymbol{\alpha}^0 \in \mathbb{R}^n : \mathbf{g}(\boldsymbol{\alpha}^0; \tilde{\mathbf{t}}) > \mathbf{0}$ *identisch.*

6.1.4 Strikte Komplementarität

Seien u_i $(i = 1, \ldots, 2(n-p))$ die Lagrange-Parameter zu Problem $P(\tilde{\mathbf{t}})$. Strikte Komplementarität liegt vor, wenn $u_i > 0$ für alle $i \in \mathcal{I}(\boldsymbol{\alpha}^*)$.

Beispiel 7 (Nichtstrikte Komplementarität)

$f := x_1^2 + x_2^2 \longrightarrow \min \qquad$ *bei* $\qquad g := x_2 \geq 0$

$l = x_1^2 + x_2^2 - u x_2,\ u \in \mathbb{R}_+$

$$\nabla l = \begin{pmatrix} 2x_1 \\ 2x_2 - u \end{pmatrix} = 0 \qquad \Longrightarrow x_1 = x_2 = 0,\ u = 0$$

Der Fall der nichtstrikten Komplementarität kann auch für lineare Nebenbedingungen auftreten. Er tritt nicht generisch auf, d.h. ist nicht stabil gegenüber Störungen (Mit anderen Worten: Die Menge der Punkte, für die strikte Komplementarität eintritt, ist offen).

Alle Aussagen über die stetig differenzierbare Abhängigkeit der Lösung von den Störungen benötigen die strikte Komplementarität. Man hat jedoch zumindest die Lipschitz-stetige Abhängigkeit der Lösung quadratischer Optimierungsprobleme bei nichtstrikter Komplementarität.

Satz 17 (J. W. Daniel [Dan73])

Es sei

$$\begin{aligned}
\mathbf{f}(\mathbf{x}) &= \mathbf{b}^T\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} && \mathbf{A} \in \mathbb{R}^{n,n} \text{ symmetrisch, positiv definit} \\
\mathbf{h}(\mathbf{x}) &= \mathbf{H}^T\mathbf{x} + \mathbf{h}^0 && \mathbf{H} \in \mathbb{R}^{n,p} \text{ spaltenregulär} \\
\mathbf{g}(\mathbf{x}) &= \mathbf{G}^T\mathbf{x} + \mathbf{g}^0 && \mathbf{G} \in \mathbb{R}^{n,m}.
\end{aligned}$$

Es existiere $\mathbf{x}^0$ *mit* $\mathbf{h}(\mathbf{x}^0) = \mathbf{0}$, $\mathbf{g}(\mathbf{x}^0) > \mathbf{0}$ *(Slater-Bedingung).* $\mathbf{x}^*$ *sei die eindeutig bestimmte Minimallösung von*

$$\mathbf{x}^* = argmin\left\{\mathbf{f}(\mathbf{x}) \,:\, \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \mathbf{x} \in \mathbb{R}^n\right\}.$$

Ferner sei

$$\begin{aligned}
\tilde{\mathbf{f}}(\mathbf{x}) &= \tilde{\mathbf{b}}^T\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\tilde{\mathbf{A}}\mathbf{x} \\
\tilde{\mathbf{h}}(\mathbf{x}) &= \tilde{\mathbf{H}}^T\mathbf{x} + \tilde{\mathbf{h}}^0 \\
\tilde{\mathbf{g}}(\mathbf{x}) &= \tilde{\mathbf{G}}^T\mathbf{x} + \tilde{\mathbf{g}}^0 \quad \text{mit} \quad \tilde{\mathbf{H}} \in \mathbb{R}^{n,p}, \tilde{\mathbf{G}} \in \mathbb{R}^{n,m}.
\end{aligned}$$

Falls dann $||\mathbf{A}-\tilde{\mathbf{A}}||$, $||\mathbf{H}-\tilde{\mathbf{H}}||$, $||\mathbf{G}-\tilde{\mathbf{G}}||$, $||\mathbf{b}-\tilde{\mathbf{b}}||$, $||\mathbf{h}^0-\tilde{\mathbf{h}}^0||$ *und* $||\mathbf{g}^0-\tilde{\mathbf{g}}^0||$ *hinreichend klein sind, dann ist auch*

$$\tilde{\mathbf{x}}^* = argmin\left\{\tilde{\mathbf{f}}(\mathbf{x}) \,:\, \tilde{\mathbf{h}}(\mathbf{x}) = \mathbf{0}, \tilde{\mathbf{g}}(\mathbf{x}) \geq \mathbf{0}, \mathbf{x} \in \mathbb{R}^n\right\}$$

eindeutig lösbar und es gilt

$$||\tilde{\mathbf{x}}^* - \mathbf{x}^*|| \leq C \max\Big\{||\mathbf{A} - \tilde{\mathbf{A}}||, ||\mathbf{H} - \tilde{\mathbf{H}}||, ||\mathbf{G} - \tilde{\mathbf{G}}||, ||\mathbf{b} - \tilde{\mathbf{b}}||, ||\mathbf{h}^0 - \tilde{\mathbf{h}}^0||, ||\mathbf{g}^0 - \tilde{\mathbf{g}}^0||\Big\}$$

mit einer Konstanten $C = C(\mathbf{x}^*)$.

6.2 Existenz einer Lösung des reduzierten Problems

Theorem 5 (Existenz einer Lösung des reduzierten Problems)
Wir betrachten die Menge der zulässigen Knotenfolgen $\{\tilde{\mathbf{t}} \in \mathbb{R}^l : \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}\}$. *Die folgenden Voraussetzungen seien erfüllt:*

(V1) Für $0 \leq q < k$ *mit* $q \leq n-1$ *gelte*

$$t_j < t_{j+k-q} \qquad (j = q+1, \ldots, n).$$

(V3) Sei $0 \leq r \leq q$. *Für das Problem der Spline-Glättung gelte* $m \geq r$ *und* $\mu > 0$.

(V4) Sei $0 \leq p \leq q$. *Die strikte Konsistenzbedingung* $\mathbf{L} < \mathbf{U}$ *für die Nebenbedingung* $\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$ *sei erfüllt.*

*Dann gilt: Das reduzierte, restringierte Glättungsproblem **RCSP** (6.6), (6.7), (6.9), (6.10) besitzt eine Lösung* $\tilde{\mathbf{t}}^*$.

Beweis. Aus (V1) folgt die Existenz der Matrizen $\mathbf{S}_r$ und $\mathbf{D}_p$ für $0 \leq r, p \leq q$. Die Matrixfunktionen $\mathbf{B}(.)$, $\mathbf{S}_r(.)$ und $\mathbf{D}_p(.)$ sind stetige Funktionen der Knoten.
Wegen (V3) hat die Matrix $\begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}$ Vollrang n für alle Knotenfolgen $\mathbf{t}$, d.h. die Hesse-Matrix von Subproblem (A) ist positiv definit.
Vorausetzung (V4) sichert schließlich die Slater-Bedingung $\mathbf{g}(\boldsymbol{\alpha}^0) > \mathbf{0}$ (siehe Bemerkung 18).
Mit

$$\begin{aligned}
\mathbf{b}(\tilde{\mathbf{t}}) &:= -\begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \\
\mathbf{A}(\tilde{\mathbf{t}}) &:= \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}^T \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \\
\mathbf{G}(\tilde{\mathbf{t}}) &:= \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}^T \\
\mathbf{g}^0 &:= -\begin{pmatrix} \mathbf{L} \\ -\mathbf{U} \end{pmatrix}
\end{aligned}$$

ist Subproblem (A) äquivalent zu

$$\underset{\boldsymbol{\alpha} \in \mathbb{R}^n}{\text{minimiere}} \ \mathbf{b}(\tilde{\mathbf{t}})^T \boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{A}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \quad \text{bei} \quad \mathbf{G}(\tilde{\mathbf{t}})^T \boldsymbol{\alpha} + \mathbf{g}^0 \geq \mathbf{0}$$

Auf dieses Problem können wir nun Satz 17 anwenden. Das entsprechende gestörte Problem erhalten wir, indem wir den Parameter $\tilde{\mathbf{t}}$ durch $\tilde{\mathbf{t}} + \delta\tilde{\mathbf{t}}$ ersetzen. Auf

Grund der Stetigkeit der Matrixfunktionen $\mathbf{B}(.)$, $\mathbf{S}_r(.)$ und $\mathbf{D}_p(.)$ ändern sich die Größen $\mathbf{b}$, $\mathbf{A}$, $\mathbf{G}$ und $\mathbf{g}^0$ bei kleinen Störungen $\delta\tilde{\mathbf{t}}$ ebenfalls nur wenig. Wir erhalten aus Satz 17, daß die Lösung $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ Lipschitz-stetig vom Parameter $\tilde{\mathbf{t}}$ abhängt, d.h. das reduzierte Funktional $f(\tilde{\mathbf{t}}) := \frac{1}{2}\left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}) \right\|^2$ ist selbst Lipschitz-stetig. Daher nimmt das stetige Funktional $f(\tilde{\mathbf{t}})$ auf der abgeschlossenen $(\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0})$ und beschränkten $(a \leq t_{p(1)}, t_{p(l)} \leq b)$ Menge der zulässigen Knotenfolgen $\left\{\tilde{\mathbf{t}} \in \mathbb{R}^l : \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}\right\}$ sein Minimum an. ∎

6.3 Äquivalenz von Ausgangsproblem und reduziertem Problem

Nachdem wir in Abschnitt 6.1 die Voraussetzungen des Hauptsatzes der Sensitivitätstheorie untersucht haben, kommen wir nun zur Anwendung von Theorem 4 auf unser spezielles Spline-Approximationsproblem.

Theorem 6 (Äquivalenz von Ausgangs- und reduziertem Problem)
Sei $\tilde{\mathbf{t}}^*$ bzw. $\mathbf{t}^*$ eine zulässige Knotenfolge, d.h. $\tilde{\mathbf{t}}^* \in \left\{\tilde{\mathbf{t}} \in \mathbb{R}^l : \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}\right\}$. Die folgenden Voraussetzungen seien erfüllt:

(V1) Für $0 \leq q < k$ mit $q \leq n-1$ gelte

$$t_j^* < t_{j+k-q}^* \qquad (j = q+1, \ldots, n).$$

(V2) Die freien Knoten $\tilde{\mathbf{t}}$ seien einfache Knoten, d.h.

$$\# t_{p(j)}^* = 1 \qquad (j = 1, \ldots, l).$$

Es gelte $k \geq 3$.

(V3) (i) Für das Problem der Spline-Approximation sei die Whitney-Schoenberg-Bedingung an der Stelle $\tilde{\mathbf{t}}^*$ erfüllt.

(ii) Sei $0 \leq r \leq q$. Für das Problem der Spline-Glättung gelte $m \geq r$ und $\mu > 0$.

(V4) Sei $0 \leq p \leq q$. Die strikte Konsistenzbedingung $\mathbf{L} < \mathbf{U}$ für die Nebenbedingung $\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}}^*)\boldsymbol{\alpha} \leq \mathbf{U}$ sei erfüllt.

(V5) Die Lagrange-Parameter $\mathbf{u}^*$ von Subproblem (A) an der Stelle $\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*)$ seien strikt komplementär.

Dann gilt:

(i) Wenn $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ eine globale Minimumstelle des Ausgangsproblems ist, dann erfüllt $\boldsymbol{\alpha}^*$ die notwendigen (und für quadratische definite Probleme auch hinreichende) Optimalitätsbedingungen erster Ordnung für das Subproblem (A), $\tilde{\mathbf{t}}^*$ ist globale Minimumstelle des reduzierten Problems und es gilt

$$f(\tilde{\mathbf{t}}^*) = \mathfrak{f}(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$$
$$\boldsymbol{\alpha}^* = \boldsymbol{\alpha}(\tilde{\mathbf{t}}^*).$$

(ii) Wenn $\tilde{\mathbf{t}}^*$ die notwendigen Optimalitätsbedingungen erster Ordnung für das reduzierte Problem erfüllt, so erfüllt $(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*)$ die notwendigen Optimalitätsbedingungen erster Ordnung für das Ausgangsproblem.

Beweis. Sei $\tilde{\mathbf{t}}^*$ eine zulässige Knotenfolge, d.h. $\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}$.

1.Differenzierbarkeit der Problemfunktionen

Aus den Voraussetzungen (V1) und (V2) folgt die zweimalige stetige Differenzierbarkeit von

$$\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) = \frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2$$

bzw.

$$\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) = \frac{1}{2} \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right\|^2$$

bezüglich $\boldsymbol{\alpha}$ sowie die stetige Differenzierbarkeit des Gradienten

$$\nabla_{\boldsymbol{\alpha}} \mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) = -\mathbf{B}(\tilde{\mathbf{t}})^T \left(\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right)$$

bzw.

$$\nabla_{\boldsymbol{\alpha}} \mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) = - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}^T \left(\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right)$$

bezüglich $\tilde{\mathbf{t}}$ in einer Umgebung $\mathcal{U}_1^*$ von $\tilde{\mathbf{t}}^*$.

2.Differenzierbarkeit der Nebenbedingungen

Die Nebenbedingung $\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}$ ist stetig differenzierbar bzgl. $\tilde{\mathbf{t}}$. Die Nebenbedingung $\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$ ist wegen (V1) und (V2) stetig differenzierbar bzgl. $\boldsymbol{\alpha}$ und $\tilde{\mathbf{t}}$ in einer Umgebung $\mathcal{U}_2^*$ von $\tilde{\mathbf{t}}^*$.

3.Voraussetzungen an Subproblem (A)

(a) Nach Voraussetzung (V3i) hat die Beobachtungsmatrix $\mathbf{B}(\tilde{\mathbf{t}}^*)$ im Falle der Spline-Approximation Vollrang n. Dann existiert eine Umgebung $\mathcal{U}_3^*$ von $\tilde{\mathbf{t}}^*$ so, daß $\mathbf{B}(\tilde{\mathbf{t}})$ Vollrang für alle $\tilde{\mathbf{t}} \in \mathcal{U}_3^*$ hat.

Im Falle der Spline-Glättung hat die Beobachtungsmatrix $\begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}$ wegen (V3ii) Vollrang für alle zulässigen Knotenfolgen $\tilde{\mathbf{t}}$.

Das quadratische Optimierungsproblem Subproblem (A) hat daher eine positiv definite Hesse-Matrix und besitzt folglich über dem nichtleeren zulässigen Bereich $\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$ eine eindeutig bestimmte Lösung $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$. Es gelten die hinreichenden Optimalitätsbedingungen zweiter Ordnung an der Stelle $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$.

(b) Aus Voraussetzung (V4) folgt nach Satz 16 die lineare Unabhängigkeit der Gradienten der aktiven Restriktionen von Subproblem (A).

(c) Die eindeutig bestimmten Lagrange-Parameter $\mathbf{u}^*$ von Subproblem (A) an der Stelle $\tilde{\mathbf{t}}^*$ sind nach Voraussetzung (V5) strikt komplementär. Dann existiert eine Umgebung $\mathcal{U}_4^*$ von $\tilde{\mathbf{t}}^*$ so, daß die eindeutig bestimmten Lagrange-Parameter $\mathbf{u}$ von Subproblem (A) an der Stelle $\tilde{\mathbf{t}} \in \mathcal{U}_4^*$ (bzw. $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$) ebenfalls strikt komplementär sind.

Im nichtleeren Durchschnitt der Umgebungen $\mathcal{U}_1^*, \ldots, \mathcal{U}_4^*$ sind somit alle Voraussetzungen von Theorem 4 erfüllt. Die Aussagen ergeben sich durch unmittelbare Anwendung des Theorems. Für den Teil (i) der Behauptung gilt zudem:

Da das Subproblem (A) ein quadratisches, definites Optimierungsproblem ist, existiert ein eindeutiges $\boldsymbol{\alpha}^*$ unter allen Paaren $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$, welche $\mathfrak{f}$ minimieren und denselben Minimalwert ergeben, und es gilt $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}(\tilde{\mathbf{t}}^*)$. ■

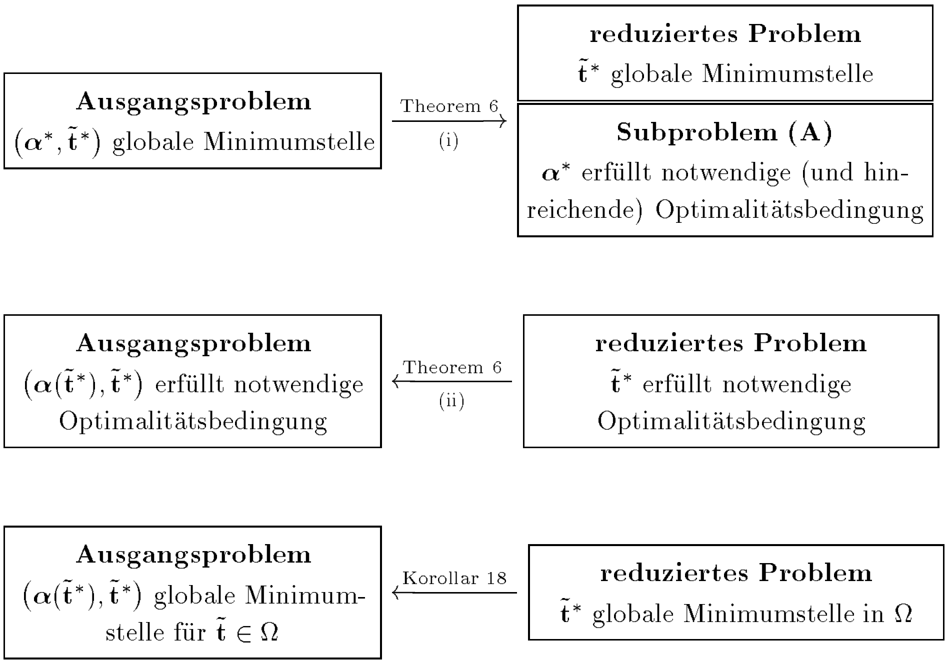


Abbildung 1: Beziehung zwischen Ausgangsproblem und reduziertem Problem

Korollar 18
Die Voraussetzungen von Theorem 6 seien in einer Umgebung Ω der zulässigen Knotenfolge $\tilde{\mathbf{t}}^*$ erfüllt. Wenn $\tilde{\mathbf{t}}^*$ eine globale Minimumstelle des reduzierten Funktionals $f(\tilde{\mathbf{t}})$ in Ω ist, so ist $(\boldsymbol{\alpha}^*, \tilde{\mathbf{t}}^*)$ globale Minimumstelle von $\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$ für $\tilde{\mathbf{t}} \in \Omega$.

Beweis. Sei $\tilde{\mathbf{t}}^*$ eine globale Minimumstelle von $f(\tilde{\mathbf{t}})$ in Ω und sei $\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*)$ die zugehörige eindeutige Lösung von Subproblem (A). Klar ist, daß

$$\mathfrak{f}\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right) = f(\tilde{\mathbf{t}}^*).$$

Annahme: Es existieren eine Knotenfolge $\tilde{\mathbf{t}}^\dagger \in \Omega$ und Koeffizienten $\boldsymbol{\alpha}^\dagger$ mit

$$\mathfrak{f}\left(\boldsymbol{\alpha}^\dagger, \tilde{\mathbf{t}}^\dagger\right) < \mathfrak{f}\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right).$$

Nach Definition des reduzierten Problems gilt aber

$$\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) \geq f(\tilde{\mathbf{t}}) \quad \forall \tilde{\mathbf{t}} \in \Omega,$$

denn $\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2$ mit „beliebigem" $\boldsymbol{\alpha}$ und $f(\tilde{\mathbf{t}}) := \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})\right\|^2$ mit „optimalem" $\boldsymbol{\alpha}$ (analog für Spline-Glättung).
Es folgt

$$f(\tilde{\mathbf{t}}^\dagger) \leq \mathfrak{f}\left(\boldsymbol{\alpha}^\dagger, \tilde{\mathbf{t}}^\dagger\right) < \mathfrak{f}\left(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*\right) = f(\tilde{\mathbf{t}}^*).$$

Dies ist ein Widerspruch zur Voraussetzung, daß $\tilde{\mathbf{t}}^*$ globale Minimumstelle von $f(\tilde{\mathbf{t}})$ in Ω ist. Also ist $(\boldsymbol{\alpha}(\tilde{\mathbf{t}}^*), \tilde{\mathbf{t}}^*)$ globale Minimumstelle von $\mathfrak{f}(\boldsymbol{\alpha}, \tilde{\mathbf{t}})$ für $\tilde{\mathbf{t}} \in \Omega$. ∎

Abbildung 1 verdeutlicht die Beziehungen zwischen Ausgangsproblem und reduziertem Problem.

Bemerkung 19
Wir betrachten das Problem der Spline-Glättung mit $m \geq r$ und $\mu > 0$. Dann besitzt die Beobachtungsmatrix Vollrang für alle Knotenfolgen $\tilde{\mathbf{t}}$ aus dem zulässigen Bereich $\{\tilde{\mathbf{t}} \in \mathbb{R}^l : \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq 0\}$. Wenn die Schranken $\mathbf{L}$ und $\mathbf{U}$ an die Ableitungen, wie in unserem Fall, nicht von den Knoten $\tilde{\mathbf{t}}$ abhängen, so lassen sich die Vorausetzungen (V1), (V2), (V3ii) und (V4) global, d.h. für jedes $\tilde{\mathbf{t}}$ aus $\{\tilde{\mathbf{t}} \in \mathbb{R}^l : \mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq 0\}$ erfüllen. Falls die Lagrange-Parameter $\mathbf{u}$ von Subproblem (A) an der Stelle $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ etwa für alle $\tilde{\mathbf{t}}$ aus der Niveaumenge $W(\tilde{\mathbf{t}}^0) := \{\tilde{\mathbf{t}} \in \mathbb{R}^l : f(\tilde{\mathbf{t}}) \leq f(\tilde{\mathbf{t}}^0)\}$ strikt komplementär sind, so gilt die Äquivalenz zwischen Ausgangsproblem und reduziertem Problem global für alle $\tilde{\mathbf{t}} \in W(\tilde{\mathbf{t}}^0)$. Die reduzierte Funktion ist dann glatt für alle $\tilde{\mathbf{t}} \in W(\tilde{\mathbf{t}}^0)$. Die Voraussetzung (V5) ist somit die einzige Bedingung, welche eine globale Charakterisierung der Äquivalenz von Ausgangsproblem und reduziertem Problem verhindert.

7 Numerische Lösung des reduzierten Problems im Fall der Spline-Approximation

Nachdem wir die Äquivalenz von Ausgangsproblem und reduziertem Problem im Sinne von Theorem 6 gezeigt haben, widmen wir uns nun der effizienten numerischen Lösung des reduzierten Problems. Wir betrachten dabei zunächst den Fall der Spline-Approximation.

Die Lösung des reduzierten Problems hat einige Vorteile gegenüber der Lösung des Ausgangsproblems:

- Die Dimension des Problems ist l gegenüber $n+l$.
- Die Nebenbedingungen des reduzierten Problems sind linear gegenüber den nichtlinearen Nebenbedingungen ($\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$) des Ausgangsproblems.
- Existierende Software zur Lösung des Subproblems (A), siehe z.B. [SK93], kann unmittelbar eingesetzt werden.
- Die Bandstruktur von $\mathbf{B}(\tilde{\mathbf{t}})$ und $\boldsymbol{\partial}\mathbf{B}(\tilde{\mathbf{t}})$ kann voll ausgenutzt werden.

Dem stehen die folgenden Nachteile gegenüber:

- Die Struktur von Gradient, Hesse-Matrix und Jacobi-Matrix des reduzierten Funktionals ist relativ kompliziert.
- Die Koeffizienten $\boldsymbol{\alpha}$ sind bei nichtstrikter Komplementarität nicht stetig differenzierbar bzgl. der Splineknoten $\mathbf{t}$.

Das reduzierte Problem ist ein nichtlineares Quadratmittelproblem mit linearen Ungleichheitsnebenbedingungen. Dieses Problem wollen wir mit einem verallgemeinerten Gauß-Newton Verfahren lösen. Dazu benötigen wir die Jacobi-Matrix des reduzierten Funktionals oder eine Approximation davon. Nach Satz 10 gilt

$$\begin{aligned}
\mathbf{J} &= \mathbf{P}^{\perp}\boldsymbol{\psi} + \mathbf{P}\boldsymbol{\phi} \\
&= \mathbf{P}_{BN}^{\perp}\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^{+}\bar{\boldsymbol{\Gamma}}\right) + \mathbf{P}_{BN}\left((\mathbf{BN})^{+}\right)^{T}\mathbf{N}^{T}\mathbf{K}
\end{aligned}$$

Im folgenden Abschnitt werden wir, analog zum Kaufman-Modell für unrestringierte separable Quadratmittelprobleme, mit der Kaufman-Approximation

$$\begin{aligned}
\mathbf{J}_K &= \mathbf{P}^{\perp}\boldsymbol{\psi} \\
&= \mathbf{P}_{BN}^{\perp}\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^{+}\bar{\boldsymbol{\Gamma}}\right)
\end{aligned}$$

der Jacobi-Matrix arbeiten, d.h. wir vernachlässigen den Term $\mathbf{P}\boldsymbol{\phi}$. Man beachte, daß diese Approximation keinen Einfluß auf den Gradienten hat.

7.1 Die Kaufman-Jacobi-Matrix im Fall der Spline-Approximation

Die Kaufman-Approximation $\mathbf{J}_K \in \mathbb{R}^{m,l}$ der Jacobi-Matrix $\mathbf{J}$ des reduzierten Funktionals ist definiert durch

$$\begin{aligned}
\mathbf{J}_K &= \mathbf{P}^{\perp}\boldsymbol{\psi} \\
&= (\mathbf{I}_m - \mathbf{P}_{BN})\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^{+}\bar{\boldsymbol{\Gamma}}\right) \\
&= \left[\mathbf{I}_m - (\mathbf{BN})(\mathbf{BN})^{+}\right]\left(-\boldsymbol{\partial}\mathbf{B}\boldsymbol{\alpha} + \mathbf{B}\begin{bmatrix} \mathbf{D}_p \\ -\mathbf{D}_p \end{bmatrix}^{+}_{i\in\mathcal{I}} \boldsymbol{\partial}\begin{bmatrix} \mathbf{D}_p \\ -\mathbf{D}_p \end{bmatrix}_{i\in\mathcal{I}} \boldsymbol{\alpha}\right)
\end{aligned}$$

Implementiert man diese Formeln naiv, indem man z.B. die Matrix **BN** explizit bildet und orthogonal faktorisiert, so wird die charakteristische Bandstruktur der beteiligten Matrizen zerstört und der Aufwand ist besonders im praktisch relevanten Fall $m \gg n$ nicht hinnehmbar. Wir streben daher nach einer Implementation, welche solche Faktorisierungen vermeidet und die Schwachbesetztkeit soweit als möglich ausnutzt.

7.2 Berechnung der Matrizen, in denen Ableitungen auftreten

Widmen wir uns zunächst der Berechnung derjenigen Matrizen, in denen Ableitungen nach $\tilde{\mathbf{t}}$ auftreten.

7.2.1 Ableitung der Beobachtungsmatrix nach den Knoten

Es gilt $\mathbf{J}_{\tilde{\mathbf{t}}} = -\boldsymbol{\partial}\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$, wobei $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ das Subproblem (A) an der Stelle $\tilde{\mathbf{t}}$ löst. Wir berechnen $\mathbf{J}_{\tilde{\mathbf{t}}}$ spaltenweise und erhalten $\mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^{\tau} = -\boldsymbol{\partial}\mathbf{B}(\tilde{\mathbf{t}})\,[\mathbf{e}^{\tau}]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ $(\tau = 1, \ldots, l)$. Wegen

$$\boldsymbol{\partial}\mathbf{B}(\tilde{\mathbf{t}})\,[\mathbf{e}^{\tau}]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}}) = \left(\sum_{j=1}^{n} \frac{\partial B_{j,k}\left(x_i, \tilde{\mathbf{t}}\right)}{\partial t_{p(\tau)}} \alpha_j(\tilde{\mathbf{t}})\right)_{i=1,\ldots,m}$$

gilt

$$\mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^{\tau} = -\left(\frac{\partial s(x_i)}{\partial t_{p(\tau)}}\right)_{i=1,\ldots,m}$$

Wenn man nun die Ableitungen eines Splines nach den Knoten zur Verfügung stellt, siehe die entprechenden Formeln in Lemma 3, so kann man $\mathbf{J}_{\tilde{\mathbf{t}}}$ spaltenweise exakt berechnen.

Algorithmus 3 (Berechnung von $\mathbf{J}_{\tilde{\mathbf{t}}} = -\boldsymbol{\partial}\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}}) \in \mathbb{R}^{m,l}$)
for $\tau := 1$ *to* l *do*

    begin

        for $j := 1$ *to* $p(\tau)$ *do*

            $t'_j := t_j$;

        for $j := p(\tau) + 1$ *to* $n + k + 1$ *do*

            $t'_j := t_{j-1}$;

        for $i := 1$ *to* m *do*

            begin

                $F_i := 0$;

                for $j := p(\tau) - k + 1$ *to* $p(\tau)$ *do*

                    $F_i := F_i - \frac{\alpha_{j-1} - \alpha_j}{t'_{j+k} - t'_j} \times B_{j,k}(x_i, \mathbf{t}')$;

            end;

        $\mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^{\tau} := \mathbf{F}$;

    end

7.2.2 Ableitung der Nebenbedingungen nach den Knoten

Für die Knotenfolge $\mathbf{t}$ gelte $t_j < t_{j+k-q}$ $(j = q+1, \ldots, n)$ mit $q \in \mathbb{Z}$, $0 \le q < k$ und $q < n-1$. Für $p \in \{0, \ldots, q\}$ seien die Matrizen $\mathbf{H}_\nu$, $\mathbf{L}_\nu$ $(\nu = 1, \ldots, p)$ und $\mathbf{D}_p$ wie in Abschnitt 2.3 definiert.

$$\mathbf{H}_\nu := (k-\nu)\,\mathrm{diag}\left(\frac{1}{t_{k+j} - t_{\nu+j}}\right)_{j=1,\ldots,n-\nu} \in \mathbb{R}^{n-\nu,n-\nu}$$

$$\mathbf{L}_\nu := \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n-\nu,n-\nu+1} \quad \text{für} \quad \nu = 1, \ldots, p$$

$$\mathbf{D}_0 := \mathbf{I} \in \mathbb{R}^{n,n}$$

$$\mathbf{D}_p := \mathbf{H}_p \mathbf{L}_p \ldots \mathbf{H}_1 \mathbf{L}_1 \in \mathbb{R}^{n-p,n} \quad \text{für} \quad p \ge 1.$$

Wir wollen nun die Fréchet-Ableitung der Matrixfunktion $\mathbf{D}_p(.)$ bzgl. der freien Knoten $\tilde{\mathbf{t}}$ bestimmen.
Offensichtlich gilt

$$\begin{aligned} \mathbf{H}_\nu &= \mathbf{H}_\nu(\tilde{\mathbf{t}}) \\ \boldsymbol{\partial}\mathbf{L}_\nu &= \mathbf{0} \qquad\qquad \nu = 1, \ldots, p. \end{aligned}$$

Für $\boldsymbol{\partial}\mathbf{D}_p \in \mathfrak{L}\left(\mathbb{R}^l, \mathfrak{L}\left(\mathbb{R}^n, \mathbb{R}^{n-p}\right)\right)$ hat man

(i) $$\begin{aligned} \boldsymbol{\partial}\mathbf{D}_0 &= \boldsymbol{\partial}\mathbf{I} \\ &= \mathbf{0} \end{aligned}$$

(ii) $$\begin{aligned} \boldsymbol{\partial}\mathbf{D}_1 &= \boldsymbol{\partial}\left(\mathbf{H}_1\mathbf{L}_1\right) \\ &= \left(\boldsymbol{\partial}\mathbf{H}_1\right)\mathbf{L}_1 + \mathbf{H}_1(\boldsymbol{\partial}\mathbf{L}_1) \\ &= \left(\boldsymbol{\partial}\mathbf{H}_1\right)\mathbf{L}_1 \end{aligned}$$

(iii) $$\begin{aligned} \boldsymbol{\partial}\mathbf{D}_2 &= \boldsymbol{\partial}\left(\mathbf{H}_2\mathbf{L}_2\mathbf{D}_1\right) \\ &= \left(\boldsymbol{\partial}\mathbf{H}_2\right)\mathbf{L}_2\mathbf{D}_1 + \mathbf{H}_2\left(\boldsymbol{\partial}\mathbf{L}_2\right)\mathbf{D}_1 + \mathbf{H}_2\mathbf{L}_2\left(\boldsymbol{\partial}\mathbf{D}_1\right) \\ &= \left(\boldsymbol{\partial}\mathbf{H}_2\right)\mathbf{L}_2\mathbf{D}_1 + \mathbf{H}_2\mathbf{L}_2\left(\boldsymbol{\partial}\mathbf{D}_1\right) \end{aligned}$$

Allgemein zeigt man leicht durch vollständige Induktion

$$\boldsymbol{\partial}\mathbf{D}_p = \begin{cases} \mathbf{0} & \text{für } p = 0, \\ \left(\boldsymbol{\partial}\mathbf{H}_p\right)\mathbf{L}_p\mathbf{D}_{p-1} + \mathbf{H}_p\mathbf{L}_p\left(\boldsymbol{\partial}\mathbf{D}_{p-1}\right) & \text{für } p \ge 1. \end{cases} \tag{7.1}$$

Damit erhalten wir

$$\boldsymbol{\partial}\mathbf{D}_p[\mathbf{e}^\tau]\mathbf{c}^n = \left(\boldsymbol{\partial}\mathbf{H}_p\right)[\mathbf{e}^\tau]\mathbf{L}_p\mathbf{D}_{p-1}\mathbf{c}^n + \mathbf{H}_p\mathbf{L}_p\left(\boldsymbol{\partial}\mathbf{D}_{p-1}\right)[\mathbf{e}^\tau]\mathbf{c}^n \in \mathbb{R}^{n-p}$$

mit dem Einheitsvektor $\mathbf{e}^\tau \in \mathbb{R}^l$, $\tau = 1, \ldots, l$ sowie einem beliebigen Vektor $\mathbf{c}^n \in \mathbb{R}^n$.
Wir betrachten nun $\boldsymbol{\partial}\mathbf{H}_\nu[\mathbf{e}^\tau]$ bzw. $\boldsymbol{\partial}\mathbf{H}_\nu[\mathbf{e}^\tau]\mathbf{c}^{n-\nu}$ für $\nu = 1, \ldots, p$; $\tau = 1, \ldots, l$ und $\mathbf{c}^{n-\nu} \in \mathbb{R}^{n-\nu}$. Es gilt

$$\boldsymbol{\partial}\mathbf{H}_\nu[\mathbf{e}^\tau]\mathbf{c}^{n-\nu} \in \mathfrak{L}\left(\mathbb{R}^{n-\nu}\right)$$

bzw. in Komponentenschreibweise

$$\boldsymbol{\partial}\mathbf{H}_\nu[\mathbf{e}^\tau]\mathbf{c}^{n-\nu} = \left(\sum_{j=1}^{n-\nu} \frac{\partial H_{ij}(\tilde{\mathbf{t}})}{\partial t_{p(\tau)}} c_j\right)_{i=1,\ldots,n-\nu.}$$

Laut Definition von $(\mathbf{H}_\nu)_{ij} = H_{ij}$ gilt

$$H_{ij}(\tilde{\mathbf{t}}) = \begin{cases} 0 & \text{falls } i \neq j, \\ \frac{k-\nu}{t_{k+j}-t_{\nu+j}} & \text{falls } i = j; i,j = 1,\ldots,n-\nu, \end{cases}$$

also

$$\frac{\partial H_{ij}(\tilde{\mathbf{t}})}{\partial t_{p(\tau)}} = 0 \qquad \text{falls } i \neq j,$$

$$\frac{\partial H_{jj}(\tilde{\mathbf{t}})}{\partial t_{p(\tau)}} = \begin{cases} -\frac{k-\nu}{(t_{k+j}-t_{\nu+j})^2} & \text{falls } p(\tau) = k+j, \\ \frac{k-\nu}{(t_{k+j}-t_{\nu+j})^2} & \text{falls } p(\tau) = \nu+j, \\ 0 & \text{sonst.} \end{cases}$$

Damit erhalten wir folgenden Algorithmus zur Berechnung von $\boldsymbol{\partial}\mathbf{H}_\nu[\mathrm{e}^\tau]\mathbf{c}^{n-\nu}$:

Algorithmus 4 (Berechnung von $\boldsymbol{\partial}\mathbf{H}_\nu[\mathrm{e}^\tau]\mathbf{c}^{n-\nu} \in \mathbb{R}^{n-\nu}$)
Sei $\nu \in \{1,\ldots,p\}$, $\tau \in \{1,\ldots,l\}$ und $\mathbf{c}^{n-\nu} = (c_1,\ldots,c_{n-\nu})^T \in \mathbb{R}^{n-\nu}$. Der Algorithmus überschreibt $\mathbf{c}^{n-\nu}$ mit $\boldsymbol{\partial}\mathbf{H}_\nu[\mathrm{e}^\tau]\mathbf{c}^{n-\nu}$.

for $j := 1$ *to* $n-\nu$ *do*

begin

if $j = p(\tau) - k$ *then*

$c_j := -\frac{k-\nu}{(t_{k+j}-t_{\nu+j})^2} \times c_j$

else

if $j = p(\tau) - \nu$ *then*

$c_j := \frac{k-\nu}{(t_{k+j}-t_{\nu+j})^2} \times c_j$

else

$c_j := 0;$

end;

Nun sind wir in der Lage, einen Algorithmus zur Berechnung der Ableitung der Nebenbedingungen nach den Knoten anzugeben:

Algorithmus 5 (Berechnung von $\mathbf{v} := \boldsymbol{\partial}\mathbf{D}_p[\mathrm{e}^\tau]\boldsymbol{\alpha} \in \mathbb{R}^{n-p}$)

$\mathbf{v} := \mathbf{0};$ $\{(\boldsymbol{\partial}\mathbf{D}_0)[\mathrm{e}^\tau]\boldsymbol{\alpha} = \mathbf{0}\}$

for $\nu := 1$ *to* p *do*

begin

$\mathbf{v}^1 := \mathbf{L}_\nu\mathbf{D}_{\nu-1}\boldsymbol{\alpha};$ $\{\mathbf{v}^1 = \mathbf{L}_\nu\mathbf{D}_{\nu-1}\boldsymbol{\alpha} \in \mathbb{R}^{n-\nu}\}$

$\mathbf{v}^1 := (\boldsymbol{\partial}\mathbf{H}_\nu)[\mathrm{e}^\tau]\mathbf{v}^1;$ $\{\mathbf{v}^1 = (\boldsymbol{\partial}\mathbf{H}_\nu)[\mathrm{e}^\tau]\mathbf{L}_\nu\mathbf{D}_{\nu-1}\boldsymbol{\alpha} \in \mathbb{R}^{n-\nu}\}$

$\mathbf{v}^2 := \mathbf{v};$ $\{\mathbf{v}^2 = (\boldsymbol{\partial}\mathbf{D}_{\nu-1})[\mathrm{e}^\tau]\boldsymbol{\alpha} \in \mathbb{R}^{n-\nu+1}\}$

$\mathbf{v}^2 := \mathbf{H}_\nu\mathbf{L}_\nu\mathbf{v}^2;$ $\{\mathbf{v}^2 = \mathbf{H}_\nu\mathbf{L}_\nu(\boldsymbol{\partial}\mathbf{D}_{\nu-1})[\mathrm{e}^\tau]\boldsymbol{\alpha} \in \mathbb{R}^{n-\nu}\}$

$\mathbf{v} := \mathbf{v}^1 + \mathbf{v}^2;$ $\{\mathbf{v} = (\boldsymbol{\partial}\mathbf{D}_\nu)[\mathrm{e}^\tau]\boldsymbol{\alpha} \in \mathbb{R}^{n-\nu}\}$

end;

Algorithmus 5 ist die Grundlage zur Berechnung der gesuchten Matrix $\bar{\mathbf{\Gamma}}$. Sie wird ebenfalls spaltenweise berechnet. Es gilt $\bar{\mathbf{\Gamma}}\mathrm{e}^\tau = -\boldsymbol{\partial}\bar{\mathbf{R}}(\tilde{\mathbf{t}})[\mathrm{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}})$, wobei $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ das

Subproblem (A) an der Stelle $\tilde{\mathbf{t}}$ löst.

Algorithmus 6 (Berechnung von $\bar{\mathbf{\Gamma}} = -\boldsymbol{\partial}\bar{\mathbf{R}}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}}) \in \mathbb{R}^{nactiv,l}$)
for $\tau := 1$ *to* l *do*

begin

$\mathbf{v} := \boldsymbol{\partial}\mathbf{D}_p[\mathbf{e}^\tau]\boldsymbol{\alpha}$;

jj:=1;

for $j := 1$ *to* $n - p$ *do*

begin

if $L_j = (\mathbf{D}_p\boldsymbol{\alpha})_j$ *then*

$\left(\bar{\mathbf{\Gamma}}\mathbf{e}^\tau\right)_{jj} = -v_j$;

$jj := jj + 1$;

end;

for $j := 1$ *to* $n - p$ *do*

begin

if $U_j = (\mathbf{D}_p\boldsymbol{\alpha})_j$ *then*

$\left(\bar{\mathbf{\Gamma}}\mathbf{e}^\tau\right)_{jj} = v_j$;

$jj := jj + 1$;

end;

end

Zusammenfassend kann man feststellen, daß die Matrizen, in denen Ableitungen nach $\tilde{\mathbf{t}}$ vorkommen, effizient berechnet werden können, da die Matrizen $\mathbf{B}(\tilde{\mathbf{t}})$ und $\bar{\mathbf{R}}(\tilde{\mathbf{t}})$ Bandstruktur haben. Alle nichtverschwindenen B-Splines an einem Punkt x_i können simultan berechnet werden. Eine analoge Aussage gilt für die Zeilen von $\bar{\mathbf{R}}(\tilde{\mathbf{t}})$ bzw. $\mathbf{D}_p(\tilde{\mathbf{t}})$.

7.3 Berechnung einer Nullraumbasis N bzw. von Vektoren $\bar{\mathbf{R}}^+\mathbf{y}$

Zur Berechnung des orthogonalen Projektors $\mathbf{P}_{BN}$ wird eine Basis $\mathbf{N}$ des Nullraumes von $\bar{\mathbf{R}}$ benötigt. Eng damit verbunden ist die Berechnung der Moore-Penrose-Inversen $\bar{\mathbf{R}}^+$ bzw. die Berechnung von Vektoren der Form $\bar{\mathbf{R}}^+\mathbf{y}$ mit $\mathbf{y} \in \mathbb{R}^{nactiv}$.

Sei $\bar{\mathbf{R}} \in \mathbb{R}^{nactiv,n}$ die Matrix mit den Zeilen von $\mathbf{R}$, für welche die Nebenbedingungen aktiv sind. Es gilt $0 \leq nactiv \leq n - p$ und die Zeilen von $\bar{\mathbf{R}}$ sind linear unabhängig, d.h. rank $\bar{\mathbf{R}} = nactiv$. Die Matrix $\bar{\mathbf{R}}$ hat eine Zeilenbandbreite von $p + 1$.

Beispiel 8
$n = 10$, $p = 2$, $nactiv = 4$, $\mathcal{I} = \{4, 6, 10, 15\}$

$$\bar{\mathbf{R}} = \begin{bmatrix} & & & x & x & x & & & & \\ & & & & & x & x & x & & \\ & x & x & x & & & & & & \\ & & & & & & x & x & x & \end{bmatrix} \in \mathbb{R}^{10,4}$$

Wegen rank $\bar{\mathbf{R}} = nactiv$ gilt $\bar{\mathbf{R}}^+ = \bar{\mathbf{R}}^T\left(\bar{\mathbf{R}}\bar{\mathbf{R}}^T\right)^{-1}$. Die Matrix $\bar{\mathbf{R}}\bar{\mathbf{R}}^T \in \mathbb{R}^{nactiv,nactiv}$ ist symmetrisch und positiv definit.

Im weiteren benötigen wir für verschiedene rechte Seiten $\mathbf{y} \in \mathbb{R}^{nactiv}$ die Vektoren $\mathbf{x} = \bar{\mathbf{R}}^{+}\mathbf{y} \in \mathbb{R}^{n}$. Genauer: $\mathbf{x}_\tau = \bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\mathbf{e}^\tau$ $(\tau = 1,\ldots,l)$. Der Vektor $\mathbf{x}$ kann als Lösung des *Minimum-Norm-Problems*

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimiere}} \quad \|\mathbf{x}\|_2 \qquad \text{bei} \qquad \bar{\mathbf{R}}\mathbf{x} = \mathbf{y}$$

interpretiert werden. Ein naheliegende Möglichkeit zur Berechnung von $\mathbf{x}$ ist die Methode der Normalgleichungen 2.Art.

Algorithmus 7 (Berechnung von $\mathbf{x} = \bar{\mathbf{R}}^{+}\mathbf{y}$, Normalgleichungen)
S1: Bilde $\bar{\mathbf{R}}\bar{\mathbf{R}}^T \in \mathbb{R}^{nactiv,nactiv}$

S2: Bilde Cholesky-Faktorisierung

$$\bar{\mathbf{R}}\bar{\mathbf{R}}^T = \mathbf{L}\mathbf{L}^T$$

mit regulärer unterer Dreiecksmatrix $\mathbf{L} \in \mathbb{R}^{nactiv,nactiv}$

S3: Für jede rechte Seite $\mathbf{y} \in \mathbb{R}^{nactiv}$

S3.1: Löse $\mathbf{L}\mathbf{L}^T\mathbf{w} = \mathbf{y}$; $\mathbf{w} \in \mathbb{R}^{nactiv}$
S3.2: Bilde $\mathbf{x} := \bar{\mathbf{R}}^T\mathbf{w}$; $\mathbf{x} \in \mathbb{R}^{n}$

Die Fehlerschranke der berechneten Lösung $\mathbf{x}$ hängt vom Quadrat der Konditionszahl von $\bar{\mathbf{R}}$ ab.

Ein Verfahren mit besseren Stabilitätseigenschaften erhält man mittels einer QR-Faktorisierung von $\bar{\mathbf{R}}^T$ (oder einer LQ-Faktorisierung von $\bar{\mathbf{R}}$). Wegen $\text{null}\left(\bar{\mathbf{R}}\right) = \text{range}\left(\bar{\mathbf{R}}^T\right)^\perp$ kann damit gleichzeitig eine orthonormale Basis $\mathbf{N}$ des Nullraumes von $\bar{\mathbf{R}}$ berechnet werden. Da die Zeilen von $\bar{\mathbf{R}} \in \mathbb{R}^{nactiv,n}$ linear unabhängig sind, ist $\bar{\mathbf{R}}^T$ spaltenregulär.

QR-Faktorisierung von $\bar{\mathbf{R}}^T$

Sei $\mathbf{Q}_1 = (\mathbf{Q}_{11}\,|\,\mathbf{Q}_{12}) \in \mathbb{R}^{n,n}$ mit $\mathbf{Q}_{11} \in \mathbb{R}^{n,nactiv}$ und $\mathbf{Q}_{12} \in \mathbb{R}^{n,n-nactiv}$ eine orthogonale Matrix mit

$$\mathbf{Q}_1^T\bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{Q}_{11}^T \\ \mathbf{Q}_{12}^T \end{bmatrix} \bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix},$$

wobei $\mathbf{R}_1 \in \mathbb{R}^{nactiv,nactiv}$ eine reguläre obere Dreiecksmatrix ist.

Dann gilt $\text{range}\left(\bar{\mathbf{R}}^T\right)^\perp = \text{range}\left(\mathbf{Q}_{12}\right)$, d.h. $\mathbf{Q}_{12} \in \mathbb{R}^{n,n-nactiv}$ ist die gesuchte Nullraummatrix $\mathbf{N}$.

Die Matrix $\bar{\mathbf{R}}^T$ enthält in jeder Spalte $p+1$ Nichtnullelemente, allerdings (abhängig von der Indexmenge $\mathcal{I}$) in ungeordneter Weise, siehe Beispiel 8. Wir verwenden einen Standardalgorithmus mittels Householder-Transformationen zur Faktorisierung von $\bar{\mathbf{R}}^T$.

Die Lösung des Minimum-Norm-Problems erhält man durch folgenden Algorithmus

Algorithmus 8 (Berechnung von $\mathbf{x} = \bar{\mathbf{R}}^{+}\mathbf{y}$ und $\mathbf{N}$, Golub's Methode)
S1: Bilde QR-Faktorisierung von $\bar{\mathbf{R}}^T \in \mathbb{R}^{n,nactiv}$

$$\begin{bmatrix} \mathbf{Q}_{11}^T \\ \mathbf{Q}_{12}^T \end{bmatrix} \bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$$

mittels Householder-Transformationen

S2: Basis $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ *des Nullraumes von* $\bar{\mathbf{R}}$

$$\mathbf{N} := \mathbf{Q}_{12}$$

S3: Für jede rechte Seite $\mathbf{y} \in \mathbb{R}^{nactiv}$

S3.1: Löse $\mathbf{R}_1^T\mathbf{w} = \mathbf{y}$; $\mathbf{R}_1 \in \mathbb{R}^{nactiv,nactiv}$ *reguläre obere Dreiecksmatrix;* $\mathbf{w} \in \mathbb{R}^{nactiv}$

S3.2: Bilde $\mathbf{x} := \mathbf{Q}_{11}^T\mathbf{w}$; $\mathbf{x} \in \mathbb{R}^n$

Der Nachteil dieser Methode für große Dimensionen ist, daß der i.allg. vollbesetzte Faktor $\mathbf{Q}_{11} \in \mathbb{R}^{n,nactiv}$ für jede rechte Seite benötigt wird.

Die Matrix $\mathbf{L} = \mathbf{R}_1^T$ ist (bis auf Vorzeichen) der Dreiecksfaktor der Cholesky-Faktorisierung der Matrix $\bar{\mathbf{R}}\bar{\mathbf{R}}^T$. Wir können deshalb $\mathbf{x} = \bar{\mathbf{R}}^+\mathbf{y} = \bar{\mathbf{R}}^T\left(\mathbf{R}_1^T\mathbf{R}_1\right)^{-1}\mathbf{y}$ über die Methode der *Semi-Normalgleichungen* (siehe [Sau72]) berechnen.

Algorithmus 9 (Berechnung von $\mathbf{x} = \bar{\mathbf{R}}^+\mathbf{y}$ und $\mathbf{N}$, Semi-Normalgleichungen)
S1: Bilde QR-Faktorisierung von $\bar{\mathbf{R}}^T \in \mathbb{R}^{n,nactiv}$

$$\begin{bmatrix} \mathbf{Q}_{11}^T \\ \mathbf{Q}_{12}^T \end{bmatrix} \bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$$

mittels Householder-Transformationen

Speichere den regulären oberen Dreiecksfaktor $\mathbf{R}_1 \in \mathbb{R}^{nactiv,nactiv}$ *ab*

S2: Basis $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ *des Nullraumes von* $\bar{\mathbf{R}}$

$$\mathbf{N} := \mathbf{Q}_{12}$$

S3: Für jede rechte Seite $\mathbf{y} \in \mathbb{R}^{nactiv}$

S3.1: Löse $\mathbf{R}_1^T\mathbf{R}_1\mathbf{w} = \mathbf{y}$; $\mathbf{w} \in \mathbb{R}^{nactiv}$

S3.2: Bilde $\mathbf{x} := \bar{\mathbf{R}}^T\mathbf{w}$; $\mathbf{x} \in \mathbb{R}^n$

Bemerkung 20

(i) Im Gegensatz zu Golub's Methode muß bei den Semi-Normalgleichungen zusätzlich zu den ohnehin benötigten Größen $\mathbf{N}$ *und* $\mathbf{R}_1$ *nur die schwach besetzte Originalmatrix* $\bar{\mathbf{R}}^T$ *an Stelle von* $\mathbf{Q}_{11}$ *aufgehoben werden.*

(ii) Die Householder-Transformationen werden in der Matrix $\mathbf{Q}_1$ *in Produktform gespeichert. Die Matrix* $\mathbf{N}$ *kann durch* $(n - nactiv)$*-malige Anwendung von*

$$\mathbf{N}\mathbf{e}^i = \mathbf{Q}_1\tilde{\mathbf{e}}^i \qquad mit \qquad \tilde{\mathbf{e}}^i = (\underbrace{0,\ldots,0}_{n-nactiv}, \mathbf{e}^{i^T})^T \quad (i = 1,\ldots,n-nactiv)$$

spaltenweise erzeugt werden.

(iii) *Für lineare Quadratmittelproblem wird in [Bjö91] gezeigt, daß der Fehler der berechneten Lösung bei der Methode der Semi-Normalgleichungen ebenfalls vom Quadrat der Konditionszahl abhängt, obwohl der berechnete Dreicksfaktor von „besserer" Qualität als der Dreiecksfaktor der Normalgleichungen ist. Bei der von Saunders [Sau72] vorgeschlagenen Methode der Semi-Normalgleichungen für das Minimum-Norm-Problem tritt dieser Effekt nicht auf. In [Pai73] wurde gezeigt, daß „the bound on the error in* $\mathbf{x}$ *is proportional to* κu *rather than* $\kappa^2 u$ *as has often been thought"(*κ*-Konditionszahl,* u*-Maschinengenauigkeit).*

7.4 Berechnung des Projektors $\mathbf{P}_{BN}^{\perp}$ bzw. von $\mathbf{P}_{BN}^{\perp}\mathbf{y}$

Bei der spaltenweisen Berechnung der Kaufman-Approximation $\mathbf{J}_K$ treten Ausdrücke der Form $(\mathbf{I}_m - \mathbf{P}_{BN})\,\mathbf{y} = \mathbf{P}_{BN}^{\perp}\mathbf{y}$ für verschiedene rechte Seiten $\mathbf{y} \in \mathbb{R}^m$ auf.

Betrachten wir zunächst den orthogonalen Projektor $\mathbf{P}_{BN}$ auf range $(\mathbf{BN})$.

$$\mathbf{P}_{BN} = (\mathbf{BN})\,(\mathbf{BN})^{+} \in \mathbb{R}^{m,m} \tag{7.2}$$

Bekanntlich gilt für spaltenreguläre Matrizen $\mathbf{A} \in \mathbb{R}^{m,n}$ mit $m \geq n$ für die Moore-Penrose-Inverse $\mathbf{A}^{+} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$. Die Matrix $\mathbf{BN} \in \mathbb{R}^{m,n-nactiv}$ ist als Produkt zweier spaltenregulärer Matrizen ebenfalls spaltenregulär. Wir erhalten daher

$$(\mathbf{BN})^{+} = \left(\mathbf{N}^T\mathbf{B}^T\mathbf{BN}\right)^{-1}\mathbf{N}^T\mathbf{B}^T.$$

Berechnet man den Projektor $\mathbf{P}_{BN}$ naiv gemäß (7.2), so hat man die i.allg. vollbesetzte Matrix $\mathbf{BN} \in \mathbb{R}^{m,n-nactiv}$ zu faktorisieren, d.h. die Bandstruktur von $\mathbf{B}$ wird durch die Multiplikation mit der Nullraummatrix $\mathbf{N}$ zerstört. Dies ist besonders für den praktisch relevanten Fall $m \gg n$ nicht hinnehmbar.

Für eine gegebene Knotenfolge $\tilde{\mathbf{t}}$ sei das Subproblem (A) etwa mit den Methoden aus [SK93] gelöst. Als Ergebnis dieses Verfahrens liegt dann eine QR-Faktorisierung der B-Spline Matrix $\mathbf{B} \in \mathbb{R}^{m,n}$ mittels zeilenweiser Givens-Drehungen vor.

$$\mathbf{Q}_0^T\mathbf{B} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}$$

$\mathbf{Q}_0 \in \mathbb{R}^{m,m}$	orthogonal
$\mathbf{R}_0 \in \mathbb{R}^{n,n}$	reguläre obere Dreiecksmatrix (Whitney-Schoenberg-Bedingung!)

Die orthogonale Matrix $\mathbf{Q}_0$ kann auf dem Platz der Bandmatrix $\mathbf{B}$ abgespeichert werden, Produkte der Form $\mathbf{Q}_0\mathbf{y}$ können billig berechnet werden (siehe [SK93]). Unter Benutzung der QR-Faktorisierung von $\mathbf{B}$ ergibt sich

$$\mathbf{B} = \mathbf{Q}_0\begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}$$
$$\mathbf{B}^T\mathbf{B} = \mathbf{R}_0^T\mathbf{R}_0$$

und schließlich

$$(\mathbf{BN})^{+} = \left(\mathbf{N}^T\mathbf{R}_0^T\mathbf{R}_0\mathbf{N}\right)^{-1}\mathbf{N}^T\left[\mathbf{R}_0^T|\mathbf{0}\right]\mathbf{Q}_0^T, \tag{7.3}$$

$$\mathbf{P}_{BN} = \mathbf{Q}_0\begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}\mathbf{N}\left(\mathbf{N}^T\mathbf{R}_0^T\mathbf{R}_0\mathbf{N}\right)^{-1}\mathbf{N}^T\left[\mathbf{R}_0^T|\mathbf{0}\right]\mathbf{Q}_0^T. \tag{7.4}$$

Wir berechnen nun die QR-Faktorisierung von $\mathbf{R}_0\mathbf{N}$ mittels Householder-Transformationen. Man beachte, daß $\mathbf{R}_0\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ als Produkt einer oberen Dreiecksmatrix und einer i.allg. vollbesetzten Nullraumbasis vollbesetzt ist (allerdings ist jetzt die Dimension n).

QR**-Faktorisierung von $\mathbf{R}_0\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$**

$$\mathbf{Q}_2^T\left(\mathbf{R}_0\mathbf{N}\right) = \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$$

$\mathbf{Q}_2 \in \mathbb{R}^{n,n}$	orthogonal
$\mathbf{R}_2 \in \mathbb{R}^{n-nactiv,n-nactiv}$	reguläre obere Dreiecksmatrix

Mit $\mathbf{R}_0\mathbf{N} = \mathbf{Q}_2 \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$ erhalten wir

$$\begin{aligned} \mathbf{N}^T\mathbf{R}_0^T\mathbf{R}_0\mathbf{N} &= \left[\mathbf{R}_2^T|\mathbf{0}\right] \underbrace{\mathbf{Q}_2^T\mathbf{Q}_2}_{\mathbf{I}} \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix} \\ &= \mathbf{R}_2^T\mathbf{R}_2 \end{aligned}$$

und mittels $\mathbf{N} = \mathbf{R}_0^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$ schließlich

$$\begin{aligned} \mathbf{N}\left(\mathbf{N}^T\mathbf{R}_0^T\mathbf{R}_0\mathbf{N}\right)^{-1}\mathbf{N}^T &= \mathbf{N}\mathbf{R}_2^{-1}\mathbf{R}_2^{-T}\mathbf{N}^T \\ &= \mathbf{R}_0^{-1}\mathbf{Q}_2 \underbrace{\begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix} \mathbf{R}_2^{-1}}_{\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}} \underbrace{\mathbf{R}_2^{-T}\left[\mathbf{R}_2^T|\mathbf{0}\right]}_{[\mathbf{I}|\mathbf{0}]} \mathbf{Q}_2^T\mathbf{R}_0^{-T} \\ &= \mathbf{R}_0^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}_2^T\mathbf{R}_0^{-T}. \end{aligned}$$

Wir definieren den Projektor

$$\mathbf{P}_{n-nactiv} := \begin{bmatrix} \mathbf{I}_{n-nactiv} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{n,n}.$$

Setzen wir dies nun in (7.4) ein, so erhält man

$$\mathbf{P}_{BN} = \mathbf{Q}_0 \underbrace{\begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix} \mathbf{R}_0^{-1}}_{\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}} \mathbf{Q}_2\mathbf{P}_{n-nactiv}\mathbf{Q}_2^T \underbrace{\mathbf{R}_0^{-T}\left[\mathbf{R}_0^T|\mathbf{0}\right]}_{[\mathbf{I}|\mathbf{0}]} \mathbf{Q}_0^T.$$

Mit dem weiteren Projektor

$$\mathbf{P}_n := \begin{bmatrix} \mathbf{I}_n \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{m,m}$$

ergibt sich

$$\mathbf{P}_{BN} = \mathbf{Q}_0\mathbf{P}_n\mathbf{Q}_2\mathbf{P}_{n-nactiv}\mathbf{Q}_2^T\mathbf{P}_n^T\mathbf{Q}_0^T.$$

Lemma 19 (Berechnung des Projektors $\mathbf{P}_{BN}^{\perp}$ bzw. von $\mathbf{P}_{BN}^{\perp}\mathbf{y}$)
Für gegebene spaltenreguläre Matrizen $\mathbf{B} \in \mathbb{R}^{m,n}$ sowie $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ seien die QR-Zerlegungen

$$\mathbf{Q}_0^T \mathbf{B} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}$$

und

$$\mathbf{Q}_2^T \left(\mathbf{R}_0 \mathbf{N}\right) = \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$$

berechnet. Dann gilt

$$\begin{aligned} \mathbf{P}_{BN} &= \mathbf{Q}_0 \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}_2^T \left[\mathbf{I}|\mathbf{0}\right] \mathbf{Q}_0^T \\ &= \mathbf{Q}_0 \mathbf{P}_n \mathbf{Q}_2 \mathbf{P}_{n-nactiv} \mathbf{Q}_2^T \mathbf{P}_n^T \mathbf{Q}_0^T. \end{aligned}$$

Für beliebige Vektoren $\mathbf{y} \in \mathbb{R}^m$ gilt

$$\mathbf{P}_{BN}^{\perp}\mathbf{y} = \mathbf{Q}_0 \left[\mathbf{I}_m - \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}_2^T \left[\mathbf{I}|\mathbf{0}\right]\right] \mathbf{Q}_0^T \mathbf{y}.$$

7.5 Ein Algorithmus zur Berechnung der Kaufman-Jacobi-Matrix

Nun sind wir in der Lage, einen vollständigen Algorithmus zur Berechnung der verwendeten Approximation der Jacobi-Matrix anzugeben. Wir berechnen die Kaufman-Approximation der Jacobi-Matrix $\mathbf{J}_K = \left(\mathbf{I}_m - \mathbf{P}_{BN}\right)\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \mathbf{B}\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right)$ spaltenweise.

Algorithmus 10 (Berechnung der Jacobi-Matrix, Kaufman-Approximation)
S1: Berechne die QR-Faktorisierung von $\mathbf{B} \in \mathbb{R}^{m,n}$ *mittels zeilenweiser Givens-Drehungen*

$$\mathbf{Q}_0^T \mathbf{B} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}$$

S2: Berechne $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ *als Lösung von*

$$\min \left\{ \frac{1}{2} \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2 \,:\, \mathbf{L} \le \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \le \mathbf{U} \,:\, \boldsymbol{\alpha} \in \mathbb{R}^n \right\}$$

S3: Setze

$$\bar{\mathbf{R}} := - \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i \in \mathcal{I}} \in \mathbb{R}^{nactiv,n}$$

S4: Berechne die QR-Faktorisierung von $\bar{\mathbf{R}}(\tilde{\mathbf{t}})^T \in \mathbb{R}^{n,nactiv}$ *mittels Householder-Transformationen*

$$\mathbf{Q}_1^T \bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{Q}_{11}^T \\ \mathbf{Q}_{12}^T \end{bmatrix} \bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$$

S5: Basis $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ *des Nullraumes von* $\bar{\mathbf{R}}$

$$\mathbf{N} := \mathbf{Q}_{12}$$

S6: Berechne die QR-Faktorisierung von $\mathbf{R}_0\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ *mittels Householder-Transformationen*

$$\mathbf{Q}_2^T (\mathbf{R}_0 \mathbf{N}) = \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$$

S7: for $\tau := 1$ *to* l *do*

S7.1: Berechne $\mathbf{v}^1 := \mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau = -\boldsymbol{\partial}\mathbf{B}(\tilde{\mathbf{t}})\,[\mathbf{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}})$; $\{\mathbf{v}^1 = \mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau \in \mathbb{R}^m\}$
S7.2: Berechne $\mathbf{v}^2 := \bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau = -\boldsymbol{\partial}\bar{\mathbf{R}}(\tilde{\mathbf{t}})\,[\mathbf{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}})$; $\{\mathbf{v}^2 = \bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^{nactiv}\}$
S7.3: Löse das System $\mathbf{R}_1^T\mathbf{R}_1\mathbf{v}^3 = \mathbf{v}^2$; *(*$\mathbf{R}_1$ *reguläre obere Dreiecksmatrix)* $\{\mathbf{v}^3 = (\mathbf{R}_1^T\mathbf{R}_1)^{-1}\bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^{nactiv}\}$
S7.4: $\mathbf{v}^4 := \bar{\mathbf{R}}^T\mathbf{v}^3$; $\{\mathbf{v}^4 = \bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^n\}$
S7.5: $\mathbf{v}^5 := \mathbf{B}\mathbf{v}^4$; $\{\mathbf{v}^5 = \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^m\}$
S7.6: $\mathbf{v}^6 := \mathbf{v}^1 + \mathbf{v}^5$; $\{\mathbf{v}^6 = \mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau + \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^m\}$
S7.7: Berechne $\mathbf{v}^7$

$$\mathbf{v}^7 = \mathbf{Q}_0 \left[\mathbf{I}_m - \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}_2^T \,[\mathbf{I}|\mathbf{0}] \right] \mathbf{Q}_0^T \mathbf{v}^6;$$

$$\{\mathbf{v}^7 = \mathbf{P}_{BN}^{\perp}\left(\mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau + \mathbf{B}\bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau\right) \in \mathbb{R}^m\}$$

S7.8: $\mathbf{J}_K\mathbf{e}^\tau := \mathbf{v}^7$;

7.6 Ein Algorithmus zur Lösung des reduzierten Problems

Abschließend geben wir ein verallgemeinertes, regularisiertes Gauß-Newton Verfahren mit Dämpfung zur Lösung des reduzierten Problems an.

Sei $\mathbf{F}(\tilde{\mathbf{t}}) := \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ die Residiumsfunktion des reduzierten Funktionals. Im folgenden werden wir das Gauß-Newton-Modell für die Residiumsfunktion $\mathbf{F}$ als Golub-Pereyra-Modell bezeichnen.

Definition 7 (Golub-Pereyra-Modell)

$$\mu_{GP}(\tilde{\mathbf{t}}+\mathbf{s}) := \frac{1}{2}\left\|\mathbf{F}(\tilde{\mathbf{t}})+\mathbf{F}'(\tilde{\mathbf{t}})\mathbf{s}\right\|^2 \approx f(\tilde{\mathbf{t}}+\mathbf{s})$$

mit

$$\mathbf{J}(\tilde{\mathbf{t}}) = \mathbf{F}'(\tilde{\mathbf{t}}) = \mathbf{P}_{BN}^{\perp}\left(\mathfrak{J}_{\tilde{\mathbf{t}}}+\mathbf{B}\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right)+\mathbf{P}_{BN}\left((\mathbf{BN})^{+}\right)^T\mathbf{N}^T\mathbf{K}$$

heißt Golub-Pereyra-Modell für das reduzierte Funktional $f(\tilde{\mathbf{t}}) = \frac{1}{2}\mathbf{F}(\tilde{\mathbf{t}})^T\mathbf{F}(\tilde{\mathbf{t}})$.

Arbeiten wir an Stelle von $\mathbf{J}$ mit der Kaufman-Approximation $\mathbf{J}_K$, so sprechen wir vom Kaufman-Modell.

Definition 8 (Kaufman-Modell)

$$\mu_{K}(\tilde{\mathbf{t}}+\mathbf{s}) := \frac{1}{2}\left\|\mathbf{F}(\tilde{\mathbf{t}})+\mathbf{J}_K(\tilde{\mathbf{t}})\mathbf{s}\right\|^2 \approx f(\tilde{\mathbf{t}}+\mathbf{s})$$

mit

$$\mathbf{J}_K(\tilde{\mathbf{t}}) = \mathbf{P}_{BN}^{\perp}\left(\mathfrak{J}_{\tilde{\mathbf{t}}}+\mathbf{B}\bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right)$$

heißt Kaufman-Modell für das reduzierte Funktional $f(\tilde{\mathbf{t}}) = \frac{1}{2}\mathbf{F}(\tilde{\mathbf{t}})^T\mathbf{F}(\tilde{\mathbf{t}})$

Zur Lösung von Subproblem (A) für eine feste Knotenfolge verwenden wir den Algorithmus aus [SK93]. Die Berechnung der Kaufman-Approximation der Jacobi-Matrix erfolgt gemäß Algorithmus 10. Werden die Ableitungen $\boldsymbol{\partial}\mathbf{B}$ bzw. $\boldsymbol{\partial}\bar{\mathbf{R}}$ exakt berechnet, so nennen wir das Verfahren **RCAP-Ka-ED** (reduced constrained approximation problem, Kaufman model, exact derivatives). Werden diese Ableitungen über finite Differenzen approximiert, so erhalten wir das Verfahren **RCAP-Ka-ID** (reduced constrained approximation problem, Kaufman model, inner discretization). Obwohl prinzipiell möglich, jedoch auf Grund der komplizierten Struktur der Jacobi-Matrix sehr schwierig zu implementieren, haben wir das Golub-Pereyra-Modell nicht mittels exakter Ableitungen **RCAP-GP-ED** (reduced constrained approximation problem, Golub-Pereyra model, exact derivatives) bzw. innerer Diskretisierung **RCAP-GP-ID** (reduced constrained approximation problem, Golub-Pereyra model, inner discretization) implementiert. Wir verwenden stattdessen die Approximation

$$\mathbf{F}'(\tilde{\mathbf{t}})[\mathrm{e}^{\tau}] \approx \frac{\mathbf{F}(\tilde{\mathbf{t}}+h_\tau\mathrm{e}^{\tau})-\mathbf{F}(\tilde{\mathbf{t}})}{h_\tau} \qquad \tau = 1,\ldots,l,$$

welche wir als **RCAP-GP-OD** (reduced constrained approximation problem, Golub-Pereyra model, outer discretization) bezeichnen, da hier die Diskretisierung im äußersten Teil erfolgt. Die Berechnung dieser Jacobi-Matrix erfordert die Lösung von Subproblem (A) an $l+1$ verschiedenen Knotenfolgen!

Die Implementierung der linearen Algebra des Gauß-Newton-Verfahrens sowie der Strahlsuche erfolgt wie in [SS95] bzw. [Sch93] beschrieben.

Algorithmus 11 (Damped Generalized Gauss-Newton Method)
S1: Choose a feasible initial knot sequence $\tilde{\mathbf{t}}^0 \in \mathbb{R}^l$
$\delta \in (0, \frac{1}{4})$

$\nu := 0$

S2: Iteration

Repeat

S2.1: $\mathbf{F} := \mathbf{F}(\tilde{\mathbf{t}}^\nu) = \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}}^\nu)\boldsymbol{\alpha}(\tilde{\mathbf{t}}^\nu)$, *where* $\boldsymbol{\alpha}(\tilde{\mathbf{t}}^\nu)$ *solves Subproblem (A)*

{ *Compute residual function of reduced functional* }

$\mathbf{J} :\approx \mathbf{F}'(\tilde{\mathbf{t}}^\nu)$

{ *Compute an approximation to the Jacobian (Kaufman model or finite differences)* }

S2.2: Compute a descent direction s^ν *from*

$$\min\left\{\tfrac{1}{2}\|\mathbf{F}+\mathbf{J}\mathrm{s}\|^2 \,:\, \mathbf{C}\mathrm{s} \geq \mathbf{h} - \mathbf{C}\tilde{\mathbf{t}}^\nu \,,\, \mathrm{s} \in \mathbb{R}^l\right\}$$

If the problem is ill-conditioned, compute s^ν *from the regularized problem*

$$\min\left\{\tfrac{1}{2}\left\|\begin{bmatrix}\mathbf{F}\\ \mathbf{0}\end{bmatrix}+\begin{bmatrix}\mathbf{J}\\ \sqrt{\lambda}\mathbf{I}\end{bmatrix}\mathrm{s}\right\|^2 \,:\, \mathbf{C}\mathrm{s} \geq \mathbf{h} - \mathbf{C}\tilde{\mathbf{t}}^\nu \,,\, \mathrm{s} \in \mathbb{R}^l\right\}$$

with $\lambda = \sqrt{l \times macheps}\,\left\|\mathbf{J}^T\mathbf{J}\right\|_1$

S2.3: Safeguarded line search

$\gamma := 1.0$

while $f(\tilde{\mathbf{t}}^\nu) - f(\tilde{\mathbf{t}}^\nu + \gamma\mathrm{s}^\nu) < -\gamma\delta\nabla f(\tilde{\mathbf{t}}^\nu)^T\mathrm{s}^\nu$ *do*

$\gamma := \gamma * \alpha$

{ *Here* α *is chosen so that* $\gamma * \alpha$ *is the minimizer of a quadratic or cubic polynomial that models* $f(\tilde{\mathbf{t}} + \gamma\mathrm{s}^\nu)$ *provided that* α *is not to close to the boundary points* 0 *or* 1 }

S2.4: $\tilde{\mathbf{t}}^{\nu+1} := \tilde{\mathbf{t}}^\nu + \gamma\mathrm{s}^\nu$

$\nu := \nu + 1$

until $\nu > \nu max$ *or* convergence

8 Numerische Lösung des reduzierten Problems im Fall der Spline-Glättung

Wir betrachten jetzt die numerische Lösung des reduzierten Problems im Fall der Spline-Glättung.

$$\underset{\tilde{\mathbf{t}}\in\mathbb{R}^l}{\text{minimiere}} \quad f(\tilde{\mathbf{t}}) := \frac{1}{2}\left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}) \right\|^2$$

bei

$$\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0},$$

wobei $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ das Subproblem (A)

$$\underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\text{minimiere}} \quad \frac{1}{2}\left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha} \right\|^2$$

bei

$$\mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U}$$

löst.
Die entsprechenden Formeln für die Residiumsfunktion, den Gradienten, die Jacobi- und Hesse-Matrix des reduzierten Funktionals f im Fall der Spline-Glättung erhält man aus den Formeln für die Spline-Approximation, indem man das Paar $\{\mathbf{B}(\tilde{\mathbf{t}}), \mathbf{y}\}$ durch die Größen $\left\{ \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix}, \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \right\}$ ersetzt.
Man beachte, daß die Nebenbedingungen und somit auch alle direkt damit zusammenhängenden Größen (wie $\bar{\mathbf{R}}$, $\mathbf{N}$ und $\bar{\mathbf{\Gamma}}$) von dieser Veränderung nicht betroffen sind. In Analogie zum vorher betrachteten Fall der Spline-Approximation definieren wir

$$\bar{\mathbf{R}} = -\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i\in\mathcal{I}} \qquad \bar{\mathbf{R}} \in \mathbb{R}^{nactiv,n}$$

$$\bar{\mathbf{\Gamma}} = -\partial \begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i\in\mathcal{I}} \boldsymbol{\alpha}(\tilde{\mathbf{t}}) \qquad \bar{\mathbf{\Gamma}} \in \mathbb{R}^{nactiv,l}$$

$$\mathbf{J}_{\tilde{\mathbf{t}}} = -\partial \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}) \qquad \mathbf{J}_{\tilde{\mathbf{t}}} \in \mathbb{R}^{m+n-r,l}$$

8.1 Die Kaufman-Jacobi-Matrix im Fall der Spline-Glättung

Für die Kaufman-Approximation $\mathbf{J}_K$ der Jacobi-Matrix erhält man mit diesen Bezeichnungen

$$\begin{aligned}
\mathbf{J}_K := \mathbf{P}^{\perp}\psi &= \mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N} \left(\mathbf{J}_{\tilde{\mathbf{t}}} + \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^+ \bar{\mathbf{\Gamma}} \right) \in \mathbb{R}^{m+n-r,l} \\
&= \left[\mathbf{I}_{m+n-r} - \left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \mathbf{N} \right) \left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \mathbf{N} \right)^+ \right] \left(\mathbf{J}_{\tilde{\mathbf{t}}} + \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^+ \bar{\mathbf{\Gamma}} \right)
\end{aligned}$$

Die Größen auf der rechten Seite sind dabei an der Stelle $\tilde{\mathbf{t}}$ zu berechnen.

8.2 Berechnung der Matrizen, in denen Ableitungen auftreten

Im Vergleich zur Spline-Approximation ändert sich lediglich die Berechnung von $\mathbf{J}_{\tilde{\mathbf{t}}} \in \mathbb{R}^{m+n-r,l}$. Im Fall der Spline-Glättung erhält man

$$\mathbf{J}_{\tilde{\mathbf{t}}} = -\partial \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}).$$

Wir berechnen $\mathbf{J}_{\tilde{\mathbf{t}}}$ erneut spaltenweise. Dabei können wir zur Berechnung der ersten m Komponenten von $\mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^{\tau}$ $(\tau = 1, \ldots, l)$ Algorithmus 3 nutzen. Zur Berechnung der letzten $n-r$ Komponenten benötigen wir die Ableitung der Glättungsmatrix $\mathbf{S}_r(\tilde{\mathbf{t}})$ nach den freien Knoten. Während es für die Approximation $\mathbf{S}_r(\tilde{\mathbf{t}}) = \tilde{\mathbf{S}}_r(\tilde{\mathbf{t}})$ des Glättungsterms noch relativ einfach ist, die Ableitungen $\partial S_{i,j}/\partial\tilde{\mathbf{t}}_{p(\tau)}$ exakt zu berechnen (siehe unten), erscheint dies für den exakten Glättungsterm $\mathbf{S}_r(\tilde{\mathbf{t}}) = \bar{\mathbf{S}}_r(\tilde{\mathbf{t}})$ selbst für kleine r als wenig aussichtsreich.

8.2.1 Ableitung der Glättungsmatrix nach den Knoten

Wir betrachten nun die Fréchet-Ableitung der exakten Glättungsmatrix

$$\bar{\mathbf{S}}_r(\tilde{\mathbf{t}}) := \bar{\mathbf{F}}_r(\tilde{\mathbf{t}})\mathbf{D}_r(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-r,n}$$

bzw. der approximierten Glättungsmatrix

$$\tilde{\mathbf{S}}_r(\tilde{\mathbf{t}}) := \tilde{\mathbf{F}}_r(\tilde{\mathbf{t}})\mathbf{D}_r(\tilde{\mathbf{t}}) \in \mathbb{R}^{n-r,n}$$

nach den freien Knoten. Die Matrix $\bar{\mathbf{F}}_r \in \mathbb{R}^{n-r,n-r}$ ist dabei als Cholesky-Faktor der Matrix $\bar{\mathbf{M}}_r = \left(\bar{m}_{ij}^{(r)}\right) \in \mathbb{R}^{n-r,n-r}$ mit

$$\bar{m}_{ij}^{(r)} := \int_a^b B_{i,k-r}(x)B_{j,k-r}(x)\,dx$$

definiert, wohingegen $\tilde{\mathbf{F}}_r \in \mathbb{R}^{n-r,n-r}$ die Form

$$\tilde{\mathbf{F}}_r := \operatorname{diag}\left(\sqrt{\frac{t_{k+j}-t_{r+j}}{k-r}}\right)_{j=1,\ldots,n-r}$$

hat. Wir erhalten daher für die Fréchet-Ableitung

$$\partial\bar{\mathbf{S}}_r = \left(\partial\bar{\mathbf{F}}_r\right)\mathbf{D}_r + \bar{\mathbf{F}}_r\left(\partial\mathbf{D}_r\right) \tag{8.1}$$

bzw.

$$\partial\tilde{\mathbf{S}}_r = \left(\partial\tilde{\mathbf{F}}_r\right)\mathbf{D}_r + \tilde{\mathbf{F}}_r\left(\partial\mathbf{D}_r\right). \tag{8.2}$$

Die Fréchet-Ableitung der Matrix $\tilde{\mathbf{F}}_r$ läßt sich explizit angeben. Dagegen ist die explizite Angabe von $\partial\bar{\mathbf{F}}_r$ sehr schwierig, da die Matrix $\bar{\mathbf{F}}_r$ selbst nur implizit als Cholesky-Faktor der Gramschen Matrix $\bar{\mathbf{M}}_r$ gegeben ist.

Wir betrachten daher nur $\partial\tilde{\mathbf{F}}_r[\mathbf{e}^{\tau}]$ bzw. $\partial\tilde{\mathbf{F}}_r[\mathbf{e}^{\tau}]\mathbf{c}^{n-r}$ für $\tau = 1, \ldots, l$ und $\mathbf{c}^{n-r} \in \mathbb{R}^{n-r}$. Es gilt

$$\partial\tilde{\mathbf{F}}_r[\mathbf{e}^{\tau}]\mathbf{c}^{n-r} \in \mathfrak{L}\left(\mathbb{R}^{n-r}\right)$$

bzw. in Komponentenschreibweise

$$\boldsymbol{\partial}\tilde{\mathbf{F}}_r[\mathrm{e}^\tau]\mathbf{c}^{n-r} = \left(\sum_{j=1}^{n-r} \frac{\partial \tilde{F}_{ij}(\tilde{\mathbf{t}})}{\partial t_{p(\tau)}} c_j \right)_{i=1,\ldots,n-r.}$$

Laut Definition von $\left(\tilde{\mathbf{F}}_r\right)_{ij} = \tilde{F}_{ij}$ gilt

$$\tilde{F}_{ij}(\tilde{\mathbf{t}}) = \begin{cases} 0 & \text{falls } i \neq j, \\ \sqrt{\frac{t_{k+j}-t_{r+j}}{k-r}} & \text{falls } i = j; i,j = 1,\ldots,n-r, \end{cases}$$

also

$$\frac{\partial \tilde{F}_{ij}(\tilde{\mathbf{t}})}{\partial t_{p(\tau)}} = 0 \qquad \text{falls } i \neq j,$$

$$\frac{\partial \tilde{F}_{jj}(\tilde{\mathbf{t}})}{\partial t_{p(\tau)}} = \begin{cases} -\frac{1}{2}\sqrt{\frac{k-r}{t_{k+j}-t_{r+j}}} & \text{falls } p(\tau) = k+j, \\ \frac{1}{2}\sqrt{\frac{k-r}{t_{k+j}-t_{r+j}}} & \text{falls } p(\tau) = r+j, \\ 0 & \text{sonst.} \end{cases}$$

Damit erhalten wir folgenden Algorithmus zur Berechnung von $\boldsymbol{\partial}\tilde{\mathbf{F}}_r[\mathrm{e}^\tau]\mathbf{c}^{n-r}$:

Algorithmus 12 (Berechnung von $\boldsymbol{\partial}\tilde{\mathbf{F}}_r[\mathrm{e}^\tau]\mathbf{c}^{n-r} \in \mathbb{R}^{n-r}$)
Sei $\tau \in \{1,\ldots,l\}$ und $\mathbf{c}^{n-r} = (c_1,\ldots,c_{n-r})^T \in \mathbb{R}^{n-r}$. Der Algorithmus überschreibt $\mathbf{c}^{n-r}$ mit $\boldsymbol{\partial}\tilde{\mathbf{F}}_r[\mathrm{e}^\tau]\mathbf{c}^{n-r}$.

for $j := 1$ *to* $n-r$ *do*

  begin

    if $j = p(\tau) - k$ *then*

      $c_j := -\frac{c_j}{2}\sqrt{\frac{k-r}{t_{k+j}-t_{r+j}}}$

    else

    if $j = p(\tau) - k$ *then*

      $c_j := \frac{c_j}{2}\sqrt{\frac{k-r}{t_{k+j}-t_{r+j}}}$

    else

      $c_j := 0;$

  end;

Schließlich können wir den Algorithmus zur Berechnung der Ableitung der approximierten Glättungsmatrix nach den Knoten angeben. Wegen (8.2) gilt für $\tau \in \{1,\ldots,l\}$ und $\boldsymbol{\alpha} \in \mathbb{R}^n$

$$\left(\boldsymbol{\partial}\tilde{\mathbf{S}}_r\right)[\mathrm{e}^\tau]\boldsymbol{\alpha} = \left(\boldsymbol{\partial}\tilde{\mathbf{F}}_r\right)[\mathrm{e}^\tau]\mathbf{D}_r\boldsymbol{\alpha} + \tilde{\mathbf{F}}_r\left(\boldsymbol{\partial}\mathbf{D}_r\right)[\mathrm{e}^\tau]\boldsymbol{\alpha} \in \mathbb{R}^{n-r}.$$

Algorithmus 13 (Berechnung von $\mathbf{v} := \left(\boldsymbol{\partial}\tilde{\mathbf{S}}_r\right)[\mathrm{e}^\tau]\boldsymbol{\alpha} \in \mathbb{R}^{n-r}$)

S1: Berechne $\mathbf{v}^1 := (\boldsymbol{\partial}\mathbf{D}_r)[\mathrm{e}^\tau]\boldsymbol{\alpha} \in \mathbb{R}^{n-r}$ *mittels Algorithmus 5;*

S2: Berechne $\mathbf{v}^1 := \tilde{\mathbf{F}}_r\mathbf{v}^1 \in \mathbb{R}^{n-r}$*;*

S3: Berechne $\mathbf{v}^2 := \mathbf{D}_r\boldsymbol{\alpha} \in \mathbb{R}^{n-r}$*;*

S4: Berechne $\mathbf{v}^2 := \left(\boldsymbol{\partial}\tilde{\mathbf{F}}_r\right)[\mathrm{e}^\tau]\mathbf{v}^2 \in \mathbb{R}^{n-r}$ *mittels Algorithmus 12;*

S5: Setze $\mathbf{v} := \mathbf{v}^1 + \mathbf{v}^2$*;* $\qquad \{\mathbf{v} = \left(\boldsymbol{\partial}\tilde{\mathbf{S}}_r\right)[\mathrm{e}^\tau]\boldsymbol{\alpha} \in \mathbb{R}^{n-r}\}$

Im Fall des exakten Glättungsterms approximieren wir $\partial S_{i,j}/\partial t_{p(\tau)}$ durch finite Differenzen.

Algorithmus 14 (Berechnung von $-\boldsymbol{\partial}\sqrt{\mu}\bar{\mathbf{S}}_r(\tilde{\mathbf{t}})\boldsymbol{\alpha}(\tilde{\mathbf{t}})$, finite Differenzen)

for $\tau := 1$ *to* l *do*

begin

$h_\tau := \epsilon_1\left(|t_{p(\tau)}| + \epsilon_2\right)$*;*

for $i := 1$ *to* $n - r$ *do*

begin

$F_i := 0$*;*

for $j := 1$ *to* n *do*

$F_i := F_i - \frac{\bar{S}_{i,j}(\tilde{\mathbf{t}}+h_\tau\mathrm{e}^\tau)-\bar{S}_{i,j}(\tilde{\mathbf{t}})}{h_\tau} \times \alpha_j$*;*

(wegen Struktur von $\bar{\mathbf{S}}_r$ *nur* k *Summanden verschieden von Null!)*

end;

$-\boldsymbol{\partial}\sqrt{\mu}\bar{\mathbf{S}}_r(\tilde{\mathbf{t}})\,[\mathrm{e}^\tau]\,\boldsymbol{\alpha}(\tilde{\mathbf{t}}) := \sqrt{\mu}\,\mathbf{F}$*;*

end

8.3 Berechnung einer Nullraumbasis N bzw. von Vektoren $\bar{\mathbf{R}}^+\mathbf{y}$

Da sich die Nebenbedingungen an Ableitungen und somit $\bar{\mathbf{R}}$ im Vergleich zur Spline-Approximation nicht geändert haben, kann man sofort die Aussagen und Algorithmen des Abschnitts 7 übernehmen.

8.4 Berechnung des Projektors $\mathbf{P}^\perp_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S\end{smallmatrix}\right]N}$ bzw. von $\mathbf{P}^\perp_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S\end{smallmatrix}\right]N}\mathbf{y}$

Es gilt

$$\mathbf{P}^\perp_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S\end{smallmatrix}\right]N} = \mathbf{I}_{m+n-r} - \left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\right)\left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\right)^+ \in \mathbb{R}^{m+n-r,m+n-r}.$$

Aus der numerischen Lösung von Subproblem (A) sind die folgenden Faktorisierungen bekannt (siehe [SK93]):

***QR*-Faktorisierung von B**

$$\mathbf{Q}_0^T\mathbf{B} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}$$

$\mathbf{Q}_0 \in \mathbb{R}^{m,m}$	orthogonal
$\mathbf{R}_0 \in \mathbb{R}^{n,n}$	obere Dreiecksmatrix (nicht notwendig regulär!)

QR-Faktorisierung von $\left[\begin{smallmatrix} R_0 \\ \sqrt{\mu} S_r \end{smallmatrix}\right]$

$$\tilde{\mathbf{Q}}^T \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}$$

$\tilde{\mathbf{Q}} \in \mathbb{R}^{2n-r,2n-r}$	orthogonal
$\tilde{\mathbf{R}} \in \mathbb{R}^{n,n}$	reguläre obere Dreiecksmatrix

Für spaltenreguläre Matrizen $\mathbf{A}$ gilt $\mathbf{A}^+ = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$. Die Matrix $\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}$ ist als Produkt zweier spaltenregulärer Matrizen ebenfalls spaltenregulär. Wir erhalten daher

$$\left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\right)^+ = \left(\mathbf{N}^T \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\right)^{-1} \mathbf{N}^T \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T.$$

Es gilt

$$\begin{aligned} \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} &= \mathbf{B}^T\mathbf{B} + \mu\mathbf{S}_r^T\mathbf{S}_r \\ &= \mathbf{R}_0^T\mathbf{R}_0 + \mu\mathbf{S}_r^T\mathbf{S}_r \\ &= \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}. \end{aligned}$$

Damit erhalten wir

$$\begin{aligned} \mathbf{P}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N} &= \left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\right)\left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\right)^+ \\ &= \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\left(\mathbf{N}^T \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\right)^{-1} \mathbf{N}^T \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T. \end{aligned}$$

Unter Benutzung der QR-Faktorisierung von $\begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}$ ergibt sich

$$\begin{aligned} \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} &= \left(\tilde{\mathbf{Q}}\begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}\right)^T \left(\tilde{\mathbf{Q}}\begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}\right) \\ &= \tilde{\mathbf{R}}^T\tilde{\mathbf{R}} \end{aligned}$$

und schließlich

$$\mathbf{P}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]} = \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\mathbf{N}\left(\mathbf{N}^T\tilde{\mathbf{R}}^T\tilde{\mathbf{R}}\mathbf{N}\right)^{-1}\mathbf{N}^T \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T.$$

Wir berechnen nun die QR-Faktorisierung von $\tilde{\mathbf{R}}\mathbf{N}$ mittels Householder-Transformationen. Man beachte, daß $\tilde{\mathbf{R}}\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ als Produkt einer oberen Dreiecksmatrix und einer i.allg. vollbesetzten Nullraumbasis vollbesetzt ist.

QR**-Faktorisierung von $\tilde{\mathbf{R}}\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$**

$$\mathbf{Q}_2^T\left(\tilde{\mathbf{R}}\mathbf{N}\right) = \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$$

$\mathbf{Q}_2 \in \mathbb{R}^{n,n}$	orthogonal
$\mathbf{R}_2 \in \mathbb{R}^{n-nactiv,n-nactiv}$	reguläre obere Dreiecksmatrix

Mit $\tilde{\mathbf{R}}\mathbf{N} = \mathbf{Q}_2 \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$ erhalten wir

$$\mathbf{N}^T\tilde{\mathbf{R}}^T\tilde{\mathbf{R}}\mathbf{N} = \mathbf{R}_2^T\mathbf{R}_2$$

und mittels $\mathbf{N} = \tilde{\mathbf{R}}^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$ schließlich

$$\begin{aligned}\mathbf{N}\left(\mathbf{N}^T\tilde{\mathbf{R}}^T\tilde{\mathbf{R}}\mathbf{N}\right)^{-1}\mathbf{N}^T &= \tilde{\mathbf{R}}^{-1}\mathbf{Q}_2 \underbrace{\begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}\mathbf{R}_2^{-1}}_{\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}} \underbrace{\mathbf{R}_2^{-T}\left[\mathbf{R}_2^T|\mathbf{0}\right]}_{[\mathbf{I}|\mathbf{0}]}\mathbf{Q}_2^T\tilde{\mathbf{R}}^{-T} \\ &= \tilde{\mathbf{R}}^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}_2^T\tilde{\mathbf{R}}^{-T}.\end{aligned}$$

Das ergibt für den Projektor $\mathbf{P}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N}$

$$\mathbf{P}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N} = \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\tilde{\mathbf{R}}^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}_2^T\tilde{\mathbf{R}}^{-T}\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T.$$

Lemma 20 (Berechnung des Projektors $\mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N}$ bzw. von $\mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N}\mathbf{y}$)
Gegeben seien die Matrizen $\mathbf{B} \in \mathbb{R}^{m,n}$ und $\mathbf{S}_r \in \mathbb{R}^{n-r,n}$, so daß für $\mu > 0$ und $m \geq r$ die Matrix $\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \in \mathbb{R}^{m+n-r,n}$ Vollrang n besitzt. Es seien die QR-Zerlegungen

$$\mathbf{Q}_0^T\mathbf{B} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}$$

$$\tilde{\mathbf{Q}}^T\begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}$$

sowie

$$\mathbf{Q}_2^T\left(\tilde{\mathbf{R}}\mathbf{N}\right) = \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$$

für eine spaltenreguläre Matrix $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ bekannt. Dann gilt

$$\mathbf{P}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N} = \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}\tilde{\mathbf{R}}^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{Q}_2^T\tilde{\mathbf{R}}^{-T}\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T.$$

Für beliebige Vektoren $\mathbf{y} \in \mathbb{R}^{m+n-r}$ *gilt*

$$\mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N}\mathbf{y} = \left[\mathbf{I}_{m+n-r} - \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \tilde{\mathbf{R}}^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}_2^T \tilde{\mathbf{R}}^{-T} \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T\right]\mathbf{y}.$$

8.5 Ein Algorithmus zur Berechnung der Kaufman-Jacobi-Matrix

Nun sind wir in der Lage, einen vollständigen Algorithmus zur Berechnung der verwendeten Approximation der Jacobi-Matrix anzugeben. Wir berechnen die Kaufman-Approximation der Jacobi-Matrix

$$\mathbf{J}_K = \left(\mathbf{I}_{m+n-r} - \mathbf{P}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N}\right)\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right)$$

spaltenweise.

Algorithmus 15 (Berechnung der Jacobi-Matrix, Kaufman-Approximation)
S1: Berechne die QR-Faktorisierung von $\mathbf{B} \in \mathbb{R}^{m,n}$ *mittels zeilenweiser Givens-Drehungen*

$$\mathbf{Q}_0^T \mathbf{B} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}$$

S2: Berechne die QR-Faktorisierung von $\begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \in \mathbb{R}^{2n-r,n}$ *mittels zeilenweiser Givens-Drehungen*

$$\tilde{\mathbf{Q}}^T \begin{bmatrix} \mathbf{R}_0 \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix}$$

S3: Berechne $\boldsymbol{\alpha}(\tilde{\mathbf{t}})$ *als Lösung von*

$$\min \left\{ \frac{1}{2} \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}) \right\|^2 \,:\, \mathbf{L} \leq \mathbf{D}_p(\tilde{\mathbf{t}})\boldsymbol{\alpha} \leq \mathbf{U} \,:\, \boldsymbol{\alpha} \in \mathbb{R}^n \right\}$$

unter Benutzung der QR-Faktorisierungen aus S1 und S2

S4: Setze

$$\bar{\mathbf{R}} := -\begin{bmatrix} \mathbf{D}_p(\tilde{\mathbf{t}}) \\ -\mathbf{D}_p(\tilde{\mathbf{t}}) \end{bmatrix}_{i \in \mathcal{I}} \in \mathbb{R}^{nactiv,n}$$

S5: Berechne die QR-Faktorisierung von $\bar{\mathbf{R}}(\tilde{\mathbf{t}})^T \in \mathbb{R}^{n,nactiv}$ *mittels Householder-Transformationen*

$$\mathbf{Q}_1^T \bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{Q}_{11}^T \\ \mathbf{Q}_{12}^T \end{bmatrix} \bar{\mathbf{R}}^T = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$$

S6: Basis $\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ *des Nullraumes von* $\bar{\mathbf{R}}$

$$\mathbf{N} := \mathbf{Q}_{12}$$

S7: Berechne die QR-Faktorisierung von $\tilde{\mathbf{R}}\mathbf{N} \in \mathbb{R}^{n,n-nactiv}$ *mittels Householder-Transformationen*

$$\mathbf{Q}_2^T \left(\tilde{\mathbf{R}}\mathbf{N}\right) = \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{0} \end{bmatrix}$$

S8: for $\tau := 1$ *to* l *do*

S8.1: Berechne $\mathbf{v}^1 := \mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau = -\boldsymbol{\partial} \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} [\mathbf{e}^\tau]\, \boldsymbol{\alpha}(\tilde{\mathbf{t}})$; $\{\mathbf{v}^1 = \mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau \in \mathbb{R}^{m+n-r}\}$

S8.2: Berechne $\mathbf{v}^2 := \bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau = -\boldsymbol{\partial}\bar{\mathbf{R}}(\tilde{\mathbf{t}})\, [\mathbf{e}^\tau]\, \boldsymbol{\alpha}(\tilde{\mathbf{t}})$; $\{\mathbf{v}^2 = \bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^{nactiv}\}$

S8.3: Löse das System $\mathbf{R}_1^T\mathbf{R}_1\mathbf{v}^3 = \mathbf{v}^2$; *(*$\mathbf{R}_1$ *reguläre obere Dreiecksmatrix)* $\{\mathbf{v}^3 = \left(\mathbf{R}_1^T\mathbf{R}_1\right)^{-1} \bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^{nactiv}\}$

S8.4: $\mathbf{v}^4 := \bar{\mathbf{R}}^T\mathbf{v}^3$; $\{\mathbf{v}^4 = \bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^n\}$

S8.5: $\mathbf{v}^5 := \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \mathbf{v}^4$; $\{\mathbf{v}^5 = \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^{m+n-r}\}$

S8.6: $\mathbf{v}^6 := \mathbf{v}^1 + \mathbf{v}^5$; $\{\mathbf{v}^6 = \mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau + \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^+\bar{\boldsymbol{\Gamma}}\mathbf{e}^\tau \in \mathbb{R}^{m+n-r}\}$

S8.7: Berechne $\mathbf{v}^7$

$$\mathbf{v}^7 = \left[\mathbf{I}_{m+n-r} - \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \tilde{\mathbf{R}}^{-1}\mathbf{Q}_2 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}_2^T \tilde{\mathbf{R}}^{-T} \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix}^T\right] \mathbf{v}^6;$$

$$\{\mathbf{v}^7 = \mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N} \left(\mathbf{J}_{\tilde{\mathbf{t}}}\mathbf{e}^\tau + \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\mathbf{e}^\tau\right) \in \mathbb{R}^{m+n-r}\}$$

S8.8: $\mathbf{J}_K\mathbf{e}^\tau := \mathbf{v}^7$*;*

8.6 Ein Algorithmus zur Lösung des reduzierten Problems

Sei $\mathbf{F}(\tilde{\mathbf{t}}) := \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}})$ die Residiumsfunktion des reduzierten Funktionals.

Definition 9 (Golub-Pereyra-Modell)

$$\mu_{GP}(\tilde{\mathbf{t}} + \mathbf{s}) := \frac{1}{2}\left\|\mathbf{F}(\tilde{\mathbf{t}}) + \mathbf{F}'(\tilde{\mathbf{t}})\mathbf{s}\right\|^2 \approx f(\tilde{\mathbf{t}} + \mathbf{s})$$

mit

$$\begin{aligned}\mathbf{J}(\tilde{\mathbf{t}}) = \mathbf{F}'(\tilde{\mathbf{t}}) = \mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N} &\left(\mathbf{J}_{\tilde{\mathbf{t}}} + \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right) \\ &+ \mathbf{P}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N} \left(\left(\begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \mathbf{N}\right)^{+}\right)^T \mathbf{N}^T\mathbf{K}\end{aligned}$$

heißt Golub-Pereyra-Modell für das reduzierte Funktional $f(\tilde{\mathbf{t}}) = \frac{1}{2}\mathbf{F}(\tilde{\mathbf{t}})^T\mathbf{F}(\tilde{\mathbf{t}})$.

Arbeiten wir an Stelle von $\mathbf{J}$ mit der Kaufman-Approximation $\mathbf{J}_K$, so sprechen wir vom Kaufman-Modell.

Definition 10 (Kaufman-Modell)

$$\mu_K(\tilde{\mathbf{t}} + \mathbf{s}) := \frac{1}{2}\left\|\mathbf{F}(\tilde{\mathbf{t}}) + \mathbf{J}_K(\tilde{\mathbf{t}})\mathbf{s}\right\|^2 \approx f(\tilde{\mathbf{t}} + \mathbf{s})$$

mit

$$\mathbf{J}_K := \mathbf{P}^{\perp}_{\left[\begin{smallmatrix} B \\ \sqrt{\mu}S \end{smallmatrix}\right]N} \left(\mathbf{J}_{\tilde{\mathbf{t}}} + \begin{bmatrix} \mathbf{B} \\ \sqrt{\mu}\mathbf{S}_r \end{bmatrix} \bar{\mathbf{R}}^{+}\bar{\mathbf{\Gamma}}\right)$$

heißt Kaufman-Modell für das reduzierte Funktional $f(\tilde{\mathbf{t}}) = \frac{1}{2}\mathbf{F}(\tilde{\mathbf{t}})^T\mathbf{F}(\tilde{\mathbf{t}})$

Zur Lösung von Subproblem (A) für eine feste Knotenfolge verwenden wir wiederum den Algorithmus aus [SK93]. In Analogie zum Fall der Spline-Approximation führen wir die Bezeichnungen **RCSP-Ka-ED** (reduced constrained smoothing problem, Kaufman model, exact derivatives), **RCSP-Ka-ID** (reduced constrained smoothing problem, Kaufman model, inner discretization) sowie **RCSP-GP-ID** (reduced constrained smoothing problem, Golub-Pereyra model, outer discretization) ein.

Algorithmus 16 (Damped Generalized Gauss-Newton Method)

S1: *Choose a feasible initial knot sequence* $\tilde{\mathbf{t}}^0 \in \mathbb{R}^l$

$\delta \in (0, \frac{1}{4})$

$\nu := 0$

S2: *Iteration*

Repeat

S2.1: $\mathbf{F} := \mathbf{F}(\tilde{\mathbf{t}}^\nu) = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{B}(\tilde{\mathbf{t}}^\nu) \\ \sqrt{\mu}\mathbf{S}_r(\tilde{\mathbf{t}}^\nu) \end{bmatrix} \boldsymbol{\alpha}(\tilde{\mathbf{t}}^\nu)$, *where* $\boldsymbol{\alpha}(\tilde{\mathbf{t}}^\nu)$ *solves Subproblem (A)*

{ *Compute residual function of reduced functional* }

$\mathbf{J} :\approx \mathbf{F}'(\tilde{\mathbf{t}}^\nu)$

{ *Compute an approximation to the Jacobian (Kaufman model or finite differences)* }

S2.2: *Compute a descent direction* s^ν *from*

$$\min \left\{ \tfrac{1}{2} \left\| \mathbf{F} + \mathbf{J}\mathrm{s} \right\|^2 \,:\, \mathbf{C}\mathrm{s} \geq \mathbf{h} - \mathbf{C}\tilde{\mathbf{t}}^\nu \,,\, \mathrm{s} \in \mathbb{R}^l \right\}$$

If the problem is ill-conditioned, compute s^ν *from the regularized problem*

$$\min \left\{ \tfrac{1}{2} \left\| \begin{bmatrix} \mathbf{F} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{J} \\ \sqrt{\lambda}\mathbf{I} \end{bmatrix} \mathrm{s} \right\|^2 \,:\, \mathbf{C}\mathrm{s} \geq \mathbf{h} - \mathbf{C}\tilde{\mathbf{t}}^\nu \,,\, \mathrm{s} \in \mathbb{R}^l \right\}$$

with $\lambda = \sqrt{l \times macheps} \left\| \mathbf{J}^T \mathbf{J} \right\|_1$

S2.3: *Safeguarded line search*

$\gamma := 1.0$

while $f(\tilde{\mathbf{t}}^\nu) - f(\tilde{\mathbf{t}}^\nu + \gamma \mathrm{s}^\nu) < -\gamma\delta\nabla f(\tilde{\mathbf{t}}^\nu)^T \mathrm{s}^\nu$ *do*

$\gamma := \gamma * \alpha$

{ *Here* α *is chosen so that* $\gamma * \alpha$ *is the minimizer of a quadratic or cubic polynomial that models* $f(\tilde{\mathbf{t}} + \gamma \mathrm{s}^\nu)$ *provided that* α *is not to close to the boundary points* 0 *or* 1 }

S2.4: $\tilde{\mathbf{t}}^{\nu+1} := \tilde{\mathbf{t}}^\nu + \gamma \mathrm{s}^\nu$

$\nu := \nu + 1$

until $\nu > \nu max$ *or* convergence

9 Numerische Tests

Dieser Abschnitt enthält einige Beispiele, welche die Leistungsfähigkeit der entwickelten Verfahren zeigen sollen. Die Algorithmen wurden in PASCAL implementiert und auf einem 486-PC in IEEE double Arithmetik getestet. Alle Tests wurden mit dem relativen Distanzmaß $\epsilon = 0.0625$ für die Knoten durchgeführt. Die Abbruchkonstanten[1] wurden wie folgt gewählt:

$$\begin{aligned}
\varepsilon_0^t &= 1.0E-10\\
\varepsilon_1^t &= 1.0E-10\\
\varepsilon_2^t &= 1.0E-10\\
\varepsilon_3^t &= 1.0E-06\\
\varepsilon_4^t &= 1.0E-03\\
\varepsilon_5^t &= 1.0E-10
\end{aligned}$$

Es sei darauf hingewiesen, daß die relativ starken Genauigkeitsforderungen, etwa $\varepsilon_3^t = 1.0E-06$, ausschließlich zu Testzwecken verwendet wurden. Beim praktischen Einsatz an realen Daten wird man diese Abbruchkonstante in der Größenordnung $10^{-2} \ldots 10^{-3}$ wählen. Eine Abbruchkonstante in dieser Größenordnung wird auch von Dierckx verwendet.

9.1 Titan-Daten

Unser erstes Testbeispiel sind die in der Spline-Literatur bekannten *Titanium Heat Data* (siehe [dB78]). Sie beschreiben eine Eigenschaft von Titan in Abhängigkeit von der Temperatur. Die Meßpunkte sind äquidistant im Intervall $[595, 1075]$ verteilt. Die Meßwerte verlaufen relativ flach bis auf den typischen Peak im Bereich $x = 900$.

Wir wollen nun die $m = 49$ Datenpunkte durch $n = 11$ B-Splines der Ordnung $k = 4$ approximieren. Wir wählen die Knoten $t_7 = 835$ und $t_{10} = 955$ fest, d.h. $l = 5$, und verteilen die Knoten t_5, t_6 sowie t_8, t_9 und t_{11} jeweils äquidistant in den Intervallen $[595, 835]$, $[835, 955]$ und $[955, 1075]$ [2].

Wir fordern die Konvexität des Splines in den Intervallen $[595, 835)$ und $[955, 1075)$, d.h. $p = 2$ und

$$\begin{aligned}
\mathbf{l} &= (0, 0, 0, -\infty, -\infty, -\infty, 0, 0)^T\\
\mathbf{u} &= (+\infty, +\infty, +\infty, +\infty, +\infty, +\infty, +\infty, +\infty)^T.
\end{aligned}$$

[1] Die entwickelten verallgemeinerten Gauß-Newton-Verfahren liefern folgende Return Codes:

1	$\left\|\mathbf{F}^{\nu+1}\right\| \le \varepsilon_0^t$
2	$\left\|\mathbf{J}^{\nu T}\mathbf{F}^{\nu}\right\| \le \varepsilon_1^t$
3	$\left\|\mathbf{F}^{\nu T}\mathbf{J}^{\nu}\mathbf{s}^{\nu}\right\| \le \varepsilon_2^t$
4	$\left\|\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu}\right\| \le \varepsilon_3^t \left(\|\mathbf{x}^{\nu}\| + \varepsilon_4^t\right)$
5	$\left\|\|\mathbf{F}^{\nu+1}\| - \|\mathbf{F}^{\nu}\|\right\| \le \varepsilon_5^t \|\mathbf{F}^{\nu}\|$
6	$\nu > \nu max$
7	failed otherwise

[2] Die Splineknoten t_j werden in den folgenden Abbildungen durch Kreuze, die Meßdaten (x_i, y_i) durch Kreise dargestellt.

	$\tilde{\mathbf{t}}_1$	**RCAP-Ka-ED**	**RCAP-GP-OD**	**RAP-Ka-ED**
t_5	675.0	6.047707 E+02	7.821307 E+02	5.959958 E+02
t_6	755.0	7.352655 E+02	7.946061 E+02	6.109336 E+02
t_8	875.0	8.756307 E+02	8.755297 E+02	8.767300 E+02
t_9	915.0	8.805912 E+02	8.804966 E+02	8.816219 E+02
t_{11}	1015.0	9.625000 E+02	9.625000 E+02	9.625000 E+02
$\|\mathbf{F}\|$	1.027678 E+00	3.457712 E-01	3.449610 E-01	3.532900 E-01
steps		6	13	9
time (ms)		149	253	116
$\|\mathbf{F}^T\mathbf{J}\mathbf{s}\|$		1.621106 E-04	2.981807 E-11	5.518354 E-12
$\|\mathbf{J}^T\mathbf{F}\|$		3.015662 E-03	2.998139 E-03	1.757170 E-03
Ret. Code		4	3	4

Tabelle 1: Titan-Daten: Spline-Approximation

	$\tilde{\mathbf{t}}_1$	**RCSP-Ka-ED**	**RCSP-GP-OD**	**RSP-Ka-ED**
t_5	675.0	7.975133 E+02	7.822991 E+02	5.959958 E+02
t_6	755.0	8.110142 E+02	7.947857 E+02	6.109336 E+02
t_8	875.0	8.751572 E+02	8.755310 E+02	8.767428 E+02
t_9	915.0	8.810366 E+02	8.804978 E+02	8.816339 E+02
t_{11}	1015.0	9.625000 E+02	9.625000 E+02	9.625000 E+02
$\|\mathbf{F}\|$	1.027722 E+00	3.469246 E-01	3.460394 E-01	3.544604 E-01
steps		7	13	9
time (ms)		149	280	143
$\|\mathbf{F}^T\mathbf{J}\mathbf{s}\|$		2.491557 E-03	3.592591 E-11	5.363720 E-10
$\|\mathbf{J}^T\mathbf{F}\|$		2.797501 E-03	2.988107 E-03	1.749176 E-03
Ret. Code		4	3	4

Tabelle 2: Titan-Daten: Spline-Glättung

Damit erhalten wir

$$\mathbf{L} = (0, 0, 0, 0, -\infty, -\infty, 0, 0, 0)^T$$
$$\mathbf{U} = (+\infty, +\infty, +\infty, +\infty, +\infty, +\infty, +\infty, +\infty, +\infty)^T.$$

In Tabelle 1 sind die Ergebnisse der Spline-Approximation mit freien Knoten mit verschiedenen Verfahren angegeben. Die Bezeichnung **RAP-Ka-ED** (reduced approximation problem, Kaufman model, exact derivatives) steht für ein Verfahren aus [SS95] zur unrestringierten Spline-Approximation mit freien Knoten. Die Lage der Knoten wurde dabei zunächst ohne Berücksichtigung der Konvexitätsnebenbedingungen optimiert, anschließend wurde zu dieser festen Knotenfolge ein formerhaltender Spline berechnet.

Tabelle 2 enthält zum Vergleich die Ergebnisse der Spline-Glättung mit freien Knoten. Es wurde $\mu = 1.0$ und $r = 2$ gewählt.

In den Abbildungen 2 und 4 findet man jeweils den Graph des approximierenden Splines vor und nach dem Optimierungsprozeß. Es ist eine deutliche Verbesserung sowohl der Form des Splines als auch des Residiums $\|\mathbf{F}\|$ zu erkennen. Das Verhalten der zweiten Ableitungen ist in den Abbildungen 3 und 5 dargestellt.

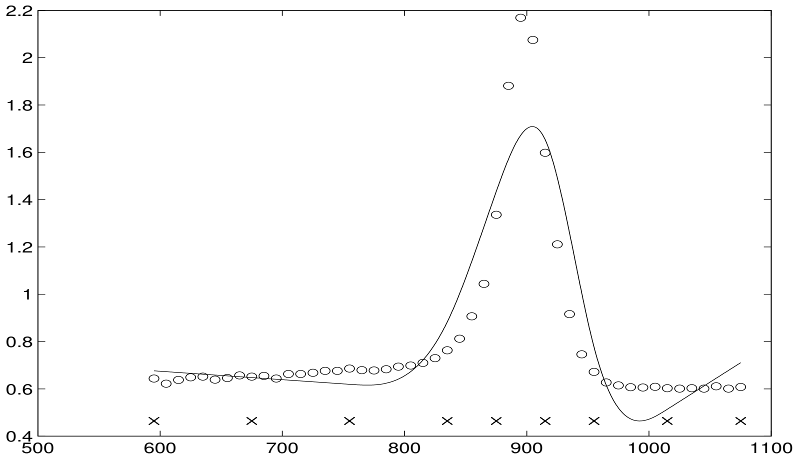


Abbildung 2: Titan-Daten: s zur Startknotenfolge $\tilde{\mathbf{t}}_1$

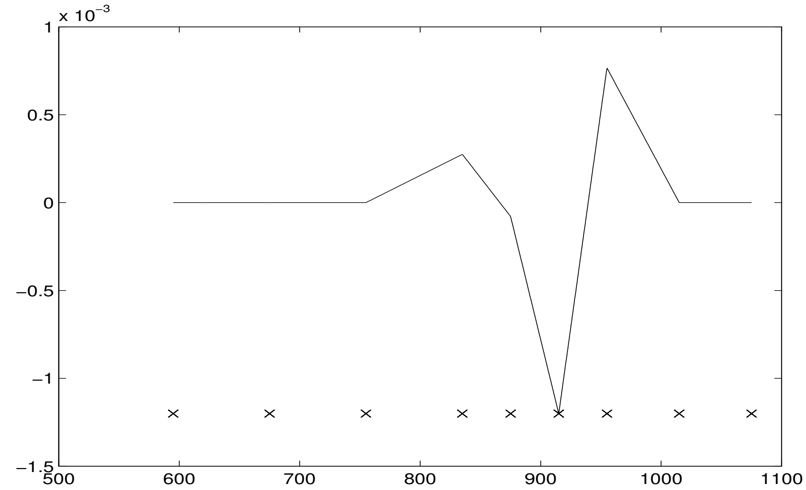


Abbildung 3: Titan-Daten: s'' zur Startknotenfolge $\tilde{\mathbf{t}}_1$

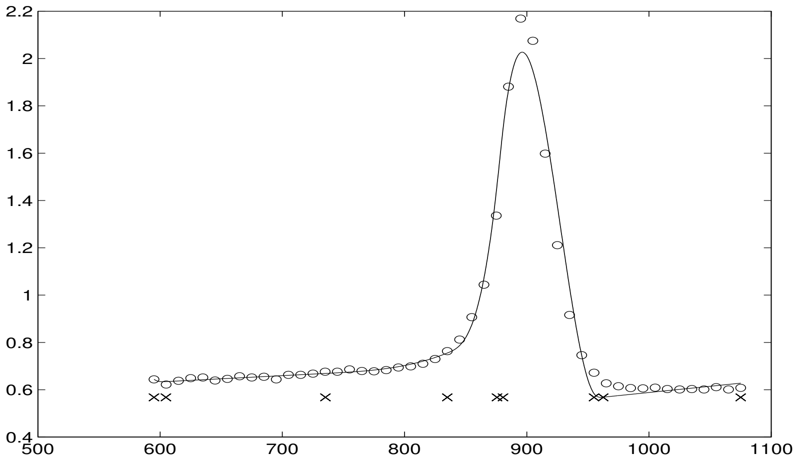


Abbildung 4: Titan-Daten: s zur optimierten Knotenfolge, **RCAP-Ka-ED**

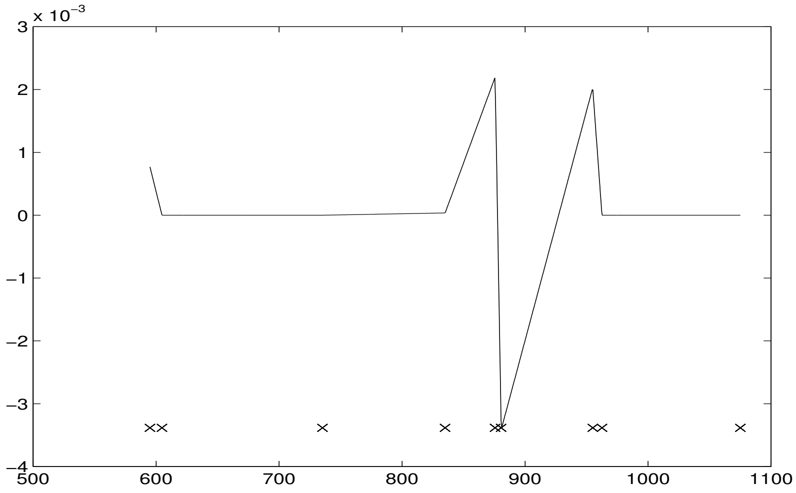


Abbildung 5: Titan-Daten: s'' zur optimierten Knotenfolge, **RCAP-Ka-ED**

9.2 Arctan-Daten

Wir betrachten Meßwerte, welche durch die Funktion

$$g(x) = \arctan(10x)$$

beschrieben werden. Mittels gleichverteilter Pseudozufallszahlen ϵ_i, $-0.075 \leq \epsilon_i \leq 0.075$, erzeugen wir die Werte

$$y_i = f(x_i)(1 + \epsilon_i)$$

an $m = 41$ äquidistanten Punkten im Intervall $[-10, 10]$.

Diese Daten wollen wir durch $n = 8$ B-Splines der Ordnung $k = 4$ approximieren. Wir wählen $l = 4$ äquidistante innere Splineknoten als Ausgangspunkt der Iteration. Der Spline s soll im Intervall $[-10, 10]$ monoton sein. Wir erhalten

$$\mathbf{l} = (0, 0, 0, 0, 0)^T$$
$$\mathbf{u} = (+\infty, +\infty, +\infty, +\infty, +\infty)^T$$

bzw.

$$\mathbf{L} = (0, 0, 0, 0, 0, 0, 0)^T$$
$$\mathbf{U} = (+\infty, +\infty, +\infty, +\infty, +\infty, +\infty, +\infty)^T$$

als Schranken des nichtlinearen Quadratmittelproblems.

Die Tabellen 3 bzw. 4 zeigen die Ergebnisse der Spline-Approximation und Spline-Glättung mit freien Knoten. Für die Spline-Glättung wurde $\mu = 1.0E-3$ und $r = 2$ gewählt.

Wie in Abbildung 6 zu erkennen ist, wird das starke Krümmungsverhalten um den Nullpunkt herum durch die äquidistante Wahl der inneren Knoten nicht angemessen modelliert. Durch die Optimierung der Lage der freien Knoten konnte eine wesentliche Verbesserung der Approximation erreicht werden (Reduktion des Residiums auf ca. 20%). In den Abbildungen 7 und 9 ist das Verhalten der ersten Ableitung und somit die Wirkung der Restriktionen ersichtlich.

Obwohl die Residuen, welche zur Startknotenfolge $\tilde{\mathbf{t}}_1$ gehören, nahezu identisch sind, kann man doch einen Unterschied der Residuen von Spline-Approximation

	$\tilde{\mathbf{t}}_1$	**RCAP-Ka-ED**	**RCAP-GP-OD**	**RAP-Ka-ED**
t_5	-6.0	-8.100201 E-01	-8.074899 E-01	-6.794899 E-01
t_6	-2.0	-1.689546 E-01	-1.946541 E-01	-5.812258 E-02
t_7	2.0	1.598707 E-01	1.855516 E-01	5.200732 E-03
t_8	6.0	8.773101 E-01	8.745770 E-01	9.550505 E-01
$\\|\mathbf{F}\\|$	2.359717 E+00	4.268960 E-01	4.268960 E-01	4.423301 E-01
steps		6	8	8
time (ms)		77	126	77
$\|\mathbf{F}^T\mathbf{J}\mathbf{s}\|$		1.664365 E-14	7.873466 E-09	1.571400 E-12
$\\|\mathbf{J}^T\mathbf{F}\\|$		1.727962 E-07	7.822950 E-07	1.278808 E-02
Ret. Code		3	4	3

Tabelle 3: Arctan-Daten: Spline-Approximation

	$\tilde{\mathbf{t}}_1$	**RCSP-Ka-ED**	**RCSP-GP-OD**	**RSP-Ka-ED**
File	arctan0.m	arctan1.m	arctan2.m	arctan3.m
t_5	-6.0	-8.760175 E-01	-9.652265 E-01	-8.838273 E-01
t_6	-2.0	-2.301281 E-01	-3.258263 E-01	-2.760825 E-01
t_7	2.0	2.743868 E-01	3.454588 E-01	2.779473 E-01
t_8	6.0	8.822404 E-01	9.809856 E-01	1.090103 E-01
$\\|\mathbf{F}\\|$	2.359790 E+00	5.230920 E-01	5.098921 E-01	5.283729 E-01
steps		4	100	13
time (ms)		66	1741	148
$\|\mathbf{F}^T\mathbf{J}\mathbf{s}\|$		8.407664 E-04	4.345827 E-06	1.0626114 E-10
$\\|\mathbf{J}^T\mathbf{F}\\|$		4.336670 E-02	1.270197 E-03	7.8463565 E-03
Ret. Code		4	6	5

Tabelle 4: Arctan-Daten: Spline-Glättung

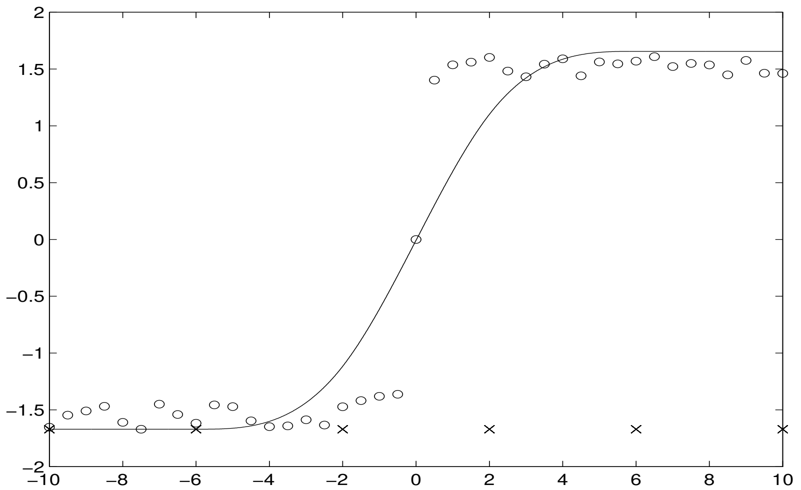


Abbildung 6: Arctan-Daten: s zur Startknotenfolge $\tilde{\mathbf{t}}_1$

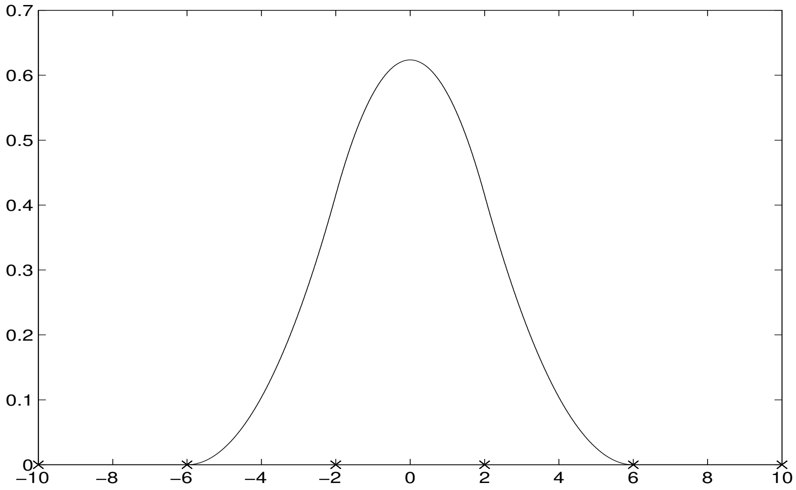


Abbildung 7: Arctan-Daten: s' zur Startknotenfolge

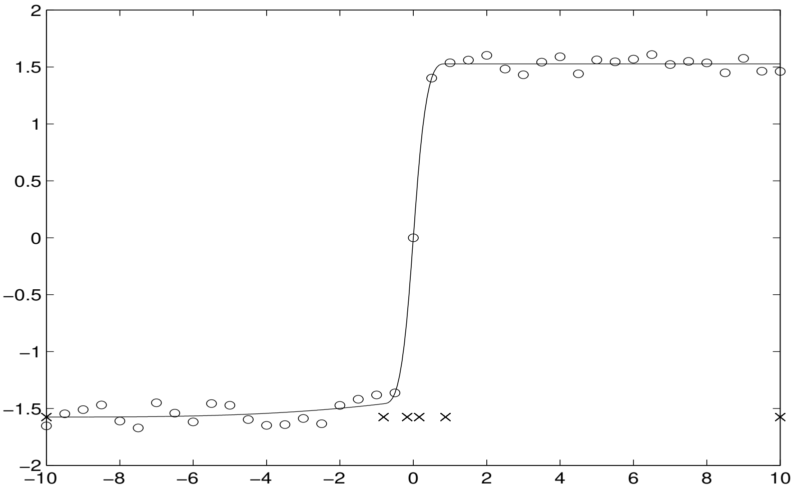


Abbildung 8: Arctan-Daten: s zur optimierten Knotenfolge, **RCAP-Ka-ED**

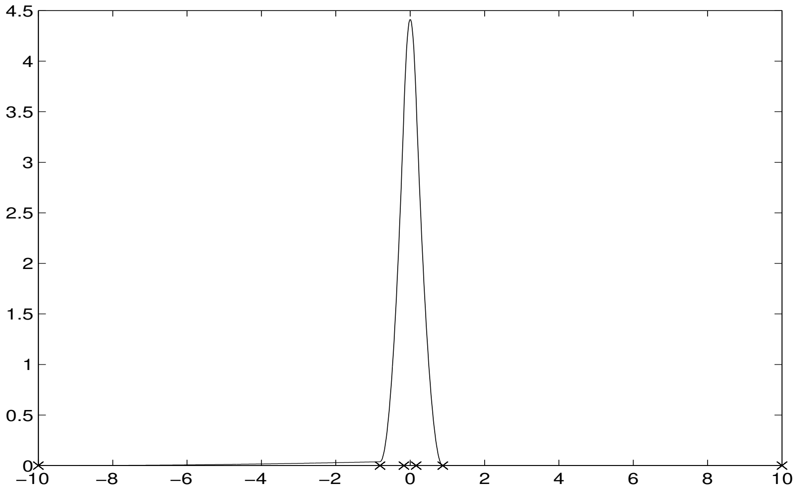


Abbildung 9: Arctan-Daten: s' zur optimierten Knotenfolge, **RCAP-Ka-ED**

und Spline-Glättung an der ermittelten Optimalstelle feststellen. Die Knoten tendieren dazu, sich um den Nullpunkt herum zu konzentrieren. Dies führt an der Optimalstelle zu einem starken Anstieg der zweiten Ableitung und damit zu einem wesentlich größeren Einfluß des Glättungstermes als bei den äquidistanten Startknoten.

Die ersten numerischen Tests lassen bereits erkennen, daß mit den Algorithmen 11 und 16 zur Quadratmittelapproximation durch B-Splines mit freien Knoten und Ungleichheitsnebenbedingungen an Ableitungen zwei robuste und effiziente Werkzeuge zur Approximation fehlerbehafteter Daten geschaffen wurden. Neben der hohen Effizienz liegt der Hauptvorteil des gewählten Zugangs darin, daß Standardverfahren der nichtlinearen Optimierung, welche auf das Ausgangsproblem angewendet werden, oftmals wegen Nichterfüllung der Anordnungsbedingung $\mathbf{C}\tilde{\mathbf{t}} - \mathbf{h} \geq \mathbf{0}$ versagen. Beispiele, welche mit dem SQP-Verfahren der MATLAB-Routine CONSTR und der Spline-Toolbox gerechnet wurden, belegen diese Aussage.

Für ein abschließendes vergleichendes Urteil über die Performance der entwickelten Verfahren sollen noch numerische Tests an realen Daten mit mehr als 100 Meßpunkten erfolgen. Ein Vergleich unserer Routinen mit der FITPACK-Routine CONCON zum automatischen Glätten mit Konvexitätsnebenbedingungen ist vorgesehen.

Literatur

[ADLM90] E. Arge, M. Dæhlen, T. Lyche, and K. Mørken, *Constrained spline approximation of functions and data based on constrained knot removal*, Algorithms for Approximation II (J. C. Mason and M. G. Cox, eds.), Chapman and Hall, 1990, pp. 4–20.

[Bjö91] Å. Björck, *Least squares methods*, Tech. report, Linköping University, Sweden, S-581 83 Linköping, Sweden, 1991.

[Bra86] D. Braess, *Nonlinear approximation theory*, Springer Verlag, 1986.

[BS74] J. H. Bigelow and N. Z. Shapiro, *Implicit function theorems for mathematical programming and for systems of inequalities*, Math. Prog. **6** (1974), 141–156.

[Dan73] J. W. Daniel, *Stability of definite quadratic programs*, Math. Prog. **5** (1973), 41–53.

[dB76] C. de Boor, *Splines as linear combinations of B-splines*, Approximation Theory II (New York) (G. G. Lorentz, C. K. Chui, and L. L. Schumaker, eds.), Academic Press, New York, 1976, pp. 1–47.

[dB78] C. de Boor, *A practical guide to splines*, Springer-Verlag, New York, Heidelberg, Berlin, 1978.

[dBLS76] C. de Boor, T. Lyche, and L. L. Schumaker, *On calculating with B-splines ii: Integration*, Numerische Methoden der Approximationstheorie (Basel) (L. Collatz, H. Werner, and G. Meinardus, eds.), Birkhäuser, Basel, 1976, pp. 123–146.

[dBR68] C. de Boor and J. R. Rice, *Least squares cubic spline approximation II – variable knots*, Technical Report CSD TR 21, Computer Science Department, Purdue University, 1968.

[Die79] P. Dierckx, *Het aanpassen van krommen en oppervlakken aan meetpunten met behulp van spline funkties*, Ph.D. thesis, Katholieke Universiteit Leuven, 1979.

[EA88] T. Elfving and L.-E. Andersson, *An algorithm for computing constrained smoothing spline functions*, Numer. Math. (1988), 583–595.

[Fia76] A. V. Fiacco, *Sensitivity analysis for nonlinear programming using penalty methods*, Math. Prog. **10** (1976), 287–311.

[Fia83] A. V. Fiacco, *Introduction to sensitivity and stability analysis in nonlinear programming*, Academic Press, 1983.

[GP73] G. H. Golub and V. Pereyra, *The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate*, SIAM J. Numer. Anal. **10** (1973), 413–432.

[Jup78] D. L. B. Jupp, *Approximation to data by splines with free knots*, SIAM J. Numer. Anal. **15** (1978), no. 2, 328–343.

[Kau75] L. Kaufman, *A variable projection method for solving separable nonlinear least squares problems*, BIT **15** (1975), no. 4, 49–57.

[Kun95] V. Kunert, *Ein Algorithmus zur Splineapproximation unter Nebenbedingungen*, Manuskript, 1995.

[LM88] T. Lyche and K. Mørken, *A data reduction strategy for splines with applications to the approximation of functions and data*, IMA J. Numer. Anal. **8** (1988), 185–208.

[MU88] C. A. Micchelli and F. I. Utreras, *Smoothing and interpolation in a convex subset of a hilbert space*, SIAM J. Sci. Stat. Comput. **9** (1988), no. 4, 728–746.

[Nür89] G. Nürnberger, *Approximation by spline functions*, Springer-Verlag, 1989.

[Pai73] C. C. Paige, *An error analysis of a method for solving matrix equations*, Math. Comp. **27** (1973), 355–359.

[Par85] T. A. Parks, *Reducible nonlinear programming problems*, Ph.D. thesis, Houston Univ., Dept. of Mathematics, Houston, 1985.

[RW80] A. Ruhe and P. Å. Wedin, *Algorithms for separable nonlinear least squares problems*, SIAM Rev. **22** (1980), no. 3, 318–337.

[Sau72] M. A. Saunders, *Large-scale linear programming using the cholesky factorization*, Tech. Report Report No. CS252, Computer Science Dept., Stanford Univ., 1972.

[Sch81] L. L. Schumaker, *Spline functions: Basic theory*, John Wiley and Sons, New York, 1981.

[Sch91] H. Schwetlick, *Nichtlineare Parameterschätzung: Modelle, Schätzkriterien und numerische Algorithmen*, GAMM-Mitteilungen (1991), 13–51.

[Sch93] T. Schütze, *Quadratmittelapproximation durch B-Splines mit freien Knoten*, Diplomarbeit, Martin-Luther-Universität Halle/Wittenberg, Fachbereich Mathematik und Informatik, 1993.

[SH88] J. W. Schmidt and W. Heß, *Positivity of cubic polynomials on intervals and positive spline interpolation*, BIT **28** (1988), 340–352.

[SK93] H. Schwetlick and V. Kunert, *Spline smoothing under constraints on derivatives*, BIT **33** (1993), 512–528.

[SS90] J. W. Schmidt and I. Scholz, *A dual algorithm for convex-concave data smoothing by cubic C^2-splines*, Numer. Math. **57** (1990), 333–350.

[SS95] H. Schwetlick and T. Schütze, *Least squares approximation by splines with free knots*, BIT **35** (1995), to appear.

[Suc91] P. Suchomski, *Method of optimal variable-knot spline interpolation in the L_2 discrete norm*, Internat. J. Systems Sci. **22** (1991), no. 11, 2263–2274.

[VBH92] A. H. Vermeulen, R. H. Bartels, and G. R. Heppler, *Integrating products of B-splines*, SIAM J. Sci. Statist. Comput. **13** (1992), no. 4, 1025–1038.