Quadratmittelapproximation durch B-Splines mit freien Knoten

Diplomarbeit

Martin-Luther-Universität Halle/Wittenberg
Fachbereich Mathematik und Informatik

eingereicht von Torsten Schütze
geb. am 24.09.1969 in Nordhausen

Betreuer: Prof. Dr. habil. H. Schwetlick
Technische Universität Dresden
Institut für Wissenschaftliches Rechnen

Halle, im Juli 1993

Schütze, Torsten

Quadratmittelapproximation durch B-Splines mit freien Knoten, 1993, 66 S., 40 Lit., 16 Abb.
Diplomarbeit, Martin-Luther-Universität Halle/S., Fachbereich Mathematik und Informatik

Die fehlerbehafteten Meßwerte einer glatten univariaten Funktion sollen durch einen Spline approximiert werden. Der Spline wird als Linearkombination von B-Splines dargestellt. Eine Teilmenge der Splineknoten, die freien Knoten, wird in den Optimierungsprozeß einbezogen. Minimiert wird die Summe der Fehlerquadrate und ein Glättungsterm. Es wird ein nichtlineares Quadratmittelproblem in den Splinekoeffizienten und den freien Knoten entwickelt. Das resultierende separable, restringierte Quadratmittelproblem wird auf ein Problem, in dem nur noch die freien Knoten auftreten, reduziert. Das reduzierte Problem wird mit einem Verallgemeinerten Gauß-Newton-Verfahren gelöst.

An dieser Stelle möchte ich Herrn Prof. Dr. habil. H. Schwetlick für die kontinuierliche Betreuung sowie für die wertvollen Hinweise und Anregungen meinen herzlichen Dank aussprechen.

k	Ordnung der B-Splines
n	Anzahl der B-Splines
m	Anzahl der Meßpunkte
l	Anzahl der freien Splineknoten
r	Ordnung der Ableitung im Glättungsterm
$(\mathbf{x}, \mathbf{y})$	Meßdaten (x_i, y_i), $i = 1, \ldots, m$
$\mathbf{p}$	$\mathbf{p} = (p(1), \ldots, p(l))$ Zeiger auf die freien Knoten
$\mathbf{t}$	(finite) Knotenfolge $\mathbf{t} = (t_1, \ldots, t_n, \ldots, t_{n+k})^T$
$\tilde{\mathbf{t}}$	freie Splineknoten $\tilde{\mathbf{t}} = (t_{p(1)}, \ldots, t_{p(l)})^T$
$B_{j,k}(x, \mathbf{t})$, $B_{j,k}(x)$	j-ter B-Spline der Ordnung k bezüglich der Knotenfolge $\mathbf{t}$ an der Stelle x
$\boldsymbol{\alpha}$	Splinekoeffizienten $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$
$\mathcal{S}_{k,\mathbf{t}}$	Splineraum der Ordnung k bezüglich der Knotenfolge $\mathbf{t}$
$\mathcal{P}_k$	Raum der Polynome k-ter Ordnung
$\mathbf{B}$	$(\mathbf{B})_{i,j} = B_{j,k}(x_i)$ Beobachtungsmatrix
$supp(.)$	Support einer Funktion, $supp f(x) := \overline{\{x \, , \, f(x) \neq 0\}}$
$macheps$	relative Maschinengenauigkeit
$\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$	Normalverteilung mit Erwartungswert $\mathbf{0}$ und Kovarianzmatrix $\sigma^2 \mathbf{I}$
$\#t_i$	Vielfachheit des Knotens t_i
$\Vert . \Vert$	Euklidische Norm
$K(\mathbf{x}, \rho)$	$K(\mathbf{x}, \rho) := \{\Vert \mathbf{y} - \mathbf{x} \Vert < \rho\}$
$\bar{K}(\mathbf{x}, \rho)$	$\bar{K}(\mathbf{x}, \rho) := \{\Vert \mathbf{y} - \mathbf{x} \Vert \leq \rho\}$

Tabelle 1: Verwendete Bezeichnungen

Inhaltsverzeichnis

1 Einleitung

Gegeben seien die Meßpunkte

$$a \leq x_1 < \ldots < x_i < x_{i+1} < \ldots < x_m \leq b, \tag{1.1}$$

und die mit Fehlern behafteten Meßwerte y_i, $i = 1, \ldots, m$, einer unbekannten glatten Funktion f. Die Messungen seien stochastisch unabhängig und normalverteilt

$$\begin{aligned} y_i &= f(x_i) + \epsilon_i, \qquad i = 1, \ldots, m, & (1.2)\\ \boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)^T &= \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}). & (1.3) \end{aligned}$$

Die Funktion f soll durch eine Funktion s approximiert werden, welche

1. möglichst einfach zu berechnen ist,
2. durch wenige Parameter beschrieben werden kann,
3. „möglichst wenig“ von den Meßwerten abweicht.

Da die Meßwerte fehlerbehaftet sind, ist es prinzipiell nicht sinnvoll, die Meßdaten zu interpolieren. Wir wollen die fehlerbehafteten Daten durch eine Summe von B-Splines approximieren. Wir machen dazu den Ansatz

$$\tilde{y}_i = \sum_{j=1}^{n} B_{j,k}(x_i)\alpha_j. \tag{1.4}$$

Die Forderung nach einer möglichst geringen Abweichung lautet dann

$$y_i - \tilde{y}_i \quad \text{„klein“} \quad i = 1, \ldots, m.$$

Unter der Voraussetzung (1.3) ist ein, im stochastischen Sinne „optimales“ Maß der Abweichung die Summe der Fehlerquadrate

$$\sum_{i=1}^{m} [y_i - \tilde{y}_i]^2 \longrightarrow \min, \tag{1.5}$$

$$\sum_{i=1}^{m} \left[y_i - \sum_{j=1}^{n} B_{j,k}(x_i)\alpha_j \right]^2 \longrightarrow \min_{\boldsymbol{\alpha}}. \tag{1.6}$$

(1.6) liefert dann eine Maximum-Likelihood-Schätzung.

Zur Definition von B-Splines benötigt man eine sogenannte Knotenfolge $\mathbf{t}$. Die Festlegung der Knoten erfolgt meist interaktiv oder auf Grund einer heuristischen Strategie (siehe Prozedur NEWNOT in [dB78]).

Ziel dieser Arbeit soll die Einbeziehung der Knotenfolge $\mathbf{t}$ (oder einer Teilmenge $\tilde{\mathbf{t}}$ davon) in den Minimierungsprozeß sein

$$\sum_{i=1}^{m} \left[y_i - \sum_{j=1}^{n} B_{j,k}(x_i, \tilde{\mathbf{t}})\alpha_j \right]^2 \longrightarrow \min_{\boldsymbol{\alpha}, \tilde{\mathbf{t}}}. \tag{1.7}$$

(1.7) ist ein hochgradig nichtlineares, restringiertes (da die Knoten $\mathbf{t}$ nicht beliebig gewählt werden können) Quadratmittelproblem spezieller Struktur.

Neben der *Splineapproximation* (1.6) betrachten wir in dieser Arbeit noch die *Splineglättung*

$$\sum_{i=1}^{m}\left[y_i-\sum_{j=1}^{n}B_{j,k}(x_i)\alpha_j\right]^2+\mu\int_a^b\left[s^{(r)}(x)\right]^2\,dx \quad\longrightarrow\quad \min_{\boldsymbol{\alpha}}. \tag{1.8}$$

Mit dem Glättungsparameter $\mu > 0$ kann die Güte der Approximation s im Vergleich zur Glattheit gesteuert werden (siehe [Wah90] für den statistischen Hintergrund und einige Verallgemeinerungen). Die Einbeziehung der Knotenfolge in den Minimierungsprozeß (1.8) liefert ein ähnlich strukturiertes nichtlineares, restringiertes Quadratmittelproblem wie (1.7).

Die Arbeit ist wie folgt aufgebaut: In Kapitel 2 werden B-Splines über eine Rekursionsformel eingeführt und einige später benötigte Eigenschaften dargestellt. Die Kapitel 3 - 5 enthalten die Grundlagen und verwendeten Algorithmen zur Lösung der auftretenden speziellen linearen und nichtlinearen Quadratmittelprobleme. Kapitel 6 beschäftigt sich mit der Splineapproximation mit freien Knoten, während in Kapitel 7 die Methoden auf die Splineglättung mit freien Knoten ausgedehnt werden. Das Listing des entwickelten Programmpaketes befindet sich in einem gesonderten Anhang.

2 B-Splines

Wir werden im folgenden B-Splines über eine Rekursionsformel definieren und einige später benötigte Eigenschaften angeben.

2.1 Definitionen

Definition 1 (Knotenfolge)
Eine Folge $\mathbf{t} = (\ldots t_{-1}, t_0, t_1, \ldots)$ *mit* $t_{j+1} \geq t_j \quad \forall j$ *und* $\lim\limits_{j\to\pm\infty} t_j = \pm\infty$ *heißt (bi-infinite) Knotenfolge.*

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

$$
\begin{aligned}
B_{j,1}(t) &:= \begin{cases} 1 & t_j \leq t < t_{j+1} \\ 0 & sonst \end{cases} \\
B_{j,k}(t) &:= \omega_{j,k}(t) \cdot B_{j,k-1}(t) + (1 - \omega_{j+1,k}(t)) \cdot B_{j+1,k-1}(t) \quad \textit{für} \quad k > 1 \\
\textit{mit} \quad \omega_{j,k}(t) &:= \begin{cases} \frac{t-t_j}{t_{j+k-1}-t_j} & falls \quad t_j < t_{j+k-1} \\ 0 & sonst. \end{cases}
\end{aligned}
$$

Für $t_j = t_{j+k}$ erkennt man durch wiederholte Anwendung der Rekursionsformel sofort

$$B_{j,k}(t) = 0.$$

Es ist deshalb vernünftig,

$$t_{j+k} \quad > \quad t_j \quad \forall j \tag{2.1}$$

zu fordern, um die trivialen B-Splines $B_{j,k} \equiv 0$ von vornherein von der Betrachtung auszuschließen.
Wir geben jetzt nach [dB78] die wichtigsten Eigenschaften der B-Splines an.

Satz 1 (Eigenschaften der B-Splines)

1. $\text{supp} B_{j,k} = [t_j, t_{j+k}]$ *(falls* $t_{j+k} > t_j$*)*
2. $B_{j,k}(t) > 0 \qquad \forall t \in (t_j, t_{j+k})$
3. $\sum_j B_{j,k}(t) = 1$
4. $B_{j,k_{|[t_j,t_{j+1}]}} \in \mathcal{P}_{k_{|[t_j,t_{j+1}]}}$

Die Glattheitseigenschaften von B-Splines hängen von der Anzahl der zusammenfallenden Knoten ab. Wir definieren daher:

Definition 3 (Knotenvielfachheit)
Der Knoten t_j *hat die Vielfachheit* ν_j*, wenn*

$$\exists i : \qquad t_i < t_j = t_{i+1} = \ldots = t_{i+\nu_j} < t_{i+\nu_j+1}$$

Bezeichnung: $\nu_j = \#t_j$

Definition 4 (Spline,Splineraum)
Die Menge $\mathcal{S}_{k,\mathbf{t}} := \left\{ \sum_j B_{j,k}\alpha_j \ , \ \alpha_j \in \mathbb{R} \right\}$ *heißt Splineraum der Ordnung k bezüglich der Knotenfolge* $\mathbf{t}$*. Die Elemente 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, zitiert nach [dBH85])
Wenn $t_j < t_{j+k}$ für alle j gilt, dann sind die B-Splines $B_{j,k}$ linear unabhängig und bilden eine Basis für den Raum $\tilde{S}_{k,\mathbf{t}}$ aller stückweisen Polynome vom Grad $k-1$ mit Bruchstellen (breakpoints) *t_j, die in t_j $(k-1-\#t_j)$-mal stetig differenzierbar sind.*

Bemerkung 1
Die Beziehung zwischen Glattheit und Knotenvielfachheit kann auch durch die folgende Formel ausgedrückt werden:
Anzahl der Glattheitsbedingungen in t_j + Knotenvielfachheit $\#t_j = k$

Beispiel:

Ordnung k	Knotenvielfachheit $\#t_j$	Glattheit in t_j
3	1	C^1
3	2	C
3	3	unstetig
3	4 (ausgeschlossen)	
4	1	C^2
4	2	C^1
4	3	C
4	4	unstetig
4	5 (ausgeschlossen)	

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

$$\mathbf{t} \quad = \quad (t_1, t_2, \ldots, t_n, t_{n+1}, \ldots, t_{n+k}). \tag{2.2}$$

Dabei bezeichnet 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

$$\mathbf{t} = (t_1, t_2, \ldots, t_n, t_{n+1}, \ldots, t_{n+k})$$

mit

$$t_1 \leq \ldots \leq t_k \leq a \qquad b \leq t_{n+1} \leq \cdots \leq t_{n+k}$$

und $t_{k+1} \leq \ldots \leq t_n \in [a, b]$. Die Knoten $t_{k+1}, \ldots, t_n$ werden als innere Knoten bezeichnet. Es gelte stets $\#t_j \leq k \; \forall j$. 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 einer Empfehlung von [Cox82] zufolge:

$$t_1 = t_2 = \ldots = t_k = a \qquad \text{und} \qquad b = t_{n+1} = t_{n+2} = \ldots = t_{n+k} \tag{2.3}$$

Bemerkung 2
Die Bedingung (2.3) ändert nur die Basis des Splineraumes, ist aber nach [Cox82] günstig für die Kondition der auftretenden Matrizen. Diese Bedingung ist für die weiteren Ausführungen nicht wesentlich.

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

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

Statt $\mathcal{S}_{k,\mathbf{t}_{|[a,b]}}$ werden wir im folgenden kurz $\mathcal{S}_{k,\mathbf{t}}$ schreiben. Die Elemente dieses Raumes bezeichnen wir mit s. Wegen Satz 1 gibt es genau eine Darstellung

$$s(t) = \sum_{j=1}^{n} B_{j,k}(t)\alpha_j \tag{2.4}$$

oder in Vektorschreibweise

$$\begin{aligned} s(t) &= \boldsymbol{\beta}(t)^T \boldsymbol{\alpha} \\ \text{mit} \quad \boldsymbol{\beta}(t) &:= (B_{1,k}(t), \ldots, B_{n,k}(t))^T \\ \text{und} \quad \boldsymbol{\alpha} &:= (\alpha_1, \ldots, \alpha_n)^T . \end{aligned}$$

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

Bemerkung 3

- *Unter der Voraussetzung* $t_{j+k} > t_j \; \forall j$ *ist die Dimension von* $\mathcal{S}_{k,\mathbf{t}}$ *gleich* n.
- *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:

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

Satz 2 (Schoenberg-Whitney 1953)
Die Matrix $\mathbf{B} \in \mathbb{R}^{m,n}$ *hat Vollrang genau dann, wenn es eine Teilfolge der Datenabszissen* (x_{i_j}) *gibt, so daß*

$$B_{j,k}\left(x_{i_j}\right) \neq 0 \qquad j = 1, \ldots, n. \tag{2.5}$$

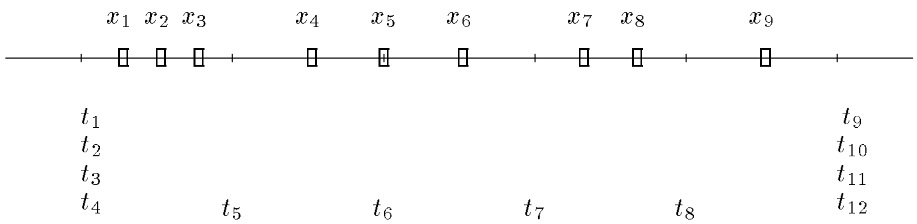
Bemerkung 4

- *Eine hinreichende Bedingung (und bei* $t_j < t_{j+k}$ *auch notwendige) Bedingung für (2.5) ist die Schoenberg-Whitney-Bedingung*

$$t_j < x_{i_j} < t_{j+k}. \tag{2.6}$$

- *In der Originalarbeit von Schoenberg/Whitney[1953] wird der Satz für den Fall der Spline-Interpolation mit* abgebrochenen Potenzfunktionen *bewiesen. Die zitierte Version stammt aus [dB78, Theorem XIII.1, Lemma XIV.2].*
- *Die Matrix* $\mathbf{B}$ *ist sogar total positiv. Aus dieser Tatsache ergeben sich neben einigen wichtigen theoretischen Eigenschaften (*variation diminishing property*) Vereinfachungen für die numerische Implementierung (siehe [dBP77]).*
- *Wegen Eigenschaft 1, Satz 1 hat die Matrix* $\mathbf{B}$ *eine Bandbreite kleiner oder gleich* k.

Beispiel: $n = 8$, $m = 9$, $k = 4$



$$\mathbf{B} = \begin{bmatrix} x & x & x & x & & & & \\ x & x & x & x & & & & \\ x & x & x & x & & & & \\ & x & x & x & x & & & \\ & & x & x & x & 0 & & \\ & & x & x & x & x & & \\ & & & x & x & x & x & \\ & & & x & x & x & x & \\ & & & & x & x & x & x \end{bmatrix} \in \mathbb{R}^{9,8}$$

2.3 Berechnung der Ableitung eines Splines

Wir wollen nun die Ableitung eines Splines berechnen.

Lemma 1 (Ableitung eines Splines)
Wenn $t_j < t_{j+k-1}$, $j = 2, \ldots, n$ gilt, dann ist die Ableitung eines Splines aus $\mathcal{S}_{k,\mathbf{t}}$ ein Spline der Ordnung $k-1$ bezüglich derselben Knotenfolge, d.h.

$$D\mathcal{S}_{k,\mathbf{t}} = \mathcal{S}_{k-1,\mathbf{t}},$$

wobei D den Differentiationsoperator bezeichnet. Gilt $s(t) = \sum_{j=1}^{n} B_{j,k}(t)\alpha_j$, so besitzt s' die Darstellung

$$s'(t) = \sum_{j=2}^{n} B_{j,k-1}(t)\alpha'_j \tag{2.7}$$

$$\text{mit} \quad \alpha'_j := \frac{\alpha_j - \alpha_{j-1}}{\frac{t_{j+k-1} - t_j}{k-1}}. \tag{2.8}$$

Bemerkung 5
Gilt $t_j = t_{j+k-1}$, so ist $B_{j,k}(t_j-) \neq B_{j,k}(t_j+)$. Damit ist $B_{j,k}$ in t_j nicht differenzierbar und somit auch $s \in \mathcal{S}_{k,\mathbf{t}}$ als Linearkombination von B-Splines i.a. nicht differenzierbar.

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

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

mit rekursiv bestimmbaren Splinekoeffizienten $\alpha_j^{(r)}$. Wir stellen diese Ableitung nach [Kun89] in Vektorform dar.

Bezeichnungen:

$$\boldsymbol{\alpha}^{(0)} := \boldsymbol{\alpha} \in \mathbb{R}^n \tag{2.10}$$

$$\boldsymbol{\alpha}^{(r)} := \left(\alpha_{r+1}^{(r)}, \ldots, \alpha_n^{(r)}\right)^T \in \mathbb{R}^{n-r} \tag{2.11}$$

Splinekoeffizienten von $s^{(r)}$

$$\boldsymbol{\beta}_r(t) := \left(B_{r+1,k-r}(t), \ldots, B_{n,k-r}(t)\right)^T \in \mathbb{R}^{n-r} \tag{2.12}$$

$$\mathbf{H}_r := (k-r)\begin{bmatrix} \frac{1}{t_{k+1}-t_{r+1}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{t_{n+k-r}-t_n} \end{bmatrix} \in \mathbb{R}^{n-r,n-r} \tag{2.13}$$

$$\mathbf{L}_r := \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & -1 & 1 & \\ & & & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n-r,n-r+1} \tag{2.14}$$

Der Spline $s^{(r)}$ besitzt dann die Darstelllung

$$s^{(r)}(t) = \boldsymbol{\beta}_r(t)^T \boldsymbol{\alpha}^{(r)}.$$

Mit (2.13), (2.14) und (2.8) erhält man

$$\begin{aligned} \boldsymbol{\alpha}^{(r)} &= \mathbf{H}_r \mathbf{L}_r \boldsymbol{\alpha}^{(r-1)} \\ &= \mathbf{H}_r \mathbf{L}_r \ldots \mathbf{H}_1 \mathbf{L}_1 \boldsymbol{\alpha} \\ &= \mathbf{D}_r \boldsymbol{\alpha} \\ \text{mit} \qquad \mathbf{D}_r &:= \mathbf{H}_r \mathbf{L}_r \ldots \mathbf{H}_1 \mathbf{L}_1 \in \mathbb{R}^{n-r,n}. \end{aligned}$$

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

$$s^{(r)}(t) = \boldsymbol{\beta}_r(t)^T \mathbf{D}_r \boldsymbol{\alpha}. \tag{2.15}$$

3 Lineare Quadratmittelprobleme

3.1 Nichtrestringierte lineare Quadratmittelprobleme

Gegeben sei das lineare Quadratmittelproblem

Problem LS:

$$\min \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\| \,:\, \mathbf{x} \in \mathbb{R}^n\} \tag{3.1}$$

mit gegebenem $\mathbf{A} \in \mathbb{R}^{m,n}$, $\mathbf{b} \in \mathbb{R}^m$ und $m \geq n$.
Zur Untersuchung des Problems LS betrachten wir die Singulärwertzerlegung von $\mathbf{A}$.

Satz 3 (Singulärwertzerlegung)
Sei $\mathbf{A} \in \mathbb{R}^{m,n}$. Dann folgt:
Es existieren orthogonale Matrizen $\mathbf{U} \in \mathbb{R}^{m,m}$, $\mathbf{V} \in \mathbb{R}^{n,n}$ und eine Diagonalmatrix

$$\mathbf{\Sigma} = \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right) \in \mathbb{R}^{m,n} \qquad r = Rang(\mathbf{A})$$

mit $\sigma_1 \geq \ldots \geq \sigma_r > 0$, so daß $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Die Zahlen $\sigma_1, \ldots, \sigma_r$ sind durch $\mathbf{A}$ eindeutig festgelegt und heißen Singulärwerte von $\mathbf{A}$.

Die Anwendung der Singulärwertzerlegung auf das Problem LS ergibt:

$$\begin{aligned} \mathbf{r} &= \mathbf{b} - \mathbf{A}\mathbf{x} \\ &= \mathbf{U}\mathbf{U}^T\mathbf{b} - \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{x} \end{aligned}$$

Unter Beachtung der Orthogonalinvarianz der Euklidischen Norm folgt:

$$\begin{aligned} \|\mathbf{r}\|^2 &= \left\|\mathbf{U}^T\mathbf{b} - \mathbf{\Sigma}\mathbf{V}^T\mathbf{x}\right\|^2 \\ &= \sum_{i=1}^{r} \left(\mathbf{u}_i^T\mathbf{b} - \sigma_i \mathbf{v}_i^T\mathbf{x}\right)^2 + \sum_{i=r+1}^{m} \left(\mathbf{u}_i^T\mathbf{b}\right)^2 \end{aligned}$$

$$\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \qquad \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$$

Der erste Summand ist stets nichtnegativ und verschwindet genau dann, wenn

$$\mathbf{x} = \sum_{i=1}^{r} \left(\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\right) \mathbf{v}_i.$$

Definition 5 (Moore-Penrose-Inverse)
Sei $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ die Singulärwertzerlegung von $\mathbf{A} \in \mathbb{R}^{m,n}$ und $\mathbf{\Sigma}^+ \in \mathbb{R}^{n,m}$ definiert durch

$$\mathbf{\Sigma}^+ := \left(\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \ddots & & 0 \\ & & \frac{1}{\sigma_r} & \\ \hline & 0 & & 0 \end{array}\right)$$

Dann heißt $\mathbf{A}^+ := \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^T \in \mathbb{R}^{n,m}$ Moore-Penrose-Inverse von $\mathbf{A}$.

Damit läßt sich die eindeutige Minimum-Norm-Lösung $\mathbf{x}^N$ durch $\mathbf{x}^N = \mathbf{A}^+\mathbf{b}$ angeben.

Wir geben jetzt noch eine wichtige Eigenschaft der Moore-Penrose-Inverse an.

Satz 4 (Moore-Penrose-Bedingungen)
Die Matrix $\mathbf{A}^+$ ist durch die folgenden Bedingungen eindeutig festgelegt:

$$\begin{array}{llll} (P1) & \mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} & (P2) & \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \\ (P3) & (\mathbf{A}\mathbf{A}^+)^T = \mathbf{A}\mathbf{A}^+ & (P4) & (\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^+\mathbf{A} \end{array}$$

Bemerkung 6
Matrizen, welche nur einen Teil der Moore-Penrose-Bedingungen erfüllen, werden als verallgemeinerte Inverse *bezeichnet und durch die erfüllten Moore-Penrose-Bedingungen gekennzeichnet. Eine Matrix $\mathbf{A}^-$, welche die Bedingungen (P1) und (P3) erfüllt, heißt auch symmetrische verallgemeinerte Inverse.*

Obwohl Moore-Penrose-Inversen sehr oft bei der theoretischen Untersuchung von Algorithmen benutzt werden, ist es oftmals billiger, bei der numerischen Implementierung mit verallgemeinerten Inversen zu arbeiten, welche nur die für den Algorithmus notwendigen Moore-Penrose-Bedingungen erfüllen (siehe auch [GP73] und [Kau75]).

Wir werden das Problem LS nicht mittels einer Singulärwertzerlegung, sondern mit einer (billigeren) QR-Faktorisierung lösen.

3.1.1 Lösung von Problem LS mittels QR-Faktorisierung

Alle in unseren Algorithmen auftretenden Matrizen haben vollen Rang $r = n$, was wir nötigenfalls durch Regularisierung in den entsprechenden nichtlinearen Quadratmittelproblemen erreichen können. Wir spezialisieren daher die QR-Faktorisierung auf diesen Fall.

Sei $\mathbf{A} \in \mathbb{R}^{m,n}$, $m \geq n$ eine Matrix mit Rang $r = n$. Dann existieren

$$\begin{array}{lcl} \mathbf{Q} & \in & \mathbb{R}^{m,m} \quad \text{orthogonal} \\ \mathbf{Q} & = & (\mathbf{Q}_1, \mathbf{Q}_2) \qquad \mathbf{Q}_1 \in \mathbb{R}^{m,n} \quad \mathbf{Q}_2 \in \mathbb{R}^{m,m-n} \\ \hat{\mathbf{R}} & = & \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix} \qquad \hat{\mathbf{R}} \in \mathbb{R}^{m,n}, \mathbf{R} \in \mathbb{R}^{n,n} \quad \text{reguläre obere Dreiecksmatrix} \\ \mathbf{P} & \in & \mathbb{R}^{n,n} \quad \text{Permutationsmatrix,} \end{array}$$

so daß

$$\begin{aligned} \mathbf{Q}^T\mathbf{A}\mathbf{P} &= \hat{\mathbf{R}} \\ \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \mathbf{A}\mathbf{P} &= \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \end{aligned}$$

Bemerkung 7

- *Im Vollrangfall ist die Permutationsmatrix $\mathbf{P}$ eigentlich nicht nötig. Sie wird jedoch zur Verbesserung der numerischen Stabilität der QR-Faktorisierung verwendet.*

- *Für vollbesetzte Matrizen $\mathbf{A} \in \mathbb{R}^{m,n}$ erzeugen wir $\mathbf{Q}$ durch eine Folge von Householder-Transformationen. Wir verwenden einen Algorithmus mit Spaltenpivotisierung (siehe [GvL89], [KS88]). Die Implementierung lehnt sich an die Prozedur* QRDC *aus* LINPACK *[DBMS79] an.*

Unter Benutzung der QR-Faktorisierung von $\mathbf{A}$ können wir nun einen Algorithmus zur Lösung von Problem LS angeben. Es gilt:

$$\begin{aligned}
\|\mathbf{r}\|^2 = \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 &= \left\|\mathbf{Q}^T\left(\mathbf{b} - \mathbf{A}\mathbf{x}\right)\right\|^2 \\
&= \left\|\mathbf{Q}^T\mathbf{b} - \left(\mathbf{Q}^T\mathbf{A}\mathbf{P}\right)\left(\mathbf{P}^T\mathbf{x}\right)\right\|^2 \\
&= \left\|\begin{pmatrix} \mathbf{Q}_1^T\mathbf{b} \\ \mathbf{Q}_2^T\mathbf{b} \end{pmatrix} - \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}\left(\mathbf{P}^T\mathbf{x}\right)\right\|^2
\end{aligned}$$

Wir definieren $\mathbf{y} := \mathbf{P}^T\mathbf{x} \in \mathbb{R}^n$, $\mathbf{c}_1 := \mathbf{Q}_1^T\mathbf{b} \in \mathbb{R}^n$, $\mathbf{c}_2 := \mathbf{Q}_2^T\mathbf{b} \in \mathbb{R}^{m-n}$ und erhalten

$$\|\mathbf{r}\|^2 = \|\mathbf{c}_1 - \mathbf{R}\mathbf{y}\|^2 + \|\mathbf{c}_2\|^2$$

$\mathbf{y}$ können wir nun aus dem oberen Dreieckssystem $\mathbf{R}\mathbf{y} = \mathbf{c}_1$ bestimmen.

Besitzt die zu transformierende Matrix eine spezielle Gestalt, so lassen sich Householder-Spiegelungen nur unter besonderen Vorkehrungen effektiv einsetzen (fill in). Wir verwenden aus diesem Grund (explizite) Givens-Drehungen. Zudem besitzen die Givens-Drehungen bei der Faktorisierung einer Matrix die deutlich besseren Fehlerkumulationskonstanten (siehe [KS88]).

3.1.2 Lineare Quadratmittelprobleme mit Bandstruktur

Bei der Approximation bzw. beim Glätten mit B-Splines treten Matrizen der folgenden Form auf:

$$(a)\quad \begin{pmatrix}
x & x & x & x & & \\
x & x & x & x & & \\
x & x & x & x & & \\
 & x & x & x & x & \\
 & x & x & x & x & \\
 & x & x & x & x & \\
 & x & x & x & x & \\
 & & x & x & x & x \\
 & & x & x & x & x \\
 & & x & x & x & x
\end{pmatrix}
\qquad
(b)\quad \begin{pmatrix}
x & x & x & x & & \\
 & x & x & x & x & \\
 & & x & x & x & x \\
 & & & x & x & x \\
 & & & & x & x \\
 & & & & & x \\
x & x & x & & & \\
 & x & x & x & & \\
 & & x & x & x & \\
 & & & x & x & x
\end{pmatrix}$$

Für eine Matrix $\mathbf{A} \in \mathbb{R}^{m,n}$ definieren wir

$$\begin{aligned}
f_i &:= \min\{j : a_{i,j} \neq 0\} \\
l_i &:= \max\{j : a_{i,j} \neq 0\}.
\end{aligned}$$

Die Zahlen f_i und l_i sind die Indizes des ersten bzw. letzten Nichtnullelementes in der i-ten Zeile.

Definition 6
Die Matrix $\mathbf{A} \in \mathbb{R}^{m,n}$ hat die (Zeilen) Bandbreite nbands, wenn

$$nbands = \max_{i=1,\ldots,m} w_i \qquad w_i := (l_i - f_i + 1)$$

Die Matrizen in (a) und (b) sind beide von der Bandbreite 4.

Definition 7
Eine Matrix ist in Standardform, wenn der Index des am weitesten rechts stehenden Nichtnullelementes l_i jeder Zeile eine nichtfallende Funktion der Zeilennummer ist.

Bemerkung 8

- *Die Matrix in (a) ist offenbar in Standardform, während (b) nicht in Standardform ist.*
- *Durch Zeilenvertauschung kann jede Matrix in Standardform gebracht werden. Wir erhalten z.B. aus (b)*

$$\begin{pmatrix} x & x & x & & & \\ & x & x & x & & \\ x & x & x & x & & \\ & & x & x & x & \\ & x & x & x & x & \\ & & & x & x & x \\ & & x & x & x & x \\ & & & x & x & x \\ & & & & x & x \\ & & & & & x \end{pmatrix}$$

eine Matrix in Standardform, indem wir, bis auf die ersten und letzten Zeilen, abwechselnd je eine Zeile der Teilmatrizen nehmen.

Betrachten wir nun das Problem LS

$$\min\{\|\mathbf{y} - \mathbf{A}\mathbf{x}\| \, : \, \mathbf{x} \in \mathbb{R}^n\}$$

mit einer Bandmatrix $\mathbf{A} \in \mathbb{R}^{m,n}$ der Bandbreite *nbands*. $\mathbf{A}$ habe Vollrang n.

Grundalgorithmus

S1: Transformation des $m \times n$ Problems in ein äquivalentes $n \times n$ Problem

S1.1: Berechnung der Givens-Faktorisierung der Matrix $\mathbf{A}$

$$\begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \mathbf{A} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \qquad \begin{array}{ll} \mathbf{R} \in \mathbb{R}^{n,n} & \text{reguläre obere Dreiecksmatrix} \\ (\mathbf{Q}_1, \mathbf{Q}_2) \in \mathbb{R}^{m,m} & \text{orthogonal} \end{array}$$

S1.2: Transformation der rechten Seite $\mathbf{b}$

$$\mathbf{c}_1 = \mathbf{Q}_1^T \mathbf{b} \qquad \mathbf{c}_1 \in \mathbb{R}^n$$

S2: Lösung des Gleichungssystems $\mathbf{R}\mathbf{x} = \mathbf{c}_1$ mit regulärer oberer Dreiecksmatrix $\mathbf{R}$

$$\mathbf{x} = \mathbf{R}^{-1}\mathbf{c}_1$$

Bemerkung 9

- *Die Schritte S1.1 und S1.2 können bei Bedarf zusammen realisiert werden. Die Speicherung und spätere Rekonstruktion der Drehungsparameter kann dann entfallen.*
- *Wenn* $\mathbf{A}$ *eine Blockbandmatrix in Standardform der Bandbreite nbands ist, hat die obere Dreiecksmatrix* $\mathbf{R}$ *diesselbe Bandbreite ([Cox81]).*
- *Bei vorliegender Rangdefizienz muß Schritt S2 geeignet modifiziert werden (siehe [Cox89]). Im Falle total positiver Matrizen, wie sie bei der Approximation mit B-Splines auftreten ([Cox82]), ist keine Pivotisierung in Schritt S1.1 nötig (siehe [dBP77]).*

Zur Berechnung der Givens-Faktorisierung der Matrix $\mathbf{A}$ verwenden wir einen Algorithmus zur zeilenweisen Givens-Faktorisierung nach [Cox81].

Algorithmus zur zeilenweise Givens-Faktorisierung
Arbeitsvektor $(\mathbf{u}, v)$ zur Speicherung der i-ten Zeile von $(\mathbf{A}, \mathbf{b})$

S1: Initialisierung

Setze $\mathbf{R} \in \mathbb{R}^{n,n}$, $\mathbf{c} \in \mathbb{R}^n$, $||\mathbf{r}||^2$ Null

S2: Berechnung

for i:=1 to m do

S2.1: Stelle $(\mathbf{a}_i^T, b_i)$ auf $(\mathbf{u}, v)$ bereit
S2.2: { Annuliere $\mathbf{u}$ durch Givens-Drehungen mit den Zeilen von $\mathbf{R}$ }
for j:=f_i to l_i do
if $u_j \neq 0$ then
S2.2.1: setup_GIVENS_rotation(r_{jj},u_j)
S2.2.2: apply_GIVENS_rotation(nbands,r_{jj},u_j)
S2.2.3: apply_GIVENS_rotation(1,c_j,v)
S2.3: Aktualisiere das Residium
$||\mathbf{r}||^2 := ||\mathbf{r}||^2 + v^2$

Bemerkung 10

- *Wenn $\mathbf{A} \in \mathbb{R}^{m,n}$ eine Bandmatrix der Bandbreite $nbands$ ist und das Problem für mehrere rechte Seiten $\mathbf{b}$ zu lösen ist, so müssen in Schritt S2.2.1 zusätzlich Informationen über die Drehungsparameter gespeichert werden. Diese können in einer Matrix $\mathbf{Q}$ der Dimension $m \times nbands$ aufgehoben werden. Die Schritte S2.2.3 und S2.3 können dann entfallen.*

- *Der Aufwand beträgt nach [Cox81] (siehe auch [Bjö 90]) für eine Matrix $\mathbf{A} \in \mathbb{R}^{m,n}$*

Matrix-Struktur	*Standardform*	*Nicht-Standardform*
volle Matrix	$\sim 2m\,n^2$	$\sim 2m\,n^2$
Bandmatrix	$\sim 2m\,nbands^2$	$\sim 2m\,n\,nbands$

flops.

- *Die Wirkungsweise des Algorithmus findet man auch in [Cox81] und [Kun89] beschrieben.*

3.2 Lineare Quadratmittelprobleme mit linearen Ungleichheitsnebenbedingungen

Bei der Lösung der später betrachteten nichtlinearen Quadratmittelprobleme muß eine Folge von linearen Quadratmittelproblemen mit linearen Ungleichheitsnebenbedingungen gelöst werden. Wir betrachten daher jetzt

Problem LSI:
$$\min\{||\mathbf{Ax} - \mathbf{b}|| \,:\, \mathbf{Cx} \geq \mathbf{h},\, \mathbf{x} \in \mathbb{R}^n\} \tag{3.2}$$

Problem LDP (Least Distance Programming):

$$\min\{||\mathbf{x}|| \,:\, \mathbf{Cx} \geq \mathbf{h},\, \mathbf{x} \in \mathbb{R}^n\} \tag{3.3}$$

Problem NNLS (Nonnegative Least Squares):

$$\min\{||\mathbf{Ax}-\mathbf{b}|| \,:\, \mathbf{x}\geq\mathbf{0},\, \mathbf{x}\in\mathbb{R}^n\} \tag{3.4}$$

mit Matrizen $\mathbf{A}\in\mathbb{R}^{m,n}$, $\mathbf{C}\in\mathbb{R}^{ncstr,n}$ und Vektoren $\mathbf{h}\in\mathbb{R}^{ncstr}$, $\mathbf{b}\in\mathbb{R}^m$.

Das allgemeine Problem LSI mit einer Vollrangmatrix $\mathbf{A}$ wird durch eine orthogonale Transformation in das Problem LDP überführt. Das Problem LDP schließlich wird in das Problem NNLS transformiert, welches mit einer aktiven Mengenstrategie gelöst wird (siehe [LH74, Chapter 23]).

Die Prozeduren zur Lösung von LDP und NNLS entnehmen wir ebenfalls den gleichnamigen FORTRAN-Programmen in [LH74]. Wir geben daher nur die Transformation von Problem LSI in Problem LDP an.

3.2.1 Transformation von Problem LSI in Problem LDP

Sei $\mathbf{A}\in\mathbb{R}^{m,n}$ eine Matrix mit Vollrang n, $\mathbf{C}\in\mathbb{R}^{ncstr,n}$, $\mathbf{b}\in\mathbb{R}^m$ und $\mathbf{h}\in\mathbb{R}^{ncstr}$.

Problem LSI:

$$\min\{||\mathbf{Ax}-\mathbf{b}|| \,:\, \mathbf{Cx}\geq\mathbf{h},\, \mathbf{x}\in\mathbb{R}^n\} \tag{3.5}$$

Sei $\mathbf{AP}=\mathbf{Q\hat{R}}$ die QR-Faktorisierung von $\mathbf{A}$, wie in Abschnitt 3.1.1 beschrieben.

$$\begin{array}{lcll}
\mathbf{Q} & \in & \mathbb{R}^{m,m} & \text{orthogonal} \\
\mathbf{Q} & = & (\mathbf{Q}_1,\mathbf{Q}_2) & \quad \mathbf{Q}_1\in\mathbb{R}^{m,n} \quad \mathbf{Q}_2\in\mathbb{R}^{m,m-n} \\
\hat{\mathbf{R}} & = & \begin{pmatrix}\mathbf{R}\\ \mathbf{0}\end{pmatrix} & \quad \hat{\mathbf{R}}\in\mathbb{R}^{m,n}, \mathbf{R}\in\mathbb{R}^{n,n} \quad \text{reguläre obere Dreiecksmatrix} \\
\mathbf{P} & \in & \mathbb{R}^{n,n} & \text{Permutationsmatrix.}
\end{array}$$

Dann gilt

$$\begin{aligned}
\mathbf{Q}^T\mathbf{AP} &= \hat{\mathbf{R}} \\
\begin{bmatrix}\mathbf{Q}_1^T\\ \mathbf{Q}_2^T\end{bmatrix}\mathbf{AP} &= \begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix}
\end{aligned}$$

Wir erhalten

$$\begin{aligned}
\varphi(\mathbf{x}) := \frac{1}{2}||\mathbf{Ax}-\mathbf{b}||^2 &= \frac{1}{2}\left\|\left(\mathbf{Q}^T\mathbf{AP}\right)\left(\mathbf{P}^T\mathbf{x}\right)-\mathbf{Q}^T\mathbf{b}\right\|^2 \\
&= \frac{1}{2}\left\|\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix}\left(\mathbf{P}^T\mathbf{x}\right)-\begin{pmatrix}\mathbf{Q}_1^T\mathbf{b}\\ \mathbf{Q}_2^T\mathbf{b}\end{pmatrix}\right\|^2
\end{aligned}$$

Wir definieren wieder $\mathbf{y}:=\mathbf{P}^T\mathbf{x}\in\mathbb{R}^n$, $\mathbf{c}_1:=\mathbf{Q}_1^T\mathbf{b}\in\mathbb{R}^n$, $\mathbf{c}_2:=\mathbf{Q}_2^T\mathbf{b}\in\mathbb{R}^{m-n}$ und erhalten

$$\varphi(\mathbf{x}) = \frac{1}{2}||\mathbf{Ry}-\mathbf{c}_1||^2+\frac{1}{2}||\mathbf{c}_2||^2.$$

Wir führen eine erneute Variablentransformation aus

$$\mathbf{z} := \mathbf{Ry}-\mathbf{c}_1 \quad \in\mathbb{R}^n.$$

Damit erhalten wir aus

$$\begin{aligned}
\varphi(\mathbf{x}) &:= \frac{1}{2}||\mathbf{Ax}-\mathbf{b}||^2 \quad \longrightarrow \text{Min!} \\
\tilde{\varphi}(\mathbf{z}) &:= \frac{1}{2}||\mathbf{z}||^2+\frac{1}{2}||\mathbf{c}_2||^2 \quad \longrightarrow \text{Min!}
\end{aligned}$$

Transformation der Nebenbedingungen

$$\begin{array}{rrcl} & \mathbf{Cx} & \geq & \mathbf{h} \\ \mathbf{y} = \mathbf{P}^T\mathbf{x} \rightsquigarrow & \mathbf{CPy} & \geq & \mathbf{h} \\ \mathbf{z} = \mathbf{Ry} - \mathbf{c}_1 \rightsquigarrow & \mathbf{CPR}^{-1}(\mathbf{z} + \mathbf{c}_1) & \geq & \mathbf{h} \\ & \mathbf{CPR}^{-1}\mathbf{z} & \geq & \mathbf{h} - \mathbf{CPR}^{-1}\mathbf{c}_1 \end{array}$$

Das Problem LSI ist damit äquivalent zu folgendem Least Distance Problem

$$\begin{array}{rrcl} & \|\mathbf{z}\| & \longrightarrow & \text{Min!} \\ \text{mit} & \mathbf{CPR}^{-1}\mathbf{z} & \geq & \mathbf{h} - \mathbf{CPR}^{-1}\mathbf{c}_1 \end{array}$$

Zusammenfassend geben wir noch einmal den Algorithmus zur Lösung von Problem LSI an.

Algorithmus LSI

{ Lösung des Problems $\|\mathbf{Ax} - \mathbf{b}\| \longrightarrow$ Min! mit $\mathbf{Cx} \geq \mathbf{h}$ mit Vollrangmatrix $\mathbf{A}$ }

S1: Berechne QR-Faktorisierung von $\mathbf{A}$

$$\begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \mathbf{A} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$$

S2: Berechne $\mathbf{c}_1 := \mathbf{Q}_1^T\mathbf{b} \quad \in \mathbb{R}^n$

S3: Berechne

$$\begin{array}{rcll} \tilde{\mathbf{C}} & := & \mathbf{CPR}^{-1} & \in \mathbb{R}^{ncstr,n} \\ \tilde{\mathbf{h}} & := & \mathbf{h} - \mathbf{CPR}^{-1}\mathbf{c}_1 & \in \mathbb{R}^{ncstr} \end{array}$$

S4: Löse das Least Distance Problem

$$\begin{array}{rrcl} & \|\mathbf{z}\| & \longrightarrow & \text{Min!} \\ \text{mit} & \mathbf{CPR}^{-1}\mathbf{z} & \geq & \mathbf{h} - \mathbf{CPR}^{-1}\mathbf{c}_1 \end{array}$$

S5: Bilde $\mathbf{z} + \mathbf{Q}_1^T\mathbf{b} = \mathbf{z} + \mathbf{c}_1$

S6: Löse das obere Dreieckssystem

$$\mathbf{Ry} = \mathbf{z} + \mathbf{c}_1$$

S7: Bilde $\mathbf{x} := \mathbf{Py}$

Bemerkung 11

- *Die QR-Faktorisierung in Schritt S1 erfolgt mit Householder-Spiegelungen mit Spaltenpivotisierung.*

- *Vor der Lösung der oberen Dreickssysteme in S3 und S6 erfolgt eine Schätzung der Konditionszahl von* $\mathbf{R}$. *Der verwendete Algorithmus folgt Algorithmus A3.3.1* CONDEST *in [DS83] (siehe auch [CMSW79]). Wenn* $cond(\mathbf{R}) > \frac{1}{\sqrt{macheps}}$ *bricht der Algorithmus ab. Die Matrix* $\mathbf{A}$ *ist dann (fast) rangdefizient. Der aufrufende Algorithmus zur Lösung nichtlinearer Quadratmittelprobleme kann dann Maßnahmen zur Regularisierung (Levenberg-Marquardt o.ä.) treffen.*

- *In [Bjö 90] wird die Transformation von Problem LSI mit rangdefizienter Matrix* $\mathbf{A}$ *in ein verallgemeinertes Least Distance Problem betrachtet. Diese Transformation beruht auf einer kompletten orthogonalen Faktorisierung*

$$\mathbf{Q}^T\mathbf{A}\mathbf{P}\mathbf{V} = \begin{bmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$\mathbf{T} \in \mathbb{R}^{r,r}$ *obere reguläre Dreiecksmatrix,* $\mathbf{V} \in \mathbb{R}^{n,n}$ *orthogonal (siehe [GvL89]) und führt auf ein verallgemeinertes Least Distance Problem*

$$\begin{array}{rcl} ||\mathbf{x}_1|| & \longrightarrow & \mathit{Min!} \\ \mathit{mit} \quad \mathbf{C}_1\mathbf{x}_1 + \mathbf{C}_2\mathbf{x}_2 & \geq & \mathbf{h}. \end{array}$$

4 Nichtlineare Quadratmittelprobleme

Die in dieser Arbeit auftretenden Optimierungsprobleme sind entweder *freie Quadratmittelprobleme*

$$\min \left\{ \frac{1}{2} \|\mathbf{F}(\mathbf{x})\|^2 \ : \ \mathbf{x} \in \mathbb{R}^n \right\} \tag{4.1}$$

mit $\mathbf{F} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $m \geq n$ oder *restringierte Quadratmittelprobleme*

$$\min \left\{ \frac{1}{2} \|\mathbf{F}_1(\mathbf{x})\|^2 \ : \ \mathbf{F}_2(\mathbf{x}) = \mathbf{0},\, \mathbf{F}_3(\mathbf{x}) \geq \mathbf{0},\, \mathbf{x} \in \mathbb{R}^n \right\} \tag{4.2}$$

mit $\mathbf{F}_i : \mathbb{R}^n \longrightarrow \mathbb{R}^{m_i}$. Wir können uns dabei hier sogar auf *nichtlineare Quadratmittelprobleme mit linearen Ungleichheitsnebenbedingungen* beschränken.

$$\min \left\{ \frac{1}{2} \|\mathbf{F}(\mathbf{x})\|^2 \ : \ \mathbf{C}\mathbf{x} \geq \mathbf{h},\, \mathbf{x} \in \mathbb{R}^n \right\} \tag{4.3}$$

mit $\mathbf{F} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $\mathbf{C} \in \mathbb{R}^{ncstr,n}$, $\mathbf{h} \in \mathbb{R}^{ncstr}$.

Die zur Lösung solcher nichtlinearer Quadratmittelprobleme verwendeteten Algorithmen sind grundsätzlich iterativ und von der Form

$$\mathbf{x}^{k+1} := \mathbf{x}^k + \mathbf{s}^k,$$

wobei $\mathbf{x}^k$ die aktuelle Iterierte bezeichnet und $\mathbf{s}^k$ der Schritt (*step*) zur nächsten Iterierten $\mathbf{x}^{k+1}$ ist.

4.1 Freie Quadratmittelprobleme

4.1.1 Das Gauß-Newton-Prinzip

Gegeben sei eine Funktion $\mathbf{F} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_m(\mathbf{x}))^T$ mit $m \geq n$. Wir betrachten das nichtlineare Quadratmittelproblem

$$\min \left\{ \sum_{i=1}^{m} [F_i(\mathbf{x})]^2 \ : \ \mathbf{x} \in \mathbb{R}^n \right\} \tag{4.4}$$

Der einfacheren Beschreibung wegen definieren wir durch die Festlegung

$$g(\mathbf{x}) \ := \ \frac{1}{2} \sum_{i=1}^{m} [F_i(\mathbf{x})]^2 = \frac{1}{2} \|\mathbf{F}(\mathbf{x})\|^2 \ \forall\, \mathbf{x} \in \mathbb{R}^n \tag{4.5}$$

eine Funktion $g : \mathbb{R}^n \longrightarrow \mathbb{R}$, mit der (4.4) in der Form

$$\min \{ g(\mathbf{x}) \ : \ \mathbf{x} \in \mathbb{R}^n \} \tag{4.6}$$

geschrieben werden kann. Falls die Funktionen F_i, $\quad i = 1, \ldots, m$, genügend glatt sind, erhalten wir

$$\begin{aligned} \nabla g(\mathbf{x}) &= \mathbf{F}'(\mathbf{x})^T \mathbf{F}(\mathbf{x}) \\ \nabla^2 g(\mathbf{x}) &= \mathbf{F}'(\mathbf{x})^T \mathbf{F}'(\mathbf{x}) + S(\mathbf{x}) \\ \text{mit} \quad S(\mathbf{x}) &:= \mathbf{F}(\mathbf{x}) \circ \mathbf{F}''(\mathbf{x}) = \sum_{i=1}^{m} F_i(\mathbf{x}) F_i''(\mathbf{x}). \end{aligned}$$

Wir bilden im Punkt $\mathbf{x} = \mathbf{x}^k$ eine quadratische Modellfunktion

$$g(\mathbf{x}+\mathbf{s}) \approx \mu_N(\mathbf{x}+\mathbf{s}) \quad := \quad g(\mathbf{x}) + \nabla g(\mathbf{x})^T \mathbf{s} + \frac{1}{2}\mathbf{s}^T \nabla^2 g(\mathbf{x})\mathbf{s}.$$

Eine naheliegende Möglichkeit zur Lösung von (4.6) besteht darin, in jedem Schritt $\mathbf{s}^k$ als Lösung des quadratischen Ersatzproblems

$$\min\left\{\mu_N(\mathbf{x}+\mathbf{s}) = g(\mathbf{x}) + \nabla g(\mathbf{x})^T \mathbf{s} + \frac{1}{2}\mathbf{s}^T \nabla^2 g(\mathbf{x})\mathbf{s} \;:\; \mathbf{s} \in \mathbb{R}^n\right\} \tag{4.7}$$

zu bestimmen. Dies führt auf das Newton-Verfahren zur Lösung von (4.6). Aus Aufwandsgründen muß jedoch die Berechnung aller partiellen Ableitungen zweiter Ordnung $\frac{\partial^2 F_i}{\partial x_j \partial x_k}$ $i = 1, \ldots, m, j = 1, \ldots, n, k = 1, \ldots n$ entfallen.

Eine weitere Möglichkeit besteht darin, den Term $\mathbf{S}(\mathbf{x})$ in der Hesse-Matrix $\nabla^2 g(\mathbf{x})$ zu vernachlässigen, da dieser Term bei Problemen mit kleinen Residuen bzw. bei schwach nichtlinearen Problemen klein ist. Dieses Vorgehen führt auf das Gauß-Newton-Modell

$$\min\left\{\mu_{GN}(\mathbf{x}+\mathbf{s}) = g(\mathbf{x}) + \nabla g(\mathbf{x})^T \mathbf{s} + \frac{1}{2}\mathbf{s}^T \mathbf{H}(\mathbf{x})\mathbf{s} \;:\; \mathbf{s} \in \mathbb{R}^n\right\} \tag{4.8}$$

mit $\mathbf{H} := \mathbf{F}'(\mathbf{x})^T \mathbf{F}'(\mathbf{x})$. Es ist äquivalent mit einer Ersetzung der Funktion $\mathbf{F}$ im Punkt $\mathbf{x}^k$ durch eine linearisierte Funktion

$$\tilde{\mathbf{F}}_k \quad := \quad \mathbf{F}(\mathbf{x}^k) + \mathbf{F}'(\mathbf{x}^k)\,(\mathbf{x} - \mathbf{x}^k).$$

Äquivalent zu (4.8) ist

$$\min\left\{\mu_{GN}(\mathbf{x}+\mathbf{s}) = \frac{1}{2}\left\|\mathbf{F} + \mathbf{F}'\mathbf{s}\right\|^2 \;:\; \mathbf{s} \in \mathbb{R}^n\right\}. \tag{4.9}$$

Wenn $\mathbf{F}'(\mathbf{x}^k)$ Vollrang hat, so besitzt das Problem (4.9) die eindeutige Lösung

$$\mathbf{s}^k \quad := \quad -\left[\mathbf{F}'(\mathbf{x}^k)^T \mathbf{F}'(\mathbf{x}^k)\right]^{-1} \mathbf{F}'(\mathbf{x}^k)^T \mathbf{F}(\mathbf{x}^k) \tag{4.10}$$

und wir erhalten die Iterationsvorschrift

$$\mathbf{x}^{k+1} \quad := \quad \mathbf{x}^k + \mathbf{s}^k. \tag{4.11}$$

Diese Iterationsvorschrift bezeichnet man als Gauß-Newton-Verfahren.

Aus den zahlreichen Konvergenzsätzen für das Gauß-Newton-Verfahren zitieren wir eine affin-invariante Version aus [DH91].

Satz 5 (Lokale Konvergenz des Gauß-Newton-Verfahrens)
Sei $D \subset \mathbb{R}^n$ offen, konvex und $\mathbf{F} : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $m \geq n$, eine stetig differenzierbare Abbildung, deren Jacobi-Matrix $\mathbf{F}'(\mathbf{x})$ für alle $\mathbf{x} \in D$ vollen Rang n habe. Es existiere eine Lösung $\mathbf{x}^* \in D$ des zugehörigen nichtlinearen Quadratmittelproblems $\|\mathbf{F}(\mathbf{x})\| \longrightarrow \min$. Ferner gebe es Konstanten $\omega > 0$ und $0 \leq \kappa^* < 1$, so daß

$$\left\|\mathbf{F}'(\mathbf{x})^+ \left[\mathbf{F}'(\mathbf{x}+s\mathbf{v}) - \mathbf{F}'(\mathbf{x})\right]\mathbf{v}\right\| \quad \leq \quad s\omega\,\|\mathbf{v}\|^2$$

für alle $s \in [0,1]$, $\mathbf{x} \in D$ und $\mathbf{v} \in \mathbb{R}^n$ mit $\mathbf{x} + \mathbf{v} \in D$, und

$$\left\|\left[\mathbf{F}'(\mathbf{x})^+ - \mathbf{F}'(\mathbf{x}^*)^+\right]\mathbf{F}'(\mathbf{x}^*)\right\| \quad \leq \quad \kappa^*\,\|\mathbf{x} - \mathbf{x}^*\|$$

für alle $\mathbf{x} \in D$. *Falls dann für einen gegebenen Startwert* $\mathbf{x}^0 \in D$

$$\rho := \left\|\mathbf{x}^0 - \mathbf{x}^*\right\| \quad < \quad 2\,\frac{1-\kappa^*}{\omega} =: \sigma$$

gilt, so bleibt die durch das Gauß-Newton-Verfahren (4.10), (4.11) definierte Folge $\{\mathbf{x}^k\}$ *in der offenen Kugel* $K(\mathbf{x}^*, \rho)$ *und konvergiert gegen* $\mathbf{x}^*$, *d.h.*

$$\left\|\mathbf{x}^k - \mathbf{x}^*\right\| < \rho \quad \textit{für} \quad k > 0 \quad \textit{und} \quad \lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^*.$$

Die Konvergenzgeschwindigkeit läßt sich abschätzen durch

$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\| \quad \le \quad \frac{\omega}{2}\left\|\mathbf{x}^k - \mathbf{x}^*\right\|^2 + \kappa^*\left\|\mathbf{x}^k - \mathbf{x}^*\right\|.$$

Insbesondere ergibt sich quadratische Konvergenz für kompatible nichtlineare Quadratmittelprobleme ($\|\mathbf{F}(\mathbf{x}^*\| = \mathbf{0}$*). Darüber hinaus ist die Lösung* $\mathbf{x}^*$ *eindeutig in* $K(\mathbf{x}^*, \rho)$.

4.1.2 Globalisierung

Wir verwenden in unseren Algorithmen eine global konvergente Modifikation des Gauß-Newton-Verfahrens. Wir führen ein gedämpftes Gauß-Newton-Verfahren mit Armijo-Goldstein-Schrittweitenstrategie aus (siehe [Sch79]).

Gedämpftes Gauß-Newton-Verfahren mit akzeptiertem Schätzwert

S1: Wähle Startwert $\mathbf{x}^0 \in \mathbb{R}^n$, $\alpha \in (0,1)$, $\delta \in (0, \frac{1}{2})$

Setze $k := 0$; $\gamma_{-1} := 1$;

S2: Iteration

Repeat

S2.1: $\mathbf{F} := \mathbf{F}(\mathbf{x}^k)$; Berechnung des Residiums
$\mathbf{J} := \mathbf{F}'(\mathbf{x}^k)$; Berechnung der Jacobi-Matrix

S2.2: Bestimme Suchrichtung $\mathbf{p}^k (= -\mathbf{s}^k)$ aus

$$\mathbf{J}\mathbf{p}^k \cong \mathbf{F}$$

S2.3: $\tilde{\gamma}_k := \gamma_k$; $\mathbf{x}^+ := \mathbf{x}^k - \tilde{\gamma}_k \mathbf{p}^k$; $\mathbf{F}^+ := \mathbf{F}(\mathbf{x}^+)$;

S2.4: Abstiegstest
if $\frac{1}{2}\|\mathbf{F}\|^2 - \frac{1}{2}\|\mathbf{F}^+\|^2 \ge \delta\tilde{\gamma}_k \mathbf{F}^T\mathbf{J}\mathbf{p}^k$

then

$\mathbf{x}^{k+1} := \mathbf{x}^+$;

$\gamma_{k+1} := \min(\gamma_k/\alpha, 1)$;

$k := k+1$;

else

$\gamma_k := \gamma_k * \alpha$; goto S2.3;

until $(k > kmax)$ or Konvergenz

Bemerkung 12

- *Die Berechnung der Suchrichtung* $\mathbf{p}^k$ *in Schritt S2.2 kann über die Normalgleichungen* $(\mathbf{J}^T\mathbf{J})\mathbf{p}^k = \mathbf{J}^T\mathbf{F}$ *mittels Cholesky-Zerlegung erfolgen. Wir wählen jedoch die direkte Lösung mittels QR-Faktorisierung der Matrix* $\mathbf{J}$. *Dazu verwenden wir Householder-Spiegelungen mit Spaltenpivotisierung (siehe Abschnitt 3.1.1).*

- *Die analytische Berechnung der Jacobi-Matrix* $\mathbf{J}$ *gestaltet sich wegen der komplizierten Gestalt der später untersuchten Funktionen* $\mathbf{F}$ *äußerst schwierig. Wir gehen deshalb einen anderen Weg und verwenden eine Approximation* $\mathbf{A}_k$ *der Jacobi-Matrix. Die einfachste und nächstliegende Möglichkeit besteht darin, die j-te Spalte von* $\mathbf{A}_k$ *als Differenzenquotient darzustellen.*

$$\mathbf{A}_k(\mathbf{x}^k)\mathbf{e}^j = \frac{\mathbf{F}(\mathbf{x}^k + h_j^k \mathbf{e}^j) - \mathbf{F}(\mathbf{x}^k)}{h_j^k} \tag{4.12}$$

Die Berechnung einer Jacobi-Matrix kostet $n+1$ *Funktionsauswertungen von* $\mathbf{F}$*. Wir benutzen diese Approximation mit* $h_j^k := \varepsilon_1^d * \left\{ \left| x_j^k \right| + \varepsilon_2^d \right\}$*. Wir wählen dabei* ε_1^d *in der Größenordnung von* $\sqrt{macheps}$ *und* $\varepsilon_2^d \approx 10^{-10}$*.*

In einigen späteren Anwendungen (siehe Abschnitt 5) wird diese Strategie innerhalb eines bekannten analytischen Ausdrucks für $\mathbf{J}$ *angewendet. Konvergenzaussagen für sogenannte* konsistente Approximationen, *d.h.*

$$\lim_{k\to\infty} \left\| \mathbf{A}_k - \mathbf{F}'(\mathbf{x}^k) \right\| = 0,$$

findet man in [Sch79].

- *Das Abbruchkriterium konkretisieren wir wie folgt:*

Repeat ... until

$(k > kmax)$ *or* $\left(\left\|\mathbf{F}(\mathbf{x}^k)\right\| \le \varepsilon_0^a\right)$ *or* $\left(\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| \le \varepsilon_1^a \left(\left\|\mathbf{x}^k\right\| + \varepsilon_2^a\right)\right)$ *or*
$\left(\left\|\mathbf{F}'(\mathbf{x}^k)^T \mathbf{F}(\mathbf{x}^k)\right\| \le \varepsilon_3^a\right)$ *or* $\left(\left|\mathbf{F}(\mathbf{x}^k)^T \mathbf{F}'(\mathbf{x}^k)\mathbf{s}\right| \le \varepsilon_4^a\right)$

Das Verfahren bricht ab, wenn die maximale Schrittanzahl $kmax$ *erreicht ist. Die Defektschranke* ε_0^a *wird vom Nutzer vorgegeben.*

$\frac{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|}{\|\mathbf{x}^k\|}$ *gibt für Q-superlinear konvergente Verfahren den relativen Fehler* ε_1^a *der Iterierten* $\mathbf{x}^k$ *asymptotisch richtig wieder.* ε_2^a *macht den Test auch im Fall* $\left\|\mathbf{x}^k\right\| = 0$ *sinnvoll und wird etwa in der Größenordnung* 10^{-3} *gewählt.*

Der Test $\left(\left\|\mathbf{F}'(\mathbf{x}^k)^T \mathbf{F}(\mathbf{x}^k)\right\| \le \varepsilon_3^a\right)$ *stellt fest, ob ein stationärer Punkt erreicht ist.* $\left(\left|\mathbf{F}(\mathbf{x}^k)^T \mathbf{F}'(\mathbf{x}^k)\mathbf{s}\right| \le \varepsilon_4^a\right)$ *stellt schließlich fest, ob überhaupt noch eine Abstiegsrichtung* $\mathbf{s}$ *gefunden werden kann.*

Die Defektschranke ε_0^a*, die Schranke* ε_3^a *für die Norm des Gradienten sowie die Schranke* ε_4^a *sind abhängig von der Größenordnung der Funktionswerte (und von der Skalierung von* $\mathbf{x}$*) und können vom Nutzer vorgegeben werden.*

- *Der Abstiegstest sichert einen hinreichenden Abstieg der Funktion* $g(\mathbf{x}) = \frac{1}{2}\|\mathbf{F}(\mathbf{x})\|^2$ *vom Punkt* $\mathbf{x}^k$ *zum Punkt* $\mathbf{x}^+ = \mathbf{x}^k - \tilde{\gamma}_k \mathbf{p}^k$*. Der Schrittweitenparameter* $\tilde{\gamma}_k$ *wird so gewählt, daß die Armijo-Goldstein-Bedingung*

$$g(\mathbf{x}^k) - g(\mathbf{x}^k - \tilde{\gamma}_k \mathbf{p}^k) \quad \ge \quad \delta \tilde{\gamma}_k \nabla g(\mathbf{x}^k)^T \mathbf{p}^k$$

erfüllt ist. Die Wahl des Verfahrensparameter $\delta \in (0, \frac{1}{2})$ *sichert, daß die Minimumstelle einer quadratischen Funktion nicht ausgeschlossen wird. Die Bestimmung eines zulässigen* $\tilde{\gamma}_k$ *kann durch eine Kombination einer Reduktions- und einer Interpolationsformel (etwa durch eine quadratische oder kubische Parabel) effizienter erfolgen. Wir haben uns in unseren Algorithmen auf die einfache Reduktionstechnik beschränkt.*

4.2 Restringierte Quadratmittelprobleme

Wir betrachten jetzt das *restringierte nichtlineare Quadratmittelproblem* (NP)

$$\min \left\{ \frac{1}{2} \|\mathbf{F}_1(\mathbf{x})\|^2 \; : \; \mathbf{F}_2(\mathbf{x}) = \mathbf{0},\, \mathbf{F}_3(\mathbf{x}) \geq \mathbf{0},\, \mathbf{x} \in \mathbb{R}^n \right\} \tag{4.13}$$

mit $\mathbf{F}_i : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^{m_i}$. $\mathbf{J}_i$ bezeichne die entsprechenden Jacobi-Matrizen $\mathbf{F}_i'$. Zur Charakterisierung einer Optimallösung von (4.13) zitieren wir Standardaussagen der nichtlinearen Optimierung. Unsere Darstellung folgt [Boc87] (siehe auch [Fle87], [GMW81], [Spe93]).

4.2.1 Optimalitätsbedingungen

Definition 8 (Lagrange-Funktion)
Die Funktion $L : \mathbb{R}^n \times \mathbb{R}^{m_2} \times \mathbb{R}^{m_3}_+ \longrightarrow \mathbb{R}$ mit

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \;\; := \;\; \frac{1}{2} \|\mathbf{F}_1(\mathbf{x})\|^2 - \boldsymbol{\lambda}^T \mathbf{F}_3(\mathbf{x}) - \boldsymbol{\mu}^T \mathbf{F}_3(\mathbf{x})$$

$\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\lambda} \in \mathbb{R}^{m_2}$, $\boldsymbol{\mu} \in \mathbb{R}^{m_3}_+$ heißt Lagrange-Funktion zu (4.13).

Wir fassen die an einer Stelle $\mathbf{z} \in D$ aktiven Ungleichheitsnebenbedingungen durch die Indexmenge $I(\mathbf{z})$ zusammen:

$$I(\mathbf{z}) \;\; := \;\; \{ i \,|\, F_{3i}(\mathbf{z}) = 0, \quad i = 1, \ldots, m_3 \}$$

Die entsprechenden Komponenten von $\mathbf{F}_3$ und ihre Ableitungen seien mit $\tilde{\mathbf{F}}_3$ bzw. $\tilde{\mathbf{J}}_3$ bezeichnet. Ferner bezeichne $\mathbf{F}_c := \begin{pmatrix} \mathbf{F}_2 \\ \tilde{\mathbf{F}}_3 \end{pmatrix}$ die zusammengesetzte Abbildung $\mathbf{F}_c : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^{m_c}$, $m_c := m_2 + \tilde{m}_3$, $\tilde{m}_3 := \#I$ und $\mathbf{J}_c := \begin{bmatrix} \mathbf{J}_2 \\ \tilde{\mathbf{J}}_3 \end{bmatrix}$ die entsprechende Jacobi-Matrix.

Definition 9 (Regulärer Punkt)
Der Punkt $\mathbf{z} \in D$ heißt regulär, wenn

$$\mathit{Rang}\, \mathbf{J}_c(\mathbf{z}) \;\; = \;\; m_c. \tag{4.14}$$

Bemerkung 13
Die in der Definition geforderte Regularitätsbedingung (auch Slater-Bedingung, constraint qualification genannt) kann abgeschwächt werden. Es muß gefordert werden, daß der Tangentenkegel $\mathcal{F}$ an einer Stelle $\mathbf{z}$ mit der Menge F aller zulässigen Richtungen der linearisierten Nebenbedingungen übereinstimmt (siehe [Fle87, S.202]). Hinreichende Bedingungen dafür sind

- *Die Vektoren $\mathbf{J}_{2i}, \quad i = 1, \ldots, m_2$ und $\tilde{\mathbf{J}}_{3i}, \quad i = 1, \ldots, \tilde{m}_3$ sind linear unabhängig ($\equiv$ 4.14).*
- *Die aktiven Nebenbedingungen $\mathbf{F}_2 = \mathbf{0} \quad \tilde{\mathbf{F}}_3 = \mathbf{0}$ sind linear.*

Einen Überblick über gebräuchliche Restriktionsqualifikationen der nichtlinearen Optimierung gibt [Spe93].

Wir können nun die notwendigen Optimalitätsbedingungen erster Ordnung formulieren.

Satz 6 (Notwendige Optimalitätsbedingungen erster Ordnung)
Sei $\mathbf{x}^* \in D$ ein regulärer Punkt und Lösung des restringierten nichtlinearen Quadratmittelproblems (4.13). Dann ist $\mathbf{x}^*$ zulässig

$$\mathbf{F}_2(\mathbf{x}^*) = \mathbf{0} \quad \mathbf{F}_3(\mathbf{x}^*) \geq \mathbf{0} \tag{4.15}$$

und es gilt:
Es existieren Lagrange-Parameter $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$ als eindeutige Lösung der Stationaritätsbedingungen

$$\frac{\partial}{\partial \mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{F}_1(\mathbf{x}^*)^T \mathbf{J}_1(\mathbf{x}^*) - {\boldsymbol{\lambda}^*}^T \mathbf{J}_2(\mathbf{x}^*) - {\boldsymbol{\mu}^*}^T \mathbf{J}_3(\mathbf{x}^*) = \mathbf{0} \tag{4.16}$$

$$\boldsymbol{\mu}^* \geq \mathbf{0} \qquad i \notin I(\mathbf{x}^*) \implies \mu_i^* = 0 \tag{4.17}$$

Ein Tupel $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, das die Zulässigkeitsbedingung (4.15) und die Stationaritätsbedingungen (4.16), (4.17) erfüllt, heißt Kuhn-Tucker-Punkt, die Bedingungen (4.15), (4.16), (4.17) Kuhn-Tucker-Bedingungen.

Definition 10
Die Richtung $\mathbf{p} \in T(\mathbf{x}^*) \subset \mathbb{R}^n$

$$T(\mathbf{x}^*) \quad := \quad \left\{ \mathbf{p} \neq \mathbf{0} \,|\, \mathbf{J}_2(\mathbf{x}^*)\mathbf{p} = \mathbf{0},\, \mu_i \mathbf{J}_{3i}(\mathbf{x}^*)\mathbf{p} = 0 \quad i \in I(\mathbf{x}^*),\, \tilde{\mathbf{J}}_3(\mathbf{x}^*)\mathbf{p} \geq \mathbf{0} \right\}$$

heißt zulässige Richtung der linearisierten Nebenbedingungen.

Schließlich geben wir noch die notwendigen Optimalitätsbedingungen zweiter Ordnung an.

Satz 7 (Notwendige Optimalitätsbedingungen zweiter Ordnung)
Falls $\mathbf{x}^* \in D$ ein regulärer Punkt und Lösung des restringierten nichtlinearen Quadratmittelproblems (4.13) ist, so gilt für alle zulässigen Richtungen der linearisierten Nebenbedingungen $\mathbf{p} \in T(\mathbf{x}^*)$

$$\mathbf{p}^T \mathbf{H}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\mathbf{p} \geq 0 \quad \mathit{mit} \quad \mathbf{H}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) := \frac{\partial^2}{\partial \mathbf{x}^2} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \tag{4.18}$$

Satz 8 (Hinreichende Optimalitätsbedingungen zweiter Ordnung)
Sei $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ Kuhn-Tucker-Punkt von (4.13) und für alle zulässigen Richtungen der linearisierten Nebenbedingungen $\mathbf{p} \in T(\mathbf{x}^*)$ sei die Hesse-Matrix der Lagrange-Funktion positiv definit

$$\mathbf{p}^T \mathbf{H}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\mathbf{p} > 0. \tag{4.19}$$

Dann ist $\mathbf{x}^*$ striktes lokales Minimum von (4.13).

4.2.2 Verallgemeinertes Gauß-Newton-Verfahren

Zur numerischen Lösung des restringierten nichtlinearen Quadratmittelproblems (4.13) betrachten wir eine Verallgemeinerung des Gauß-Newton-Verfahrens (siehe [Boc87]). Grundlage ist die wiederholte Lösung eines linearisierten restringierten Quadratmittelproblems.

Linearisiertes restringiertes Quadratmittelproblem $LP(\mathbf{x})$:

$$\min \left\{ \frac{1}{2} \left\| \mathbf{F}_1(\mathbf{x}) + \mathbf{J}_1(\mathbf{x})\mathbf{s} \right\|^2 \,:\, \mathbf{F}_2(\mathbf{x}) + \mathbf{J}_2(\mathbf{x})\mathbf{s} = \mathbf{0},\, \mathbf{F}_3(\mathbf{x}) + \mathbf{J}_3(\mathbf{x})\mathbf{s} \geq \mathbf{0},\, \mathbf{s} \in \mathbb{R}^n \right\} \tag{4.20}$$

Grundalgorithmus Verallgemeinertes Gauß-Newton-Verfahren

S1: Wähle Startpunkt $\mathbf{x}^0 \in \mathbb{R}^n$

S2: For $k = 0, 1, \ldots$ do

S2.1: Bestimme $\mathbf{s}^k$ als Lösung des linearisierten restringierten Quadratmittelproblems $LP(\mathbf{x}^k)$

$$\left\|\mathbf{F}_1(\mathbf{x}^k) + \mathbf{J}_1(\mathbf{x}^k)\mathbf{s}\right\|^2 \longrightarrow \min$$

unter den Nebenbedingungen

$$\begin{aligned} \mathbf{F}_2(\mathbf{x}^k) + \mathbf{J}_2(\mathbf{x}^k)\mathbf{s} &= \mathbf{0} \\ \mathbf{F}_3(\mathbf{x}^k) + \mathbf{J}_3(\mathbf{x}^k)\mathbf{s} &\geq \mathbf{0} \end{aligned}$$

S2.2: Setze $\mathbf{x}^{k+1} := \mathbf{x}^k + \mathbf{s}^k$;

Betrachten wir die hinreichenden Optimalitätsbedingungen (4.19) für das linearisierte restringierte Quadratmittelproblem (4.20). Die Hesse-Matrix der Lagrange-Funktion ist gerade $\mathbf{J}_1(\mathbf{x})^T\mathbf{J}_1(\mathbf{x})$. Die Bedingung

$$\mathbf{p}^T\mathbf{J}_1^T\mathbf{J}_1\mathbf{p} > 0 \quad \text{für alle}\, \mathbf{p} \in Z := \{\mathbf{p} \neq \mathbf{0} \,:\, \mathbf{J}_c\mathbf{p} = \mathbf{0}\} \tag{4.21}$$

ist äquivalent mit

$$Rang \begin{bmatrix} \mathbf{J}_1 \\ \mathbf{J}_c \end{bmatrix} = n. \tag{4.22}$$

Sei $(\mathbf{s}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ Kuhn-Tucker-Punkt des linearisierten restringierten Quadratmittelproblems (4.20) an der Stelle $\mathbf{x}^*$ und die hinreichende Bedingung (4.21) bzw. (4.22) sei erfüllt. Bock [Boc87] zeigt, daß es dann eine Umgebung X von $\mathbf{x}^*$ gibt, so daß $\forall \mathbf{x} \in X$ das linearisierte restringierte Quadratmittelproblem $LP(\mathbf{x})$ genau einen Kuhn-Tucker-Punkt $(\mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ besitzt und das $\mathbf{s}$ striktes Minimum ist. Unter der Voraussetzung der strikten Komplementarität der Kuhn-Tucker-Vorzeichenbedingung (4.17)

$$i \in I(\mathbf{x}^*) \quad \Longrightarrow \quad \mu_i^* > 0 \tag{4.23}$$

gilt: $I_{LP(\mathbf{x})}(\mathbf{s})$ ist konstant auf X, d.h. die Indexmenge der aktiven Ungleichheitsnebenbedingungen ändert sich nicht mehr.

Bemerkung 14
Insbesondere kann bei Vorliegen strikter Komplementarität kein Zigzagging, *d.h. das Oszillieren zwischen verschiedenen Teilmengen der Nebenbedingungen, auftreten, welches die Konvergenz eines numerischen Algorithmus drastisch verschlechtern kann (siehe [Fle87, S.268]).*

Damit ist gezeigt, daß das lokale Konvergenzverhalten des ungleichungrestringierten Problems dem des gleichungsrestringierten Problems entspricht, in dem die Gleichungs- und aktiven Ungleichungsnebenbedingungen zusammengefaßt werden. Wir beschränken uns deshalb im weiteren auf solche Probleme mit dem Grundgebiet $D = X$ in der Umgebung der Lösung.

Voraussetzungen:
Für alle $\mathbf{x} \in D$ gelte

$$Rang\, \mathbf{J}_c(\mathbf{x}) = m_c \tag{4.24}$$

$$Rang\, \mathbf{J}(\mathbf{x}) = n \qquad \mathbf{J}(\mathbf{x}) := \begin{bmatrix} \mathbf{J}_1(\mathbf{x} \\ \mathbf{J}_c(\mathbf{x}) \end{bmatrix} \tag{4.25}$$

Bock [Boc87] zeigt nun, daß sich die Lösung dieses gleichungsrestringierten linearen Quadratmittelproblems durch eine verallgemeinerte Inverse ausdrücken läßt. Mit dem folgenden Satz aus [Boc87] ist damit eine einheitliche Konvergenztheorie für freie und restringierte Quadratmittelprobleme möglich.

Satz 9 (Existenz einer verallgemeinerten Inversen)
Betrachte das gleichungsrestringierte lineare Quadratmittelproblem

$$||\mathbf{J}_1\mathbf{s} + \mathbf{F}_1||^2 \quad \longrightarrow \quad \min \tag{4.26}$$

unter der Nebenbedingung

$$\mathbf{J}_c\mathbf{s} + \mathbf{F}_c \quad = \quad \mathbf{0}. \tag{4.27}$$

$\mathbf{J}_1$ *und* $\mathbf{J}_c$ *erfüllen die Voraussetzungen (4.24), (4.25). Dann gilt:*

(i) Für beliebiges $\mathbf{F} := \begin{pmatrix} \mathbf{F}_1 \\ \mathbf{F}_c \end{pmatrix} \in \mathbb{R}^{m_1+m_c}$ *existiert genau ein Kuhn-Tucker-Punkt* $(\mathbf{s}, \boldsymbol{\lambda}_c)$ *von (4.26), (4.27), und* $\mathbf{s}$ *ist striktes Minimum.*

(ii) Es existiert eine lineare Abbildung $\mathbf{J}^- : \mathbb{R}^{m_1+m_c} \longrightarrow \mathbb{R}^n$*, so daß* $\mathbf{s} = -\mathbf{J}^-\mathbf{F}$ *Lösung von (4.26), (4.27) ist für beliebiges* $\mathbf{F} \in \mathbb{R}^{m_1+m_c}$*.*

(iii) Der Lösungsoperator $\mathbf{J}^-$ *ist eine verallgemeinerte Inverse, d.h.*

$$(P2) \qquad \mathbf{J}^-\mathbf{J}\mathbf{J}^- = \mathbf{J}^-.$$

Bemerkung 15

- *Unter den Voraussetzungen (4.24) und (4.25) gelten auch die Moore-Penrose Bedingungen (P1) und (P4).*
- *Unter den Voraussetzungen (4.24), (4.25) ist* $\mathbf{J}(\mathbf{x})^-$ *stetig differenzierbar auf* X *(vgl. Abschnitt 5).*

Die lokale Konvergenz des verallgemeinerten Gauß-Newton-Verfahrens folgt nun aus Satz 9 und dem folgenden Satz.

Satz 10 (Kontraktionssatz, [Boc87])
Sei $\mathbf{J}(\mathbf{x})^-$ *verallgemeinerte Inverse zur Ableitung* $\mathbf{J}(\mathbf{x})$ *von* $\mathbf{F} \in C^1(D)$*, und die Lipschitzbedingungen an* $\mathbf{J}$ *bzw.* $\mathbf{J}^-$

$$\begin{aligned} \left\|\mathbf{J}(\mathbf{y})^- \left[\mathbf{J}\left(\mathbf{x} + t(\mathbf{y} - \mathbf{x})\right) - \mathbf{J}(\mathbf{x})\right](\mathbf{y} - \mathbf{x})\right\| &\leq \omega t \left\|\mathbf{y} - \mathbf{x}\right\|^2 \quad \mathit{mit} \quad \omega < \infty & (4.28) \\ \left\|\left[\mathbf{J}(\mathbf{z})^- - \mathbf{J}(\mathbf{x})^-\right]\mathbf{R}(\mathbf{x})\right\| &\leq \kappa(\mathbf{x})\left\|\mathbf{z} - \mathbf{x})\right\| \leq \kappa\left\|\mathbf{z} - \mathbf{x}\right\| & (4.29) \\ \mathbf{R}(\mathbf{x}) &:= \mathbf{F}(\mathbf{x}) - \mathbf{J}(\mathbf{x})\mathbf{J}(\mathbf{x})^-\mathbf{F}(\mathbf{x})\, mit\, \kappa < 1 & \end{aligned}$$

seien erfüllt ($\forall t \in [0,1], \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in D,\ \mathbf{x} - \mathbf{y} = \mathbf{J}(\mathbf{x})^-\mathbf{F}(\mathbf{x})$*).*
Für alle Startwerte $\mathbf{x}^0 \in D$ *mit*

$$\begin{aligned} \delta_0 &:= \frac{\alpha_0 \omega}{2} + \kappa < 1 \qquad \alpha_k := \left\|\mathbf{J}(\mathbf{x}^k)^-\mathbf{F}(\mathbf{x}^k)\right\| \\ D_0 &:= \bar{K}\left(\mathbf{x}^0, \frac{\alpha_0}{1 - \delta_0}\right) \subset D \end{aligned}$$

folgt dann:

(i) die Iteration $\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^k$*,* $\Delta\mathbf{x}^k := -\mathbf{J}(\mathbf{x}^k)^-\mathbf{F}(\mathbf{x}^k)$ *ist wohldefiniert und bleibt in* D_0*,*

(ii) $\mathbf{x}^k \longrightarrow \mathbf{x}^* \in D_0$ *mit* $\mathbf{J}(\mathbf{x}^*)^- \mathbf{F}(\mathbf{x}^*) = \mathbf{0}$, *und es gilt die*

(iii) a priori Abschätzung

$$\left\| \mathbf{x}^k - \mathbf{x}^* \right\| \leq \delta_0^k \frac{\alpha_0}{1 - \delta_0}.$$

(iv) Die Konvergenz ist linear

$$\left\| \Delta \mathbf{x}^{k+1} \right\| \quad \leq \quad \left(\frac{\alpha_k \omega}{2} + \kappa \right) \left\| \Delta \mathbf{x}^k \right\|.$$

Bemerkung 16

- *Die Bedingung $\kappa < 1$ kennzeichnet im statistischen Sinne vernünftig gestellte Quadratmittelprobleme. Bock [Boc87] zeigt, daß unter der Voraussetzung $\kappa < 1$ der berechnete Fixpunkt $\mathbf{x}^*$ nicht nur stationärer Punkt des Problems, sondern auch lokales Minimum und stabil gegen lokale Störungen ist. Umgekehrt wird gezeigt, daß andernfalls Störungen existieren, für die der Fixpunkt zwar stationär, aber kein Minimum ist.*

- *Alle Annahmen und Voraussetzungen des Kontraktionssatzes sind affin invariant.*

4.2.3 Globalisierung

Zur Globalisierung des soeben beschriebenen Verallgemeinerten Gauß-Newton-Verfahrens führen wir eine Dämpfung aus:

$$\mathbf{x}^{k+1} \quad := \quad \mathbf{x}^k + \alpha_k \mathbf{s}^k. \tag{4.30}$$

Der Schrittweitenparameter α_k wird dabei so bestimmt, daß eine sogenannte *Niveaufunktion* einen ausreichenden Abstieg signalisiert. Neben dem Wert der Zielfunktion muß bei restringierten Quadratmittelproblemen auch berücksichtigt werden, wie stark die Nebenbedingungen verletzt werden.

Bock [Boc87] betrachtet die Niveaufunktion

$$T(\mathbf{x}) \quad := \quad \frac{1}{2} \left\| \mathbf{F}_1(\mathbf{x}) \right\|^2 + \sum_{i=1}^{m_2} \beta_i \left| F_{2i}(\mathbf{x}) \right| + \sum_{i=1}^{m_3} \gamma_i \left| \min\left(0, F_{3i}(\mathbf{x})\right) \right|.$$

Die Bestimmung des Schrittweitenparameters α_k erfolgt über exakte Strahlsuche.

Gedämpftes Verallgemeinertes Gauß-Newton-Verfahren

S1: Wähle Startpunkt $\mathbf{x}^0 \in \mathbb{R}^n$; $k := 0$;

S2: Löse $LP(\mathbf{x}^k)$;

S3: Bestimme α_k, so daß

$$T\left(\mathbf{x}^k + \alpha_k \mathbf{s}^k\right) \quad = \quad \min\left\{ T\left(\mathbf{x}^k + \alpha \mathbf{s}\right) \,:\, \alpha \in [0, 1] \right\}$$

S3: Setze $\mathbf{x}^{k+1} := \mathbf{x}^k + \alpha_k \mathbf{s}^k$; $k := k + 1$; Goto S2;

Bemerkung 17

Die in Schritt S3 angenommene exakte Strahlsuche kann durch eine näherungsweise ersetzt werden (etwa Armijo-Goldstein-Schrittweitenstrategie).

Mit dieser Niveaufunktion beweist Bock globale Konvergenz des Verallgemeinerten Gauß-Newton-Verfahrens. Auf dem Grundgebiet D seien dabei die Regularitätsannahmen (4.24), (4.25) erfüllt.

Satz 11 (Globale Konvergenz des Verallgemeinerten GN-Verfahrens)
(i) Sei $\mathbf{x}^0 \in D$ beliebig, und die zugehörige Niveaumenge

$$D_0 \quad := \quad \left\{\mathbf{x} \,|\, T(\mathbf{x}) \le T(\mathbf{x}^0)\right\} \subset D$$

sei kompakt.

(ii) Für alle $\mathbf{x} \in D_0$ gelte für den eindeutigen Kuhn-Tucker-Punkt $(\mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ des linearisierten restringierten Problems $LP(\mathbf{x})$

$$|\lambda_i| < \beta_i \qquad |\mu_i| < \gamma_i$$

Dann erzeugt das gedämpfte Verallgemeinerte Gauß-Newton-Verfahren eine Folge $\{\alpha_k\}$, so daß die relaxierte Iteration (4.30) mit Startwert $\mathbf{x}^0$ einen Kuhn-Tucker-Punkt des nichtlinearen Problems (4.13) als Häufungspunkt besitzt.

4.2.4 Nichtlineare Quadratmittelprobleme mit linearen Ungleichheitsnebenbedingungen

In diesem Abschnitt schränken wir uns auf *nichtlineare Quadratmittelprobleme mit linearen Ungleichheitsnebenbedingungen* ein.

$$\min\left\{\frac{1}{2}\left\|\mathbf{F}(\mathbf{x})\right\|^2 \,:\, \mathbf{C}\mathbf{x} \ge \mathbf{h},\, \mathbf{x} \in \mathbb{R}^n\right\} \tag{4.31}$$

mit $\mathbf{F} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $\mathbf{C} \in \mathbb{R}^{ncstr,n}$, $\mathbf{h} \in \mathbb{R}^{ncstr}$.

Wright [Wri92] unterscheidet die aktuell verwendeten Algorithmen zur Lösung von (4.31) in zwei Klassen. Auf der einen Seite gibt es die *active set* Zugänge. Bei diesen Algorithmen wird in jeder *äußeren* Iteration eine Modellfunktion von $g(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{F}(\mathbf{x})\right\|^2$ gebildet. In jeder inneren Iteration wird diese Modellfunktion dann über dem zulässigen Bereich $\{\mathbf{x} \in \mathbb{R}^n \,|\mathbf{C}\mathbf{x} \ge \mathbf{h}\}$ minimiert. Bei der zweiten Klasse, den sogenannten *gradient projection methods*, wird in jeder Iteration eine Suchrichtung $\mathbf{p}$ gewählt. Anschließend wird auf dem Strahl $\mathbf{P}(\mathbf{x} - \alpha\mathbf{p})$ gesucht, wobei $\alpha > 0$ und $\mathbf{P}$ die Projektion auf den zulässigen Bereich ist.

Die von uns verwendete Strategie folgt dem im vorigen Abschnitt dargelegten Verallgemeinerten Gauß-Newton-Verfahren und läßt sich damit der ersten Klasse zuordnen. Einen allgemeinen Überblick über linear restringierte Optimierungsprobleme gibt [Fle87, Chapter 21].

Bei der Anwendung der Methoden aus Abschnitt 4.2.2 auf das Problem (4.31) ergeben sich einige Vereinfachungen. So ist die Regularitätsforderung (constraint qualification) automatisch erfüllt. Als Spezialisierung des linearisierten restringierten Problems LS auf (4.31) erhalten wir

Linearisiertes restringiertes Quadratmittelproblem $LSI(\mathbf{x})$:

$$\min\left\{\frac{1}{2}\left\|\mathbf{F}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{s}\right\|^2 \,:\, \mathbf{C}(\mathbf{x} + \mathbf{s}) \ge \mathbf{h},\, \mathbf{s} \in \mathbb{R}^n\right\} \tag{4.32}$$

Dieses Problem kann unter der Voraussetzung *Rang* $\mathbf{J}(\mathbf{x}) = n$ mit den in Abschnitt 3.2 entwickelten Methoden gelöst werden.

Bemerkung 18
Die Vollrangbedingung Rang $\mathbf{J}(\mathbf{x}) = n$ *ist analog zu der Vollrangbedingung für das Gauß-Newton-Verfahren für freie Quadratmittelprobleme. Sie sichert zugleich die Rangbedingung (4.25) Rang* $\begin{bmatrix} \mathbf{J}(\mathbf{x}) \\ \mathbf{C}_a \end{bmatrix} = n$ *(* $\mathbf{C}_a$ *aktive Nebenbedingungen) und damit die Existenz eines strikten lokalen Minimums* $\mathbf{s}$.

Die Globalisierung erfolgt wieder über Dämpfung. Es ergeben sich jedoch auch hier einige wichtige Vereinfachungen. Die Terme

$$\sum_{i=1}^{m_2} \beta_i \left| F_{2i}(\mathbf{x}) \right| + \sum_{i=1}^{m_3} \gamma_i \left| \min\left(0, F_{3i}(\mathbf{x})\right) \right|$$

in der Definition der Niveaufunktion entfallen, da das Verfahren nur mit zulässigen Punkten arbeitet. Wir arbeiten stattdessen wieder mit der Armijo-Goldstein-Schrittweitenstrategie für die Funktion $g(\mathbf{x}) = \frac{1}{2} \left\| \mathbf{F}(\mathbf{x}) \right\|^2$.
Abschließend formulieren wir das verwendete Verallgemeinerte Gauß-Newton-Verfahren zur Lösung von (4.31).

Gedämpftes Verallgemeinertes Gauß-Newton-Verfahren

S1: Wähle (zulässigen) Startwert $\mathbf{x}^0 \in \mathbb{R}^n$, $\alpha \in (0,1)$, $\delta \in (0, \frac{1}{2})$
Setze $k := 0$; $\gamma_{-1} := 1$;

S2: Iteration
Repeat

S2.1: $\mathbf{F} := \mathbf{F}(\mathbf{x}^k)$; Berechnung des Residiums
$\mathbf{J} := \mathbf{F}'(\mathbf{x}^k)$; Berechnung der Jacobi-Matrix
S2.2: Bestimme Abstiegsrichtung $\mathbf{s}^k (= -\mathbf{p}^k)$ aus

$$\begin{array}{rccc} & \left\| \mathbf{F} + \mathbf{J}\mathbf{s} \right\|^2 & \longrightarrow & \min \\ \text{mit} & \mathbf{C}\left(\mathbf{x}^k + \mathbf{s}\right) & \geq & \mathbf{h} \end{array}$$

S2.3: $\tilde{\gamma}_k := \gamma_k$; $\mathbf{x}^+ := \mathbf{x}^k + \tilde{\gamma}_k \mathbf{s}^k$; $\mathbf{F}^+ := \mathbf{F}(\mathbf{x}^+)$;
S2.4: Abstiegstest
if $\frac{1}{2} \left\| \mathbf{F} \right\|^2 - \frac{1}{2} \left\| \mathbf{F}^+ \right\|^2 \geq \delta \tilde{\gamma}_k \mathbf{F}^T \mathbf{J} \mathbf{p}^k$
then
$\mathbf{x}^{k+1} := \mathbf{x}^+$;
$\gamma_{k+1} := \min(\gamma_k / \alpha, 1)$;
$k := k + 1$;
else
$\gamma_k := \gamma_k * \alpha$; goto S2.3;

until ($k > kmax$) or Konvergenz

Bemerkung 19

- *Die Berechnung der Abstiegsrichtung* $\mathbf{s}^k$ *in Schritt S2.2 erfolgt mittels des Algorithmus LSI aus Abschnit 3.2.*
- *Die Übertragung der Armijo-Goldstein-Schrittweitenstrategie auf linear restringierte Probleme bereitet keinerlei Schwierigkeiten, da alle Punkte längs des Strahls* $\mathbf{x}^k - \alpha \mathbf{p}^k$, $\alpha \in [0,1]$ *zulässig sind. Bei nichtlinearen Restriktionen ist die Verfolgung eines zulässigen Kurvenbogens* $\mathbf{x}(\alpha)$ *mittels eines Korrekturterms notwendig, was die „Strahl"suche i.a. sehr verteuert. Eine Armijo-Goldstein-Schrittweitenstrategie für solche nichtlinear restringierten Probleme beschreibt Spellucci [Spe93].*

- *Die Bemerkungen zum gedämpften Gauß-Newton-Verfahren mit akzeptiertem Schätzwert gelten sinngemäß.*

4.2.5 Regularisierung

Die Durchführbarkeit und Konvergenz von (Verallgemeinerten) Gauß-Newton-Verfahren benötigt als wesentliche Voraussetzung die Vollrangeigenschaft $Rang\,\mathbf{J}(\mathbf{x}) = n$ der Jacobi-Matrix. Wenn die Matrix $\mathbf{J}$ (fast) rangdefizient ist, so ist der Algorithmus LSI zur Lösung des linearisierten restringierten Teilproblems nicht durchführbar und bricht ab.

Im Fall der freien Quadratmittelprobleme erfolgt eine Regularisierung meist über ein Trust-Region-Verfahren, d.h.

$$\min\left\{\mu_{GN}(\mathbf{x}+\mathbf{s}) = g(\mathbf{x}) + \nabla g(\mathbf{x})^T\mathbf{s} + \frac{1}{2}\mathbf{s}^T\mathbf{H}(\mathbf{x})\mathbf{s} \;:\; \mathbf{s}\in\mathbb{R}^n \text{ bei } \|\mathbf{s}\|^2 \le \Delta^2\right\} \tag{4.33}$$

mit dem Trust-Region-Radius $\Delta = \Delta_k > 0$. Bis auf Ausnahmefälle ergibt sich die Lösung von (4.33) als Lösung von

$$(\mathbf{H} + \lambda\mathbf{I})\,\mathbf{s} = -\nabla g,$$

wobei der Lagrange-Parameter $\lambda \ge 0$ so gewählt werden muß, daß $\mathbf{H} + \lambda\mathbf{I}$ positiv definit ist und außerdem $\|\mathbf{s}(\lambda)\| = \Delta$ gilt.

Für das Gauß-Newton-Modell erhalten wir

$$\min\left\{\left\|\begin{bmatrix}\mathbf{J}\\ \sqrt{\lambda}\mathbf{I}\end{bmatrix}\mathbf{s} + \begin{pmatrix}\mathbf{F}\\ \mathbf{0}\end{pmatrix}\right\|^2 = \|\mathbf{F}+\mathbf{J}\mathbf{s}\|^2 + \lambda\|\mathbf{s}\|^2 \;:\; \mathbf{s}\in\mathbb{R}^n\right\} \tag{4.34}$$

Steuert man den Trust-Region-Radius über den Lagrange-Parameter λ spricht man vom Levenberg-Marquardt-Verfahren. Für Trust-Region-Verfahren kann man unter schwachen Voraussetzungen globale Konvergenz beweisen, d.h. der Trust-Region-Ansatz dient nicht nur zur Regularisierung, sondern auch zur Globalisierung.

Dennis/Schnabel [DS83, S.151] beschreiben eine Strategie zur Regularisierung des Newton-Verfahrens bei Verwendung eines gedämpften Verfahrens. Falls die Hesse-Matrix $\mathbf{H} = \nabla^2 g(\mathbf{x})$ des quadratischen Modells schlecht konditioniert ist, wird das quadratische Modell

$$\min\left\{\mu_N(\mathbf{x}+\mathbf{s}) = g(\mathbf{x}) + \nabla g(\mathbf{x})^T\mathbf{s} + \frac{1}{2}\mathbf{s}^T\mathbf{H}(\mathbf{x})\mathbf{s} \;:\; \mathbf{s}\in\mathbb{R}^n\right\} \tag{4.35}$$

gestört:

$$\mathbf{H}_{reg} \;:=\; \mathbf{H} + \sqrt{n \times macheps}\,\|\mathbf{H}\|_1 \times \mathbf{I}$$

Diese Strategie ist in dem in [DS83] enthaltenen Programmpaket realisiert. Dort findet sich auch die theoretische Fundierung der Wahl der Größe des Lagrange-Parameters.

Wir verwenden diese Strategie zur Regularisierung des Verallgemeinerten Gauß-Newton-Verfahrens. Der Algorithmus LSI führt eine QR-Faktorisierung von $\mathbf{J}$ durch und schätzt die Kondition der oberen Dreicksmatrix $\mathbf{R}$. Falls die Konditionszahl größer als $\frac{1}{\sqrt{macheps}}$ ist, bricht der Algorithmus ab und teilt dies dem aufrufenden Algorithmus mit. Wir berechnen dann

$$\lambda \;:=\; \sqrt{n \times macheps}\,\left\|\mathbf{J}^T\mathbf{J}\right\|_1$$

und lösen das Problem

$$\min \left\{ \left\| \begin{bmatrix} \mathbf{J} \\ \sqrt{\lambda}\mathbf{I} \end{bmatrix} \mathbf{s} + \begin{pmatrix} \mathbf{F} \\ \mathbf{0} \end{pmatrix} \right\|^2 : \mathbf{C}\left(\mathbf{x}^k + \mathbf{s}\right) \geq \mathbf{h},\, \mathbf{s} \in \mathbb{R}^n \right\}. \tag{4.36}$$

Bemerkung 20

- *Die soeben beschriebene Regularisierung wurde in den verwendeten Algorithmen implementiert und liefert gute numerische Ergebnisse.*

- *Die Verwendung eines Trust-Region-Verfahrens zur Lösung von nichtlinearen restringierten Quadratmittelproblemen beschreiben [HF79]. Die Steuerung des Trust-Region-Radius erfolgt dabei nicht über den Lagrange-Parameter λ, sondern über Δ, da Δ nicht so empfindlich bei der Änderung der aktiven Menge ist. Trust-Region-Verfahren für Probleme mit linearen Ungleichheitsnebenbedingungen findet man in [Gay84] und [WH85].*

5 Separable Quadratmittelprobleme

5.1 Problemformulierung

Ein nichtlineares Quadratmittelproblem, bei denen ein Teil der Unbekannten linear auftritt, wird als *(linear) separabel* bezeichnet.

$$\min\left\{\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha}) := \frac{1}{2}\left\|\mathbf{F}\left(\tilde{\mathbf{t}},\boldsymbol{\alpha}\right)\right\|^2 = \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \;:\; \tilde{\mathbf{t}}\in\mathbb{R}^l, \boldsymbol{\alpha}\in\mathbb{R}^n\right\} \tag{5.1}$$

$\mathbf{y}\in\mathbb{R}^m$ sei dabei ein vorgegebener Vektor und $\mathbf{B}(.) : \mathbb{R}^l \longrightarrow \mathbb{R}^{m\times n}$ eine hinreichend glatte Matrixfunktion. Es gelte $m \geq n + l$, d.h. das System (5.1) sei überbestimmt.

Separable Quadratmittelprobleme sind ein Spezialfall sog. *reduzibler* nichtlinearer Optimierungsprobleme, bei denen eine natürliche Unterscheidung der Variablen in zwei Gruppen auftritt (siehe [Par85] und [RW80]).

$$\min\left\{f(\tilde{\mathbf{t}},\boldsymbol{\alpha}) \;:\; \tilde{\mathbf{t}}\in\mathbb{R}^l, \boldsymbol{\alpha}\in\mathbb{R}^n\right\} \tag{5.2}$$

Die Variablen, welche den Vektor $\boldsymbol{\alpha}$ darstellen, sind dabei so gewählt, daß das *reduzierte* (n-dimensionale) *Problem*

$$\min\left\{f(\tilde{\mathbf{t}}_c,\boldsymbol{\alpha}) \;:\; \tilde{\mathbf{t}}_c\in\mathbb{R}^l \quad \text{fest}, \boldsymbol{\alpha}\in\mathbb{R}^n\right\} \tag{5.3}$$

effektiv und akkurat gelöst werden kann.

Typische Beispiele für separable Quadratmittelprobleme sind:

- Parameterschätzung bei Spektren

$$\frac{1}{2}\sum_{i=1}^{m}\left[y_i - r\left(x_i,\tilde{\mathbf{t}},\boldsymbol{\alpha}\right)\right]^2 \longrightarrow \text{Min!}$$

$$r\left(x,\tilde{\mathbf{t}},\boldsymbol{\alpha}\right) = \sum_{j=1}^{k}\alpha_j \exp\left(-\left(\frac{x-\tilde{t}_j}{\tilde{t}_{j+k}}\right)^2\right) + \alpha_{k+1} + \alpha_{k+2}x + \alpha_{k+3}x^2$$

  mit den $k+3$ linearen Parametern $(\alpha_1,\ldots,\alpha_{k+3})^T$ und den $2k$ nichtlinearen Parametern $(\tilde{t}_1,\ldots,\tilde{t}_{2k})^T$

- B-Splines mit freien Knoten

$$\frac{1}{2}\sum_{i=1}^{m}\left[y_i - r\left(x_i,\tilde{\mathbf{t}},\boldsymbol{\alpha}\right)\right]^2 \longrightarrow \text{Min!}$$

$$r\left(x,\tilde{\mathbf{t}},\boldsymbol{\alpha}\right) = \sum_{j=1}^{n} B_{j,k}(x,\tilde{\mathbf{t}})\alpha_j$$

  wobei $B_{j,k}(x,\tilde{\mathbf{t}})$ der j-te normalisierte B-Spline der Ordnung k an der Stelle x zur Knotenfolge $\tilde{\mathbf{t}}$ ist.

5.2 Gauß-Newton-ähnliche Verfahren für separable Quadratmittelprobleme

Wir benutzen das Gauß-Newton-Verfahren, um die Ideen zur effektiven Lösung von (5.1) darzulegen. Der Gauß-Newton-Schritt $\mathbf{s} \in \mathbb{R}^{n+l}$ wird aus der Gleichung

$$\mathbf{J}^T\mathbf{J}\mathbf{s} = -\mathbf{J}^T\mathbf{F} \tag{5.4}$$

berechnet. Die Jacobi-Matrix $\mathbf{J} \in \mathbb{R}^{m,n+l}$ der Residiuenfunktion $\mathbf{F}$ ist

$$\begin{aligned} \mathbf{J} &= [\mathbf{J}_\alpha\ \mathbf{J}_{\tilde{t}}] \\ \text{mit} \quad \mathbf{J}_\alpha &= \mathbf{B}\left(\tilde{\mathbf{t}}\right) \\ \text{und} \quad \mathbf{J}_{\tilde{t}} &= -\mathbf{DB}\left(\tilde{\mathbf{t}}\right)\boldsymbol{\alpha}. \end{aligned}$$

$\mathbf{DB}$ bezeichne die Fréchet-Ableitung der Matrixfunktion $\mathbf{B}(.)$.

> Eine $m \times n$ Matrix-Funktion $\mathbf{B}\left(\tilde{\mathbf{t}}\right)$ ist eine nichtlineare Abbildung zwischen dem linearen Raum der Parameter $\mathbb{R}^l$ und dem Raum der linearen Abbildungen $\mathfrak{L}(\mathbb{R}^n, \mathbb{R}^m)$. $\mathbf{DB}\left(\tilde{\mathbf{t}}\right)$ ist also ein Element von $\mathfrak{L}\left(\mathbb{R}^l, \mathfrak{L}(\mathbb{R}^n, \mathbb{R}^m)\right)$ und kann somit als dreidimensionaler Tensor interpretiert werden.

Wir können nun den Gauß-Newton-Schritt $\mathbf{s} \in \mathbb{R}^{n+l}$ in einen Schritt $\mathbf{s}_\alpha$ bezüglich der Variablen $\boldsymbol{\alpha}$ und einen Schritt $\mathbf{s}_{\tilde{t}}$ bezüglich $\tilde{\mathbf{t}}$ aufspalten. Schreiben wir die Normalgleichungen (5.4) aus, so erhalten wir folgende Gleichungen für $\mathbf{s}_\alpha$ und $\mathbf{s}_{\tilde{t}}$.

$$\begin{bmatrix} \mathbf{J}_\alpha^T\mathbf{J}_\alpha & \mathbf{J}_\alpha^T\mathbf{J}_{\tilde{t}} \\ \mathbf{J}_{\tilde{t}}^T\mathbf{J}_\alpha & \mathbf{J}_{\tilde{t}}^T\mathbf{J}_{\tilde{t}} \end{bmatrix} \begin{pmatrix} \mathbf{s}_\alpha \\ \mathbf{s}_{\tilde{t}} \end{pmatrix} = \begin{pmatrix} -\mathbf{J}_\alpha^T\mathbf{F} \\ -\mathbf{J}_{\tilde{t}}^T\mathbf{F} \end{pmatrix}$$

$$\mathbf{B}^T\mathbf{B}\mathbf{s}_\alpha + \mathbf{B}^T\mathbf{J}_{\tilde{t}}\mathbf{s}_{\tilde{t}} = -\mathbf{B}^T\mathbf{F} \tag{5.5}$$

$$\mathbf{J}_{\tilde{t}}^T\mathbf{B}\mathbf{s}_\alpha + \mathbf{J}_{\tilde{t}}^T\mathbf{J}_{\tilde{t}}\mathbf{s}_{\tilde{t}} = -\mathbf{J}_{\tilde{t}}^T\mathbf{F} \tag{5.6}$$

Indem wir die Gleichung (5.5) von links mit $\mathbf{J}_{\tilde{t}}^T\left(\mathbf{B}^+\right)^T$ multiplizieren und das Resultat von (5.6) subtrahieren, erhalten wir wegen der Beziehung $\mathbf{B}^{+T}\mathbf{B}^T\mathbf{B} = \left(\mathbf{B}\mathbf{B}^+\right)^T\mathbf{B} \overset{(P3)}{=} \mathbf{B}\mathbf{B}^+\mathbf{B} \overset{(P1)}{=} \mathbf{B}$ die Gleichung

$$\mathbf{J}_{\tilde{t}}^T\left(\mathbf{I} - \mathbf{B}\mathbf{B}^+\right)\mathbf{J}_{\tilde{t}}\mathbf{s}_{\tilde{t}} = -\mathbf{J}_{\tilde{t}}^T\left(\mathbf{I} - \mathbf{B}\mathbf{B}^+\right)\mathbf{F}.$$

Definieren wir den orthogonalen Projektor $\mathbf{P}_B^\perp := \mathbf{I} - \mathbf{B}\mathbf{B}^+$, so ergibt sich

$$\mathbf{J}_{\tilde{t}}^T\mathbf{P}_B^\perp\mathbf{J}_{\tilde{t}}\mathbf{s}_{\tilde{t}} = -\mathbf{J}_{\tilde{t}}^T\mathbf{P}_B^\perp\mathbf{F}.$$

Wegen der Symmetrie und Idempotenz eines Projektors ist dazu äquivalent

$$\left(\mathbf{P}_B^\perp\mathbf{J}_{\tilde{t}}\right)^T\left(\mathbf{P}_B^\perp\mathbf{J}_{\tilde{t}}\right)\mathbf{s}_{\tilde{t}} = -\left(\mathbf{P}_B^\perp\mathbf{J}_{\tilde{t}}\right)^T\mathbf{P}_B^\perp\mathbf{F}.$$

Dies sind aber genau die Normalgleichungen zu

$$\mathbf{P}_B^\perp\mathbf{J}_{\tilde{t}}\mathbf{s}_{\tilde{t}} \cong -\mathbf{P}_B^\perp\mathbf{F}.$$

Die eindeutige Normallösung ist

$$\mathbf{s}_{\tilde{t}} = -\left(\mathbf{P}_B^\perp\mathbf{J}_{\tilde{t}}\right)^+\mathbf{P}_B^\perp\mathbf{F}.$$

Aus der Gleichung (5.5) erhalten wir dann

$$\mathbf{s}_\alpha = -\mathbf{B}^+\left(\mathbf{F} + \mathbf{J}_{\tilde{t}}\mathbf{s}_{\tilde{t}}\right).$$

Jetzt sind wir in der Lage, das Gauß-Newton-Verfahren für das unseparierte Problem (5.1) zu formulieren:

Algorithmus G: Gauß-Newton-Verfahren für das unseparierte Problem

S1: Wähle Startpunkt $\tilde{\mathbf{t}}^{(0)} \in \mathbb{R}^l$, $\boldsymbol{\alpha}^{(0)} \in \mathbb{R}^n$

S2: For $k = 0, 1, \ldots$ do

S2.1: Konvergenztest

S2.2: Berechne

$$\begin{aligned} \mathbf{s}_{\tilde{t}}^{(k)} &:= -\left(\mathbf{P}_B^{\perp}\mathbf{J}_{\tilde{t}}\right)^+ \mathbf{P}_B^{\perp}\mathbf{F} \\ \mathbf{s}_{\alpha}^{(k)} &:= -\mathbf{B}^+\left(\mathbf{F} + \mathbf{J}_{\tilde{t}}\mathbf{s}_{\tilde{t}}^{(k)}\right) \end{aligned}$$

S2.3: Setze

$$\begin{aligned} \tilde{\mathbf{t}}^{(k+1)} &:= \tilde{\mathbf{t}}^{(k)} + \mathbf{s}_{\tilde{t}}^{(k)} \\ \boldsymbol{\alpha}^{(k+1)} &:= \boldsymbol{\alpha}^{(k)} + \mathbf{s}_{\alpha}^{(k)} \end{aligned}$$

Das separable Quadratmittelproblem (5.1) besitzt offensichtlich mehr Struktur als ein beliebiges nichtlineares Quadratmittelproblem. Das wesentliche Merkmal separabler Quadratmittelprobleme ist die Tatsache, daß für feste $\tilde{\mathbf{t}}_c \in \mathbb{R}^l$ die Lösung des *reduzierten Problems*

$$\min\left\{\frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}}_c)\boldsymbol{\alpha}\right\|^2 \,:\, \tilde{\mathbf{t}}_c \in \mathbb{R}^l \quad \text{fest}, \ \boldsymbol{\alpha} \in \mathbb{R}^n\right\} \tag{5.7}$$

sehr effizient berechnet werden kann. Das reduzierte Problem (5.7) ist ein lineares Quadratmittelproblem der Dimension n, welches mit Standardorthogonalisierungstechniken gelöst werden kann.

Eine naheliegende Veränderung des Gauß-Newton-Verfahrens besteht darin, abwechselnd über einer Variablengruppe zu minimieren, während die Variablen der anderen Gruppe konstant gehalten werden. Diese Methode ist in der Literatur als *Methode der alternierenden Variablen* bekannt.

Algorithmus 1: Methode der alternierenden Variablen

S1: Wähle Startpunkt $\tilde{\mathbf{t}}^{(0)} \in \mathbb{R}^l$

S2: For $k = 0, 1, \ldots$ do

S2.1: Konvergenztest

S2.2: Löse das lineare Quadratmittelproblem

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

um $\boldsymbol{\alpha}^{(k)}$ zu erhalten

S2.3: Berechne

$$\mathbf{s}_{\tilde{t}}^{(k)} := -\mathbf{J}_{\tilde{t}}^+\mathbf{F}$$

S2.4: Setze

$$\tilde{\mathbf{t}}^{(k+1)} := \tilde{\mathbf{t}}^{(k)} + \mathbf{s}_{\tilde{t}}^{(k)}$$

Bemerkung 21
Algorithmus 1 ist der Algorithmus NIPALS, *welcher von Wold und Lyttkens [WL69] stammt und unter Statistikern verbreitet ist.*

Betrachten wir nun das lineare Quadratmittelproblem in Schritt S2.2 in Algorithmus 1. Dieses lineare Quadratmittelproblem hat im allgemeinen, rankdefizienten Fall viele Lösungen. Wir wählen die eindeutige Normallösung

$$\boldsymbol{\alpha}_{opt} = \boldsymbol{\alpha}_{opt}(\tilde{\mathbf{t}}) = \mathbf{B}(\tilde{\mathbf{t}})^{+}\mathbf{y} \tag{5.8}$$

Falls die Matrix $\mathbf{B}$ Vollrang l hat, so erhalten wir

$$\mathbf{B}(\tilde{\mathbf{t}})^{+} = \left[\mathbf{B}(\tilde{\mathbf{t}})^T\mathbf{B}(\tilde{\mathbf{t}})\right]^{-1}\mathbf{B}(\tilde{\mathbf{t}})^T$$

Für jedes $\tilde{\mathbf{t}} \in \mathbb{R}^l$ erhalten wir den optimalen Parameter $\boldsymbol{\alpha}_{opt}(\tilde{\mathbf{t}})$ durch (5.8). Setzen wir diesen optimalen Parameter in das Ausgangsproblem

$$\min\left\{\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha}) := \frac{1}{2}\left\|\mathbf{F}\left(\tilde{\mathbf{t}},\boldsymbol{\alpha}\right)\right\|^2 = \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \;:\; \tilde{\mathbf{t}} \in \mathbb{R}^l,\, \boldsymbol{\alpha} \in \mathbb{R}^n\right\} \tag{5.9}$$

ein, so erhalten wir folgendes *reduzierte Problem*:

$$\min\left\{\phi(\tilde{\mathbf{t}}) := \frac{1}{2}\Big\|\underbrace{\left[\mathbf{I}-\mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^{+}\right]\mathbf{y}}_{=:\mathbf{G}(\tilde{\mathbf{t}})}\Big\|^2 = \frac{1}{2}\left\|\mathbf{G}(\tilde{\mathbf{t}})\right\|^2 \;:\; \tilde{\mathbf{t}} \in \mathbb{R}^l\right\} \tag{5.10}$$

Da $\mathbf{P}_B^{\perp} := \mathbf{I} - \mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^{+}$ der orthogonale Projektor auf *im* $\mathbf{B}(\tilde{\mathbf{t}})^{\perp}$ ist, wird das Funktional $\phi(\tilde{\mathbf{t}})$ auch als *variable projection functional* und die beschriebene Reduktionstechnik als *Verfahren der variablen Projektion* bezeichnet (siehe [GP73]).

Das reduzierte Problem (5.10) enthält nur noch die l nichtlinear auftretenden Parameter $\tilde{\mathbf{t}}$, ist dafür aber von einer größeren Komplexität als das Ausgangsproblem.

Für die Anwendung des Verfahrens der variablen Projektion bleibt die Zulässigkeit der Transformation zu zeigen. Außerdem muß für die Anwendung von Gauß-Newton-ähnlichen Verfahren die Jacobi-Matrix $\mathbf{G}'$ von $\mathbf{G}(\tilde{\mathbf{t}})$ berechnet werden.

5.3 Der Übergang zum reduzierten Funktional

Das folgende Theorem von Golub/Pereyra [GP73] rechtfertigt den Übergang vom Ausgangsfunktional $\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha})$ zum reduzierten Funktional $\phi(\tilde{\mathbf{t}})$.

Theorem 2 (Golub/Pereyra [GP73])
Seien $\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha})$ und $\phi(\tilde{\mathbf{t}})$ wie oben definiert. $\mathbf{B}(\tilde{\mathbf{t}})$ habe in der offenen Menge $\Omega \subset \mathbb{R}^l$ konstanten Rang $r \leq \min(m,n)$.

(a) *Wenn $\tilde{\mathbf{t}}^*$ ein kritischer Punkt (oder eine globale Minimumstelle in Ω) von $\phi(\tilde{\mathbf{t}})$ ist und*

$$\boldsymbol{\alpha}^* = \mathbf{B}(\tilde{\mathbf{t}}^*)^{+}\mathbf{y}, \tag{5.11}$$

so ist $(\tilde{\mathbf{t}}^*,\boldsymbol{\alpha}^*)$ ein kritischer Punkt (oder eine globale Minimumstelle für $\tilde{\mathbf{t}} \in \Omega$) von $\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha})$, und es gilt

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

(b) *Wenn* $(\tilde{\mathbf{t}}^*, \boldsymbol{\alpha}^*)$ *eine globale Minimumstelle von* $\psi(\tilde{\mathbf{t}}, \boldsymbol{\alpha})$ *für* $\tilde{\mathbf{t}} \in \Omega$ *ist, so ist* $\tilde{\mathbf{t}}^*$ *eine globale Minimumstelle von* $\phi(\tilde{\mathbf{t}})$ *in* Ω*, und es gilt*

$$\phi(\tilde{\mathbf{t}}^*) \quad = \quad \psi(\tilde{\mathbf{t}}^*, \boldsymbol{\alpha}^*).$$

Wenn es ein eindeutiges $\boldsymbol{\alpha}^*$ *unter den Minimumstellen von* $\psi(\tilde{\mathbf{t}}, \boldsymbol{\alpha})$ *gibt, so muß* $\boldsymbol{\alpha}^*$ *die Beziehung (5.11) erfüllen.*

Bemerkung 22
Da die Moore-Penrose-Inverse $\mathbf{B}^+$ *bei rangerhöhenden Störungen nicht stetig ist (siehe z.B. [KS88]), ist die Forderung nach einem lokal konstanten Rang der Matrix* $\mathbf{B}$ *natürlich. Wir machen daher im weiteren die Voraussetzung*

> *Die Matrix* $\mathbf{B}(\tilde{\mathbf{t}})$ *habe lokal konstanten Rang* $r \leq \min(m, n)$ *für alle* $\tilde{\mathbf{t}} \in \Omega \subset \mathbb{R}^l$*, wobei* Ω *eine offene Menge ist, welche die gesuchte Lösung enthält.*

5.4 Die Bestimmung der Jacobi-Matrix

Zur Berechnung der Jacobi-Matrix $\mathbf{G}'$ benötigen wir zunächst die Fréchet-Ableitung eines orthogonalen Projektors $\mathbf{P}_{A(\tilde{t})}$ einer differenzierbaren $m \times n$ Matrix-Funktion $\mathbf{A}(\tilde{\mathbf{t}})$ von lokal konstantem Rang r. Wir werden dabei zunächst nicht die Moore-Penrose-Inverse $\mathbf{A}^+$ und den zugehörigen Projektor $\mathbf{P}_A = \mathbf{A}\mathbf{A}^+$ betrachten, sondern eine verallgemeinerte Inverse $\mathbf{A}^-$, welche nicht alle vier Moore-Penrose-Bedingungen erfüllt. Der zugehörige Projektor $\mathbf{P}_A = \mathbf{A}\mathbf{A}^-$ wird dabei nicht geändert.

Lemma 2 (Golub/Pereyra [GP73])
Sei $\mathbf{A}^-(\tilde{\mathbf{t}})$ *eine* $n \times m$ *Matrixfunktion, so daß (P1)* $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ *und (P3)* $\left(\mathbf{A}\mathbf{A}^-\right)^T = \mathbf{A}\mathbf{A}^-$*. Dann gilt:*

$$\mathbf{D}\mathbf{P}_A = \mathbf{P}_A^\perp \left(\mathbf{D}\mathbf{A}\right) \mathbf{A}^- + \left(\mathbf{P}_A^\perp \left(\mathbf{D}\mathbf{A}\right) \mathbf{A}^-\right)^T$$

Beweis: Wegen $\mathbf{P}_A\mathbf{A} = \mathbf{A}$ gilt

$$\mathbf{D}\left(\mathbf{P}_A\mathbf{A}\right) \quad = \quad \left(\mathbf{D}\mathbf{P}_A\right)\mathbf{A} + \mathbf{P}_A\left(\mathbf{D}\mathbf{A}\right) = \mathbf{D}\mathbf{A},$$

d.h.

$$\left(\mathbf{D}\mathbf{P}_A\right)\mathbf{A} \quad = \quad \mathbf{D}\mathbf{A} - \mathbf{P}_A\left(\mathbf{D}\mathbf{A}\right) = \mathbf{P}_A^\perp\left(\mathbf{D}\mathbf{A}\right)$$

und daher wegen $\mathbf{P}_A = \mathbf{A}\mathbf{A}^-$ schließlich

$$(i) \qquad \left(\mathbf{D}\mathbf{P}_A\right)\mathbf{P}_A \quad = \quad \left(\mathbf{D}\mathbf{P}_A\right)\mathbf{A}\mathbf{A}^- = \mathbf{P}_A^\perp\left(\mathbf{D}\mathbf{A}\right)\mathbf{A}^-.$$

Wegen $(ii) \quad \left(\left(\mathbf{D}\mathbf{P}_A\right)\mathbf{P}_A\right)^T = \mathbf{P}_A\left(\mathbf{D}\mathbf{P}_A\right)$ erhalten wir aus (i) und (ii) letztendlich

$$\begin{aligned}\mathbf{D}\mathbf{P}_A = \mathbf{D}\left(\mathbf{P}_A^2\right) \quad &= \quad \left(\mathbf{D}\mathbf{P}_A\right)\mathbf{P}_A + \mathbf{P}_A\left(\mathbf{D}\mathbf{P}_A\right) \\ &= \quad \mathbf{P}_A^\perp\left(\mathbf{D}\mathbf{A}\right)\mathbf{A}^- + \left(\mathbf{P}_A^\perp\left(\mathbf{D}\mathbf{A}\right)\mathbf{A}^-\right)^T\end{aligned}$$

□

Bemerkung 23

- *Das Lemma gilt natürlich auch, wenn die verallgemeinerte Inverse* $\mathbf{A}^-$ *durch die Moore-Penrose-Inverse* $\mathbf{A}^+$ *ersetzt wird.*
- *Wegen* $\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{P}_A$ *gilt* $\mathbf{D}\mathbf{P}_A^\perp = -\mathbf{D}\mathbf{P}_A$.

Wir sind nun in der Lage, den Gradienten des reduzierten Funktionals

$$\begin{aligned}\phi(\tilde{\mathbf{t}}) &= \frac{1}{2}\left\|\mathbf{P}^{\perp}_{B(\tilde{t})}\mathbf{y}\right\|^2\\ &= \frac{1}{2}\left\langle \mathbf{P}^{\perp}_{B}\mathbf{y}, \mathbf{P}^{\perp}_{B}\mathbf{y}\right\rangle\end{aligned}$$

zu berechnen. Es gilt:

$$\begin{aligned}\nabla\phi(\tilde{\mathbf{t}}) &= \left\langle \mathbf{P}^{\perp}_{B}\mathbf{y}, \mathbf{D}\left(\mathbf{P}^{\perp}_{B}\mathbf{y}\right)\right\rangle\\ &= -\left\langle \mathbf{P}^{\perp}_{B}\mathbf{y}, \left(\mathbf{D}\mathbf{P}_{B}\right)\mathbf{y}\right\rangle\\ &= -\left\langle \mathbf{P}^{\perp}_{B}\mathbf{y}, \left[\mathbf{P}^{\perp}_{B}(\mathbf{DB})\mathbf{B}^{-} + \left(\mathbf{B}^{-}\right)^{T}(\mathbf{DB})^{T}\mathbf{P}^{\perp}_{B}\right]\mathbf{y}\right\rangle\\ &= -\mathbf{y}^{T}\mathbf{P}^{\perp}_{B}\left[\mathbf{P}^{\perp}_{B}(\mathbf{DB})\mathbf{B}^{-} + \left(\mathbf{B}^{-}\right)^{T}(\mathbf{DB})^{T}\mathbf{P}^{\perp}_{B}\right]\mathbf{y}\end{aligned}$$

Beachte: Diese Darstellung für den Gradienten des reduzierten Funktionals gilt für beliebige orthogonale Projektoren $\mathbf{P}^{\perp}_{B} = \mathbf{I} - \mathbf{BB}^{-}$, sofern $\mathbf{B}^{-}$ die Moore-Penrose-Bedingungen (P1) und (P3) erfüllt.

Wir nehmen jetzt an, daß $\mathbf{B}^{-}$ zusätzlich die Bedingung (P2) $\mathbf{B}^{-}\mathbf{BB}^{-} = \mathbf{B}^{-}$ erfüllt. Dann gilt:

$$\begin{aligned}\mathbf{P}^{\perp}_{B}\left(\mathbf{B}^{-}\right)^{T} &= \left[\mathbf{I} - \mathbf{BB}^{-}\right]\left(\mathbf{B}^{-}\right)^{T}\\ &= \left(\mathbf{B}^{-}\right)^{T} - \mathbf{BB}^{-}\left(\mathbf{B}^{-}\right)^{T}\\ &= \left(\mathbf{B}^{-}\right)^{T} - \left(\mathbf{BB}^{-}\right)^{T}\left(\mathbf{B}^{-}\right)^{T}\\ &= \left(\mathbf{B}^{-}\right)^{T} - \left(\mathbf{B}^{-}\mathbf{BB}^{-}\right)^{T} = 0\end{aligned}$$

Falls $\mathbf{B}^{-}$ die Moore-Penrose-Bedingungen (P1), (P2) und (P3) erfüllt, erhalten wir für den Gradienten des reduzierten Funktionals $\phi(\tilde{\mathbf{t}}) = \frac{1}{2}\left\|\mathbf{P}^{\perp}_{B(\tilde{t})}\mathbf{y}\right\|^2$ den Ausdruck

$$\nabla\phi(\tilde{\mathbf{t}}) = -\mathbf{y}^{T}\mathbf{P}^{\perp}_{B(\tilde{t})}\left(\mathbf{DB}(\tilde{\mathbf{t}})\right)\mathbf{B}(\tilde{\mathbf{t}})^{-}\mathbf{y}. \tag{5.12}$$

Kommen wir nun zur angekündigten Berechnung der Jacobi-Matrix $\mathbf{G}'$ der reduzierten Funktion $\mathbf{G}$.

$$\begin{aligned}\mathbf{G}(\tilde{\mathbf{t}}) &= \left[\mathbf{I} - \mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^{+}\right]\mathbf{y} = \mathbf{P}^{\perp}_{B(\tilde{\mathbf{t}})}\mathbf{y}\\ \mathbf{G}'(\tilde{\mathbf{t}}) &= \mathbf{D}\left(\mathbf{P}^{\perp}_{B}\mathbf{y}\right) = -\left(\mathbf{DP}_{B}\mathbf{y}\right)\\ &= -\mathbf{P}^{\perp}_{B}(\mathbf{DB})\mathbf{B}^{+}\mathbf{y} - \left(\mathbf{P}^{\perp}_{B}(\mathbf{DB})\mathbf{B}^{+}\right)^{T}\mathbf{y}\end{aligned}$$

Wir erhalten daher für beliebige $\mathbf{s} \in \mathbb{R}^{l}$

$$\mathbf{G}'(\tilde{\mathbf{t}})\mathbf{s} = \left\{\mathcal{A}(\tilde{\mathbf{t}})[\mathrm{s}] + \left(\mathcal{A}(\tilde{\mathbf{t}})[\mathrm{s}]\right)^{T}\right\}\mathbf{y} \tag{5.13}$$

$$\text{mit}\quad \mathcal{A}(\tilde{\mathbf{t}})[\mathrm{s}] := -\mathbf{P}^{\perp}_{B}\left(\mathbf{B}'(\tilde{\mathbf{t}})[\mathrm{s}]\right)\mathbf{B}(\tilde{\mathbf{t}})^{+} \tag{5.14}$$

$$= -\left[\mathbf{I} - \mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^{+}\right]\left(\mathbf{B}'(\tilde{\mathbf{t}})[\mathrm{s}]\right)\mathbf{B}(\tilde{\mathbf{t}})^{+} \tag{5.15}$$

Nun können wir das Gauß-Newton-Modell für das reduzierte Problem (5.10) bilden:

$$\mu_{GP}(\tilde{\mathbf{t}} + \mathbf{s}) := \frac{1}{2}\left\|\mathbf{G} + \mathbf{G}'\mathbf{s}\right\|^2 \tag{5.16}$$

Bemerkung 24

- *Im weiteren werden wir dieses Modell als Golub/Pereyra-Modell bezeichnen.*
- *In der Originalarbeit von Golub/Pereyra [GP73] wird anstatt mit der Pseudoinversen* $\mathbf{B}^+$ *mit einer symmetrischen verallgemeinerten Inversen* $\mathbf{B}^-$, *welche nur die Bedingungen (P1), (P2) und (P3) erfüllt, gearbeitet. Diese verallgemeinerte Inverse wird über eine QR-Transformation auf Trapezform berechnet (siehe etwa [GvL89]).*

Jetzt sind wir in der Lage, das Gauß-Newton-Verfahren für das reduzierte Problem (5.10) zu formulieren.

Algorithmus 2: Gauß-Newton-Verfahren für das reduzierte Problem

S1: Wähle Startpunkt $\tilde{\mathbf{t}}^{(0)} \in \mathbb{R}^l$

S2: For $k = 0, 1, \ldots$ do

S2.1: Konvergenztest, Falls Konvergenz goto S3;

S2.2: Berechne

$$\mathbf{s}_{\tilde{t}}^{(k)} := -\left\{-\left[\mathbf{P}_B^\perp(\mathbf{DB})\mathbf{B}^+ + \left(\mathbf{P}_B^\perp(\mathbf{DB})\mathbf{B}^+\right)^T\right]\mathbf{y}\right\}^+ \mathbf{P}_B^\perp \mathbf{y}$$

S2.3: Setze

$$\tilde{\mathbf{t}}^{(k+1)} := \tilde{\mathbf{t}}^{(k)} + \mathbf{s}_{\tilde{t}}^{(k)}$$

S3: Setze $\boldsymbol{\alpha}^* := \mathbf{B}(\tilde{\mathbf{t}})\mathbf{y}$

Bemerkung 25
Eine effiziente Implementierung dieses Verfahrens findet sich in Krogh [Kro74]. Dort wird angenommen, daß $\mathbf{B}$ *Vollrang hat.*

Sowohl in den Implementierungen von Golub/Pereyra als auch in der von Krogh muß der Tensor $\mathbf{DB}$ gebildet werden, da er einmal als $\mathbf{DB}$ als auch als $(\mathbf{DB})^T$ auftritt. Wie oben (5.12) gezeigt wurde, trägt der Term

$$\left(\mathcal{A}(\tilde{\mathbf{t}})\right)^T \mathbf{y} := -\left(\mathbf{P}_B^\perp(\mathbf{DB})\mathbf{B}^+\right)^T \mathbf{y}$$

jedoch gar nicht zum Gradienten $\nabla\phi(\tilde{\mathbf{t}})$ bei. Eine genaue Analyse zeigt, daß für das Golub/Pereyra-Modell (5.16) gilt

$$||\mathbf{G} + \mathbf{G}'\mathbf{s}||^2 = ||\mathbf{G} + \mathcal{A}[\mathbf{s}]\mathbf{y}||^2 + \mathbf{s}^T\mathbf{T}_G\mathbf{s}, \tag{5.17}$$

wobei die Matrix $\mathbf{T}_G$ gemäß

$$\mathbf{u}^T\mathbf{T}_G\mathbf{v} := \left\{(\mathcal{A}[\mathbf{u}])^T\mathbf{y}\right\}^T\left\{(\mathcal{A}[\mathbf{u}])^T\mathbf{y}\right\} \tag{5.18}$$

$$:= \mathbf{G}^T\left(\mathbf{B}'[\mathbf{u}]\right)\mathbf{B}^+\mathbf{B}^{+T}\left(\mathbf{B}'[\mathbf{v}]\right)^T\mathbf{G} \tag{5.19}$$

definiert ist. Der Beitrag des Terms $\left(\mathcal{A}(\tilde{\mathbf{t}})[\mathbf{s}]\right)^T\mathbf{y}$ zur Hesse-Matrix $\nabla^2\phi(\tilde{\mathbf{t}})$ des reduzierten Funktionals ist $\mathbf{T}_G$, dieser ist aber von der Größenordnung $\mathcal{O}\left(||\mathbf{G}||^2\right)$. Er ist somit für kleines $||\mathbf{G}||$ kleiner als der beim Gauß-Newton-Verfahren ohnehin weggelassene Term $\mathbf{S} = \mathbf{G} \circ \mathbf{G}''$. Entsprechend der Gauß-Newton-Philosophie wird der Term $\mathbf{T}_G$ daher ebenfalls vernachlässigt. Dies führt auf das vereinfachte Modell

$$\mu_K(\tilde{\mathbf{t}} + \mathbf{s}) := \frac{1}{2}||\mathbf{G} + \mathbf{J}\mathbf{s}||^2 \tag{5.20}$$

$$\text{mit} \quad \mathbf{J}\mathbf{s} := \mathcal{A}(\tilde{\mathbf{t}})[\mathbf{s}]\mathbf{y} = -\left[\mathbf{I} - \mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^+\right]\left(\mathbf{B}'(\tilde{\mathbf{t}})[\mathbf{s}]\right)\mathbf{B}^+\mathbf{y}, \tag{5.21}$$

welches nach Kaufman [Kau75] benannt ist.

Bemerkung 26
Die soeben beschriebene Vereinfachung wurde erstmals von Kaufman [Kau75] vorgenommen. Sie wird dort unter Benutzung einer QR-Transformation auf Trapezform hergeleitet. Eine direkte Herleitung, an welche sich die obige Darstellung anlehnt, findet sich in [RW80]. Der Kaufman-Zugang eignet sich nach [Bjö 90] besser zur Verallgemeinerung auf restringierte Probleme (siehe auch [KP78]).

Algorithmus 3: Gauß-Newton-Verfahren für das red. Problem (Kaufman-Modell)

S1: Wähle Startpunkt $\tilde{\mathbf{t}}^{(0)} \in \mathbb{R}^l$

S2: For $k = 0, 1, \ldots$ do

S2.1: Konvergenztest, Falls Konvergenz goto S3;

S2.2: Berechne

$$\mathbf{s}_{\tilde{t}}^{(k)} \quad := - \left\{ - \left[\mathbf{P}_B^{\perp} (\mathbf{DB}) \mathbf{B}^+ \right] \mathbf{y} \right\}^+ \mathbf{P}_B^{\perp} \mathbf{y}$$

S2.3: Setze

$$\tilde{\mathbf{t}}^{(k+1)} \quad := \quad \tilde{\mathbf{t}}^{(k)} + \mathbf{s}_{\tilde{t}}^{(k)}$$

S3: Setze $\boldsymbol{\alpha}^* := \mathbf{B}(\tilde{\mathbf{t}})\mathbf{y}$

5.5 Konvergenzraten und Aufwand

Ruhe/Wedin [RW80] haben die asymptotischen Konvergenzraten der vorgestellten Algorithmen untersucht und geben hinreichende Bedingungen für die lokale Konvergenz der ungedämpften Verfahren an. Sei

$$\epsilon \quad := \quad \frac{\left\| \mathbf{y} - \mathbf{B}\left(\tilde{\mathbf{t}}\right) \boldsymbol{\alpha} \right\|_2}{\left\| \mathbf{B}\left(\tilde{\mathbf{t}}\right) \boldsymbol{\alpha} \right\|_2} = \frac{\left\| \left[\mathbf{I} - \mathbf{B}\left(\tilde{\mathbf{t}}\right) \mathbf{B}\left(\tilde{\mathbf{t}}\right)^+ \right] \mathbf{y} \right\|_2}{\left\| \mathbf{B}\left(\tilde{\mathbf{t}}\right) \mathbf{B}\left(\tilde{\mathbf{t}}\right)^+ \mathbf{y} \right\|_2}$$

ein Maß für die relative Größe des Residiums im Lösungspunkt. Das Residium wird als klein betrachtet, wenn ϵ klein ist.
Für jede Fixpunktiteration

$$z_{k+1} := h(z_k), \quad \lim_{k \to \infty} z_k = z^*,$$

ist die R-Konvergenzrate als

$$R = \rho \left(h'(z^*) \right)$$

definiert (ρ – Spektralradius).

Satz 12 (asymptotische Konvergenzraten)
Die R-Konvergenzraten der Algorithmen hängen in folgender Weise von ϵ ab:

R_G	$= R_{II} + \mathcal{O}(\epsilon^2)$	*Gauß-Newton unsepariertes Problem*
R_I	$= \mathcal{O}(1)$	*Alternierende Variablen*
R_{II}	$= \mathcal{O}(\epsilon)$	*Golub/Pereyra-Modell*
R_{III}	$= R_{II} + \mathcal{O}(\epsilon^2)$	*Kaufman-Modell*

Jeder der vier Algorithmen ist also im allgemeinen R-linear konvergent. Wenn der Gauß-Newton-Algorithmus für das unseparierte Problem superlinear konvergiert, so konvergieren auch die Algorithmen II und III superlinear. Die Methode der alternierenden Variablen dagegen konvergiert stets nur R-linear.

In Kaufman [Kau75] und Ruhe/Wedin [RW80] werden die Algorithmen für freie Optimierungsprobleme verglichen und festgestellt, daß der Kaufman-Algorithmus mit weniger arithmetischen Operationen nicht mehr Iterationen als der ursprüngliche Algorithmus von Golub/Pereyra benötigt. Sie konvergieren beide schneller als der Gauß-Newton-Algorithmus für das unseparierte Problem. Dabei ist jedoch zu beachten, daß stets nur bestimmte Implementierungen der Algorithmen verglichen werden, und daß das Abschneiden der Algorithmen stark vom Aufwand zur Berechnung der Residiumfunktion abhängig ist.

Ein Vorteil der Algorithmen für separable Quadratmittelprobleme liegt in der Tatsache, daß keine Anfangswerte für die linearen Parameter benötigt werden. Eine eventuelle spezielle Struktur (Schwachbesetztheit) von $\mathbf{B}'$ (dies ist insbesondere bei der Approximation mit B-Splines der Fall) kann in den Algorithmen von Golub-Pereyra und Kaufman gut ausgenutzt werden.

6 Splineapproximation mit freien Knoten

6.1 Problemstellung

Gegeben seien die Meßdaten (x_i, y_i), $i = 1, \ldots, m$, mit

$$a \leq x_1 < \ldots < x_m \leq b$$

einer unbekannten glatten Funktion f. Die mit Fehlern behafteten Meßwerte

$$\begin{aligned} y_i &= f(x_i) + \epsilon_i, \qquad i = 1, \ldots, m \\ \boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)^T &= \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned}$$

sollen durch eine einfach zu berechnende Funktion s approximiert werden.
Wir wählen als Ansatzfunktionen die B-Splines $B_{1,k}, \ldots, B_{n,k}$ der Ordnung k bezüglich einer Knotenfolge $\mathbf{t} = (t_1, \ldots, t_{n+k})$. Dabei gelte $m \geq n$. Für die Knotenfolge $\mathbf{t}$ gelte

$$t_1 = \ldots = t_k = a < t_{k+1} \leq \ \ldots \ \leq t_n < b = t_{n+1} = \ldots = t_{n+k} \tag{6.1}$$

$$\text{und} \qquad t_{j+k} > t_j \qquad j = 1, \ldots, n. \tag{6.2}$$

Die Ansatzfunktion s lautet dann

$$s(x) := \sum_{j=1}^{n} B_{j,k}(x, \mathbf{t}) \alpha_j \, , \ \alpha_j \in \mathbb{R}.$$

Die Bezeichnung $B_{j,k}(x, \mathbf{t})$ deutet dabei an, daß die B-Splines von der Knotenfolge $\mathbf{t}$ abhängig sind.
Unter der Voraussetzung (6.2) hat der Splineraum

$$\mathcal{S}_{k,\mathbf{t}} := \left\{ \sum_{j=1}^{n} B_{j,k}(x, \mathbf{t}) \alpha_j \, , \ \alpha_j \in \mathbb{R} \right\}$$

die Dimension n. Für die „wahren" Meßwerte $\tilde{y}_i$ erhalten wir den Ansatz

$$\tilde{y}_i = \sum_{j=1}^{n} B_{j,k}(x_i, \mathbf{t}) \alpha_j$$

und mittels der diskreten l_2-Norm das Minimierungsproblem

$$\sum_{i=1}^{m} [y_i - \tilde{y}_i]^2 \longrightarrow \min$$

$$\sum_{i=1}^{m} \left[y_i - \sum_{j=1}^{n} B_{j,k}(x_i, \mathbf{t}) \alpha_j \right]^2 \longrightarrow \min . \tag{6.3}$$

Wir definieren die Beobachtungsmatrix

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

und erhalten mit den Splinekoeffizienten $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{R}^n$ die zu (6.3) äquivalente Matrixformulierung

$$\|\mathbf{y} - \mathbf{B}(\mathbf{t})\boldsymbol{\alpha}\|^2 \quad \longrightarrow \quad \min. \tag{6.4}$$

Die Matrix $\mathbf{B}(\mathbf{t})$ hat unter der Bedingung (6.2) genau dann Vollrang n, wenn die Whitney-Schoenberg-Bedingung (2.6)

$$t_j < x_{i_j} < t_{j+k}$$

erfüllt ist.

Die Splineknoten $\mathbf{t}$ *und* die Splinekoeffizienten $\boldsymbol{\alpha}$ sollen jetzt so bestimmt werden, daß $\|\mathbf{y} - \mathbf{B}(\mathbf{t})\boldsymbol{\alpha}\|^2$ möglichst klein wird. Um größere Flexibilität zu erreichen, wählen wir nur eine Teilmenge $\tilde{\mathbf{t}}$, die sogenannten *freien Knoten*, der Splineknoten $\mathbf{t}$ aus.

l sei die Anzahl der freien Knoten und

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

sei der Vektor der freien Knoten. $\mathbf{p} = (p(1), \ldots, p(l))^T$ ist ein Indexvektor, welcher auf die freien Knoten zeigt. Es gelte

$$k < p(1) < p(2) < \ldots p(l) < n + 1. \tag{6.5}$$

Bemerkung 27

- *Die Bedingung (6.5) besagt, daß nur die inneren Knoten $t_{k+1}, \ldots, t_n$ freie Knoten sein können. Diese Beschränkung erscheint wegen der besonderen Wahl der Knotenfolge (6.1) natürlich, da Informationen über das Meßintervall $[a, b]$ zumeist vorliegen.*
- *Wegen (6.5) gilt $l \leq n - k$.*

An Stelle des Minimierungsproblems (6.4) erhalten wir jetzt

$$\left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \quad \longrightarrow \quad \min. \tag{6.6}$$

Die freien Splineknoten $\tilde{\mathbf{t}}$ können jedoch nicht beliebig gewählt werden. Es muß stets gelten

$$t_{j+k} \quad > \quad t_j \qquad j = 1, \ldots, n, \tag{6.7}$$

oder bei stärkeren Glattheitsforderungen sogar

$$t_{j+k-r} \quad > \quad t_j \qquad j = r + 1, \ldots, n. \tag{6.8}$$

Diese Nebenbedingungen für die Splineknoten werden wir im nächsten Abschnitt formulieren.

6.2 Anordnungsbedingung für die Splineknoten

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

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

Da bei der Minimierung von (6.6) nur die freien Knoten $\tilde{\mathbf{t}}$ verändert werden, lautet diese Bedingung

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

Für die numerische Praxis ist diese Bedingung ungünstig ((6.7) bzw. (6.8) wird nicht erfüllt!). Wir fordern daher

$$\begin{aligned} t_{p(j)} - t_{p(j)-1} &\geq \epsilon & & (6.9)\\ t_{p(j)+1} - t_{p(j)} &\geq \epsilon \qquad j = 1, \ldots, l. & & (6.10) \end{aligned}$$

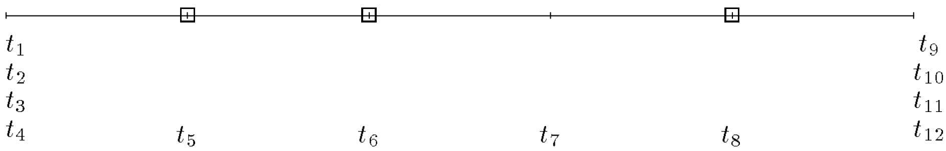
mit einer vorgegebenen Konstanten $0 < \epsilon \ll 1$.

Die Bedingungen (6.9), (6.10) können äquivalent in Matrixform angegeben werden

$$\mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h} \tag{6.11}$$

mit $\mathbf{C} \in \mathbb{R}^{ncstr,l}$, $\tilde{\mathbf{t}} \in \mathbb{R}^{l}$, $\mathbf{h} \in \mathbb{R}^{ncstr}$. $ncstr$ ist die Anzahl der linear unabhängigen Nebenbedingungen. 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: $k = 4$, $n = 8$, $l = 3$, $p(1) = 5$, $p(2) = 6$, $p(3) = 8$



$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{pmatrix} t_5 \\ t_6 \\ t_8 \end{pmatrix} \geq \begin{pmatrix} t_4 + \epsilon \\ \epsilon \\ -t_7 + \epsilon \\ t_7 + \epsilon \\ -t_9 + \epsilon \end{pmatrix}$$

$$\mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h} \qquad \mathbf{C} \in \mathbb{R}^{5,3}, \mathbf{h} \in \mathbb{R}^{5}$$

Bemerkung 28

- *Es gilt* $l + 1 \leq ncstr \leq 2l$.
- *Prinzipiell läßt sich auch*

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

  mit $2l$ *Konstanten* $\underline{\epsilon}_j, \overline{\epsilon}_j$ *fordern. Das Problem liegt dann in der Wahl geeigneter Konstanten. Deshalb haben wir im vorliegenden Programmpaket lediglich (6.9) und (6.10) mit einer, in Abhängigkeit von der durchschnittlichen Größe der Knoten zu wählenden Konstante* ϵ *realisiert.*
- *Die Matrix* $\mathbf{C}$ *enthält maximal zwei Nichtnullelemente in jeder Zeile.*
- *Mit der Wahl eines beliebig kleinen, aber positiven* ϵ *ist – zumindest theoretisch in rundungsfehlerfreier Arithmetik – die höchstmögliche Glattheit des Splines an den Stellen* $t_{p(j)}$, $j = 1, \ldots, l$, *sichergestellt, da die freien Knoten dann einfache Knoten sind.*

6.3 Untersuchung der Struktur des Problems

Jetzt können wir das vollständige Minimierungsproblem formulieren.

$$\min\left\{\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha}) := \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \;:\; \mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}\,,\, \tilde{\mathbf{t}} \in \mathbb{R}^l, \boldsymbol{\alpha} \in \mathbb{R}^n\right\} \tag{6.12}$$

Bei diesem Problem handelt es sich um ein separables Quadratmittelproblem mit linearen Ungleichheitsnebenbedingungen. Die Splinekoeffizienten $\boldsymbol{\alpha} \in \mathbb{R}^n$ treten linear auf, die freien Splineknoten $\tilde{\mathbf{t}} \in \mathbb{R}^l$ treten nichtlinear auf.

Wenn die Whitney-Schoenberg-Bedingung erfüllt ist, hat die Beobachtungsmatrix $\mathbf{B}$ Vollrang n und (6.12) liefert für jedes feste zulässige $\tilde{\mathbf{t}} \in \mathbb{R}^l$ die eindeutige Lösung $\boldsymbol{\alpha}_{opt}(\tilde{\mathbf{t}}) = \mathbf{B}(\tilde{\mathbf{t}})^+\mathbf{y}$.

6.3.1 Der Übergang zum reduzierten Funktional beim freien Problem

Wir betrachten zunächst das freie Optimierungsproblem

$$\min\left\{\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha}) := \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \;:\; \tilde{\mathbf{t}} \in \mathbb{R}^l, \boldsymbol{\alpha} \in \mathbb{R}^n\right\}. \tag{6.13}$$

Wir setzen voraus, daß die Whitney-Schoenberg-Bedingung überall erfüllt ist. Theorem 2 von Golub/Pereyra rechtfertigt dann den Übergang zum reduzierten Funktional $\phi(\tilde{\mathbf{t}})$

$$\min\left\{\phi(\tilde{\mathbf{t}}) := \frac{1}{2}\Big\|\underbrace{\left[\mathbf{I}-\mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^+\right]\mathbf{y}}_{=:\mathbf{G}(\tilde{\mathbf{t}})}\Big\|^2 = \frac{1}{2}\left\|\mathbf{G}(\tilde{\mathbf{t}})\right\|^2 \;:\; \tilde{\mathbf{t}} \in \mathbb{R}^l\right\}. \tag{6.14}$$

6.3.2 Der Übergang zum reduzierten Funktional beim restringierten Problem

Die Zulässigkeit des Überganges zum reduzierten Funktional bezieht sich bis jetzt nur auf freie Quadratmittelprobleme. Bei dem zu lösenden Problem (6.12) handelt es sich aber um ein separables Quadratmittelproblem mit linearen Ungleichheitsnebenbedingungen an die nichtlinearen Variable $\tilde{\mathbf{t}}$.

Kaufman hat in [GP76] gezeigt, daß dieser Übergang ebenfalls zulässig ist.

Satz 13 (Kaufman, zitiert nach [Par85])
Ein Problem der Form

$$\min\left\{\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha}) := \frac{1}{2}\left\|\mathbf{y}-\mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 \;:\; \mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}\,,\, \tilde{\mathbf{t}} \in \mathbb{R}^l, \boldsymbol{\alpha} \in \mathbb{R}^n\right\}$$

ist äquivalent mit der Lösung von

$$\min\left\{\phi(\tilde{\mathbf{t}}) := \frac{1}{2}\left\|\left[\mathbf{I}-\mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^+\right]\mathbf{y}\right\|^2 = \frac{1}{2}\left\|\mathbf{P}_B^\perp(\tilde{\mathbf{t}})\right\|^2 \;:\; \mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}\,,\, \tilde{\mathbf{t}} \in \mathbb{R}^l\right\}$$

gefolgt von

$$\boldsymbol{\alpha}^* \;=\; \mathbf{B}(\tilde{\mathbf{t}})^+\mathbf{y}.$$

Bemerkung 29

- *Die Aussage des Satzes ist anschaulich klar, da die variable Projektion ja bezüglich der linearen Variablen $\boldsymbol{\alpha}$ erfolgt, und nicht bezüglich der restringierten nichtlinearen Variablen $\tilde{\mathbf{t}}$.*

- *Den Fall von nichtlinearen Gleichungsnebenbedingungen an* $\tilde{\mathbf{t}}$ *untersuchen [KP78]. Das Problem läßt sich dann in ein verallgemeinertes separables Quadratmittelproblem überführen und mit den Methoden aus Kapitel 5 lösen.*

  Zusätzliche Nebenbedingungen auch an die linearen Variablen erschweren die algorithmische Lösung sehr. Diese allgemeinen Probleme wurden in [Par85] untersucht. Dort findet sich auch eine Klassifizierung nach den verschiedenen Typen von Nebenbedingungen.

Wir können nun an Stelle von (6.12) das restringierte, reduzierte Problem

$$\min\left\{\phi(\tilde{\mathbf{t}}) := \frac{1}{2}\big\|\underbrace{[\mathbf{I}-\mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^{+}]\,\mathbf{y}}_{=:\mathbf{G}(\tilde{\mathbf{t}})}\big\|^2 = \frac{1}{2}\big\|\mathbf{G}(\tilde{\mathbf{t}})\big\|^2 \;:\; \mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}\,,\, \tilde{\mathbf{t}} \in \mathbb{R}^l\right\} \tag{6.15}$$

betrachten. Die effektive Lösung von (6.15) ist Gegenstand des restlichen Kapitels. Dabei sind die folgenden Besonderheiten zu beachten und – soweit möglich – auszunutzen.

- $\mathbf{B}$ hat stets Vollrang n und besitzt eine Bandbreite kleiner oder gleich k
- $\mathbf{B}'$ hat ebenfalls eine Bandbreite kleiner oder gleich k
- $\mathbf{B}(\tilde{\mathbf{t}})^{+}\mathbf{y}$ kann sehr effektiv mittels zeilenweiser Givens-Drehungen berechnet werden
- $\mathbf{C}$ hat maximal zwei Nichtnullelemente in jeder Zeile

Grundlage für die Lösung von (6.15) ist ein Verallgemeinertes Gauß-Newton-Verfahren.

6.4 Existenz einer Lösung des reduzierten Problems

Satz 14 (Existenz einer Lösung des reduzierten Problems)
Die Knotenfolge $\mathbf{t}$ *erfülle überall die Whitney-Schoenberg-Bedingung und es gelten (6.9), (6.10). Dann besitzt das reduzierte Problem (6.15) eine Lösung* $\tilde{\mathbf{t}}^*$.

Beweis: Die Beobachtungsmatrix und die Moore-Penrose-Inverse sind unter den gestellten Voraussetzungen stetige – sogar differenzierbare – Funktionen der Splineknoten. Das stetige Funktional $\phi(\tilde{\mathbf{t}})$ nimmt daher über der abgeschlossenen und beschränkten ($a \leq t_{p(1)},\, t_{p(l)} \leq b$) Menge $\{\tilde{\mathbf{t}} \in \mathbb{R}^l \,:\, \mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}\}$ nach dem Satz von Weierstraß ihr Minimum an. □

Die Charakterisierung und Berechnung einer eindeutigen Bestapproximation $\tilde{\mathbf{t}}^*$ ist ein schwieriges Problem. Wir beschränken uns aus praktischen Gesichtspunkten daher auf die Bestimmung von lokal guten Approximationen. Unser Ziel ist es, die Lage der Knoten $\tilde{\mathbf{t}}$ so zu verbessern, daß wir mit weniger, aber gut plazierten Knoten die notwendige Approximationsgüte erreichen. Es können also solange Knoten entfernt werden, bis die Approximationsgüte in der Größenordnung der Standardabweichung der Meßdaten ist.

6.5 Die Berechnung der Jacobi-Matrix

Zur Anwendung eines Verallgemeinerten Gauß-Newton-Verfahrens zur Lösung von (6.15) benötigen wir die Jacobi-Matrix $\mathbf{G}' \in \mathbb{R}^{m,l}$ der reduzierten Funktion $\mathbf{G}$:

$\mathbb{R}^l \longrightarrow \mathbb{R}^m$ mit $\mathbf{G}(\tilde{\mathbf{t}}) := \left[\mathbf{I} - \mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^+\right]\mathbf{y}$. Die reduzierte Funktion $\mathbf{G}(\tilde{\mathbf{t}})$ ist unter den gestellten Vorausssetzungen – Whitney-Schoenberg-Bedingung erfüllt, Knotenfolge genügt (6.9),(6.10) – differenzierbar und es gilt

$$\begin{aligned} \mathbf{G}'(\tilde{\mathbf{t}})\mathbf{s} &= \left\{\mathcal{A}(\tilde{\mathbf{t}})[\mathbf{s}] + \left(\mathcal{A}(\tilde{\mathbf{t}})[\mathbf{s}]\right)^T\right\}\mathbf{y} \\ \text{mit}\quad \mathcal{A}(\tilde{\mathbf{t}})[\mathbf{s}] &:= -\left[\mathbf{I} - \mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^+\right]\left(\mathbf{B}'(\tilde{\mathbf{t}})[\mathbf{s}]\right)\mathbf{B}(\tilde{\mathbf{t}})^+. \end{aligned}$$

Die Jacobi-Matrix $\mathbf{G}'$ wird in dem entwickelten Programmpaket auf zwei unterschiedliche Arten approximiert.

6.5.1 Die Berechnung der Jacobi-Matrix im Golub/Pereyra-Modell

Im Golub/Pereyra-Modell wird das reduzierte Funktional $\phi(\tilde{\mathbf{t}} + \mathbf{s})$ durch die quadratische Modellfunktion

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

ersetzt. Wegen des Auftretens des Tensors $\mathcal{A}(\tilde{\mathbf{t}})^T$ eignet sich dieses Modell nur bedingt für die numerische Approximation der Ableitungen durch Differenzenquotienten. Wir approximieren daher direkt

$$\mathbf{G}'(\tilde{\mathbf{t}})\mathbf{e}^k \approx \frac{\mathbf{G}(\tilde{\mathbf{t}} + h_k\mathbf{e}^k) - \mathbf{G}(\tilde{\mathbf{t}})}{h_k}$$

Diese Approximation benötigt $l + 1$ Funktionsberechnungen von $\mathbf{G}$.

6.5.2 Die Berechnung der Jacobi-Matrix im Kaufman-Modell

Im Kaufman-Modell wird das reduzierte Funktional $\phi(\tilde{\mathbf{t}} + \mathbf{s})$ durch die quadratische Modellfunktion

$$\begin{aligned} \mu_K(\tilde{\mathbf{t}} + \mathbf{s}) &:= \frac{1}{2}\left\|\mathbf{G}(\tilde{\mathbf{t}}) + \mathbf{J}'(\tilde{\mathbf{t}})\mathbf{s}\right\|^2 \\ \text{mit}\quad \mathbf{G}'(\tilde{\mathbf{t}})\mathbf{s} \approx \mathbf{J}(\tilde{\mathbf{t}})\mathbf{s} &:= \left\{\mathcal{A}(\tilde{\mathbf{t}})[\mathbf{s}]\right\}\mathbf{y} \\ &= -\left[\mathbf{I} - \mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^+\right]\left(\mathbf{B}'(\tilde{\mathbf{t}})[\mathbf{s}]\right)\mathbf{B}(\tilde{\mathbf{t}})^+\mathbf{y} \end{aligned}$$

ersetzt. Die Approximation $\mathbf{J} \in \mathbb{R}^{m,l}$ der Jacobi-Matrix $\mathbf{G}'$ wird spaltenweise berechnet. Wir erhalten

$$\begin{aligned} \mathbf{J}\mathbf{e}^k &= -\left[\mathbf{I} - \mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^+\right]\left(\mathbf{B}'(\tilde{\mathbf{t}})[\mathbf{e}^k]\right)\mathbf{B}(\tilde{\mathbf{t}})^+\mathbf{y} \\ \mathbf{B}(\tilde{\mathbf{t}}) &= B_j\left(x_i, \tilde{\mathbf{t}}\right)\mathbf{e}^i \circ \mathbf{e}^j \qquad i = 1, \ldots, m; j = 1, \ldots, n \\ \mathbf{B}'(\tilde{\mathbf{t}}) &= \frac{\partial B_j\left(x_i, \tilde{\mathbf{t}}\right)}{\partial t_{p(k)}}\mathbf{e}^i \circ \mathbf{e}^j \circ \mathbf{e}^k \qquad k = 1, \ldots, l \\ \mathbf{B}'(\tilde{\mathbf{t}})\left[\mathbf{e}^k\right] &= \frac{\partial B_j\left(x_i, \tilde{\mathbf{t}}\right)}{\partial t_{p(k)}}\delta_{k,k}\,\mathbf{e}^i \circ \mathbf{e}^j. \end{aligned}$$

Hierbei haben wir von der Einsteinschen Summenkonvention und der Darstellung eines Tensors als dyadisches Produkt Gebrauch gemacht.
Die Matrix $\mathbf{B}'(\tilde{\mathbf{t}})\left[\mathbf{e}^k\right]$ wird numerisch über Differenzenquotienten berechnet.

$$\left(\mathbf{B}'(\tilde{\mathbf{t}})\left[\mathbf{e}^k\right]\right)_{i,j} \approx \frac{B_j\left(x_i, \tilde{\mathbf{t}} + h_k\mathbf{e}^k\right) - B_j\left(x_i, \tilde{\mathbf{t}}\right)}{h_k}$$

Man beachte dabei die Blockbandstruktur von $\mathbf{B}$!

Bemerkung 30

- *Ausdrücke der Form* $\mathbf{B}(\tilde{\mathbf{t}})^{+}\mathbf{u}$ *werden stets über die QR-Faktorisierung von* $\mathbf{B}(\tilde{\mathbf{t}})$ *berechnet und niemals über die direkte Bildung von* $\mathbf{B}(\tilde{\mathbf{t}})^{+}$.
- $\mathbf{B}(\tilde{\mathbf{t}})$ *wird nur einmal faktorisiert. Die Drehungsparameter werden dann abgespeichert. Die wiederholte Berechnung von* $\mathbf{B}(\tilde{\mathbf{t}})\mathbf{u}$ *für verschiedene* $\mathbf{u}$ *kann dann sehr effektiv erfolgen.*
- *Die Approximation von* $\mathbf{B}'(\tilde{\mathbf{t}})\left[\mathbf{e}^{k}\right]$ *wird sehr effektiv durchgeführt, da der verwendete Algorithmus zur Berechnung von B-Splines simultan alle nichtverschwindenen B-Splines an der Stelle* x_i *berechnet.*

6.6 Ein Verfahren zur Lösung des Problems

Jetzt können wir ein Verfahren zur Lösung des restringierten, reduzierten Problems

$$\min\left\{\phi(\tilde{\mathbf{t}}) := \frac{1}{2}\Big\|\underbrace{\left[\mathbf{I}-\mathbf{B}(\tilde{\mathbf{t}})\mathbf{B}(\tilde{\mathbf{t}})^{+}\right]\mathbf{y}}_{=:\mathbf{G}(\tilde{\mathbf{t}})}\Big\|^{2} = \frac{1}{2}\left\|\mathbf{G}(\tilde{\mathbf{t}})\right\|^{2} \,:\, \mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}\,,\, \tilde{\mathbf{t}} \in \mathbb{R}^{l}\right\} \tag{6.16}$$

angeben. Wir verwenden ein gedämpftes Verallgemeinertes Gauß-Newton-Verfahren mit der in 4.2.5 beschriebenen Regularisierung. Wir formulieren das Verfahren noch einmal mit den aktuellen Bezeichnungen.

Gedämpftes Verallgemeinertes Gauß-Newton-Verfahren mit Regularisierung

S1: Wähle (zulässigen) Startwert $\tilde{\mathbf{t}}^{0} \in \mathbb{R}^{l}$, $\alpha \in (0,1)$, $\delta \in (0, \frac{1}{2})$

Setze $k := 0$; $\gamma_{-1} := 1$;

S2: Iteration

Repeat

S2.1: $\mathbf{G} := \mathbf{G}(\tilde{\mathbf{t}}^{k}) = \left[\mathbf{I}-\mathbf{B}(\tilde{\mathbf{t}}^{k})\mathbf{B}(\tilde{\mathbf{t}}^{k})^{+}\right]\mathbf{y}$; Berechnung des Residiums
$\mathbf{J} :\approx \mathbf{G}'(\tilde{\mathbf{t}}^{k})$; Berechnung einer Approximation der Jacobi-Matrix (Golub/Pereyra- oder Kaufman-Modell)

S2.2: Bestimme Abstiegsrichtung $\mathbf{s}^{k}$ aus

$$\begin{array}{rccc} & \|\mathbf{G}+\mathbf{J}\mathbf{s}\|^{2} & \longrightarrow & \min \\ \text{mit} & \mathbf{C}\left(\tilde{\mathbf{t}}^{k}+\mathbf{s}\right) & \geq & \mathbf{h} \end{array}$$

Falls das Problem schlecht konditioniert ist, löse das regularisierte Problem

$$\begin{array}{rccc} & \left\|\begin{pmatrix}\mathbf{G}\\ \mathbf{0}\end{pmatrix}+\begin{bmatrix}\mathbf{J}\\ \sqrt{\lambda}\mathbf{I}\end{bmatrix}\mathbf{s}\right\|^{2} & \longrightarrow & \min \\ \text{mit} & \mathbf{C}\left(\tilde{\mathbf{t}}^{k}+\mathbf{s}\right) & \geq & \mathbf{h} \end{array}$$

S2.3: $\tilde{\gamma}_k := \gamma_k$; $\tilde{\mathbf{t}}^{+} := \tilde{\mathbf{t}}^{k} + \tilde{\gamma}_k\mathbf{s}^{k}$; $\mathbf{G}^{+} := \mathbf{G}(\tilde{\mathbf{t}}^{+})$;

S2.4: Abstiegstest

if $\frac{1}{2}\|\mathbf{G}\|^2 - \frac{1}{2}\left\|\mathbf{G} + \tilde{\gamma}_k \mathbf{J}\mathbf{s}^k\right\|^2 \geq -\delta\tilde{\gamma}_k \mathbf{G}^T\mathbf{J}\mathbf{s}^k$
 then
  $\tilde{\mathbf{t}}^{k+1} := \tilde{\mathbf{t}}^+;$
  $\gamma_{k+1} := \min(\gamma_k/\alpha, 1);$
  $k := k+1;$
 else
  $\gamma_k := \gamma_k * \alpha;$ goto S2.3;

until $(k > kmax)$ or Konvergenz

Bemerkung 31

- *Die Lösung des linearen Quadratmittelproblems mit linearen Ungleichheitsnebenbedingungen in Schritt S2.2 erfolgt mittels des Algorithmus LSI (siehe 3.2).*
- *Die Feststellung, wann ein Problem in S2.2 in unserem Sinne schlecht konditioniert ist, sowie die Wahl des Regularisierungsparameters λ ist in 4.2.5 beschrieben.*
- *Für das Abbruchkriterium gilt das in 4.1.2 zum Gauß-Newton-Verfahren gesagte. Insbesondere ist bei restringierten Problemen der, bei freien Problemen mit $\left\|\mathbf{G}'(\tilde{\mathbf{t}}^k)\mathbf{G}(\tilde{\mathbf{t}}^k)\right\| \leq \varepsilon_3^a$ äquivalente, Test $\left(\left|\mathbf{G}(\tilde{\mathbf{t}}^k)^T\mathbf{G}'(\tilde{\mathbf{t}}^k)\mathbf{s}\right| \leq \varepsilon_4^a\right)$ wichtig. Er stellt fest, ob noch eine zulässige Abstiegsrichtung* $\mathbf{s}$ *gefunden werden kann.*
- *Für die numerischen Tests wählen wir $\delta = 0.2$ und $\alpha = 0.2$.*

6.7 Testbeispiele

Dieser Abschnitt erhält einige Beispiele aus der Literatur sowie ein selbstgewähltes Beispiel, welche die Leistungsfähigkeit des entwickelten Verfahrens zeigen sollen. Alle Testrechnungen wurden auf einem 486-PC in IEEE double Arithmetik mit einer relativen Maschinengenauigkeit von $macheps \approx 2.22E-16$ durchgeführt. Die Abbruchkonstanten[1] (siehe 4.1.2) wurden wie folgt gewählt:

$$\begin{aligned}\varepsilon_0^a &= 1.0E-06\\ \varepsilon_1^a &= 1.0E-04\\ \varepsilon_3^a &= 1.0E-10\\ \varepsilon_4^a &= 1.0E-10\end{aligned}$$

Zum Vergleich des entwickelten Verfahrens mit Standardoptimierungsverfahren führen wir noch eine Minimierung des reduzierten Funktionals mit der MATLAB-Prozedur CONSTR durch. Die Abbruchkonstanten wurden auf die vorgegebenen Standardwerte gesetzt.

[1] Die entwickelten Verallgemeinerten Gauß-Newton-Verfahren liefern folgende Return Codes:

1	$\left\|\mathbf{G}\right\| \leq \varepsilon_0^a$
2	$\left\|\mathbf{J}^T\mathbf{G}\right\| \leq \varepsilon_3^a$
3	$\left\lvert\mathbf{G}^T\mathbf{J}\mathbf{p}\right\rvert \leq \varepsilon_4^a$
4	$\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| \leq \varepsilon_1^a \left(\left\|\mathbf{x}^k\right\| + \varepsilon_2^a\right)$
5	$k > kmax$
6	failed otherwise

6.7.1 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. $m = 49$ 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. Zunächst wählen wir die $l = 7$ inneren Splineknoten äquidistant im Intervall $[595, 1075]$ verteilt, was angesichts des Peaks sicher keine sehr gute Wahl ist. [2]
Die Konstante ϵ, welche den minimalen Abstand zwischen den freien Splineknoten angibt, wurde auf $\epsilon = 0.0$ für das Kaufman-Modell und $\epsilon = 0.1$ für die MATLAB-Routine gesetzt.

Bemerkung 32
Die theoretische Durchführbarkeit ist zwar nur für den Fall $\epsilon > 0.0$ gesichert. Die von uns entwickelte Routine ist aber relativ unempfindlich gegenüber der Wahl von ϵ, während die MATLAB-Routine bei zu kleinem ϵ wegen der Rundungsfehler abbricht.

	Startpunkt	MATLAB-CONSTR	Kaufman-Modell
t_5	655	656.648608	603.122987
t_6	715	722.545037	805.055831
t_7	775	843.889596	866.128923
t_8	835	874.837667	866.128923
t_9	895	897.730378	897.867208
t_{10}	955	916.990605	916.774224
t_{11}	1015	972.317384	973.773457
$\|\|\mathbf{r}\|\|$	$7.924657E-01$	$7.370720E-02$	$5.651648E-02$
Schritte		484 Func. calls	24
Zeit		12.037 MFlops	0.56 sec
$\|\mathbf{G}^T\mathbf{J}\mathbf{s}\|$			$2.30011E-11$
$\|\|\mathbf{J}^T\mathbf{G}\|\|$			$3.98777E-08$
Return Code		erfolgreich	3

Ein erfahrener Nutzer würde natürlich die Splineknoten geschickter wählen. Wir beginnen daher die Iteration noch einmal mit einer *guten* Knotenfolge.

	Startpunkt	MATLAB-CONSTR	Kaufman-Modell
t_5	750	802.354053	802.345327
t_6	825	865.366264	859.461095
t_7	885	865.366264	879.534951
t_8	890	905.535024	879.534951
t_9	920	905.635024	910.350046
t_{10}	935	934.584027	910.350046
t_{11}	975	969.940811	974.653448
$\|\|\mathbf{r}\|\|$	$1.054318E-01$	$5.875918E-02$	$3.934182E-02$
Schritte		704 Func. calls	17
Zeit		17.706 MFlops	0.40 sec
$\|\mathbf{G}^T\mathbf{J}\mathbf{s}\|$			$5.93399E-11$
$\|\|\mathbf{J}^T\mathbf{G}\|\|$			$1.49747E-08$
Return Code		Abbruch	3

[2] Die Splineknoten t_j werden in den folgenden Abbildungen durch Rechtecke, die Meßdaten (x_i, y_i) durch Kreise dargestellt.

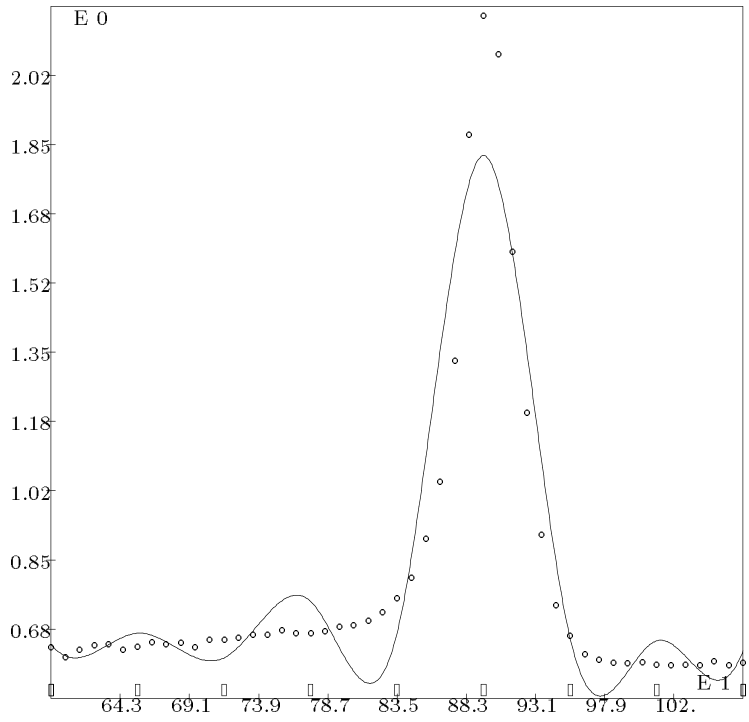


Abbildung 1: Titanium Heat Data, $k = 4$, $n = 11$, equidistant knots

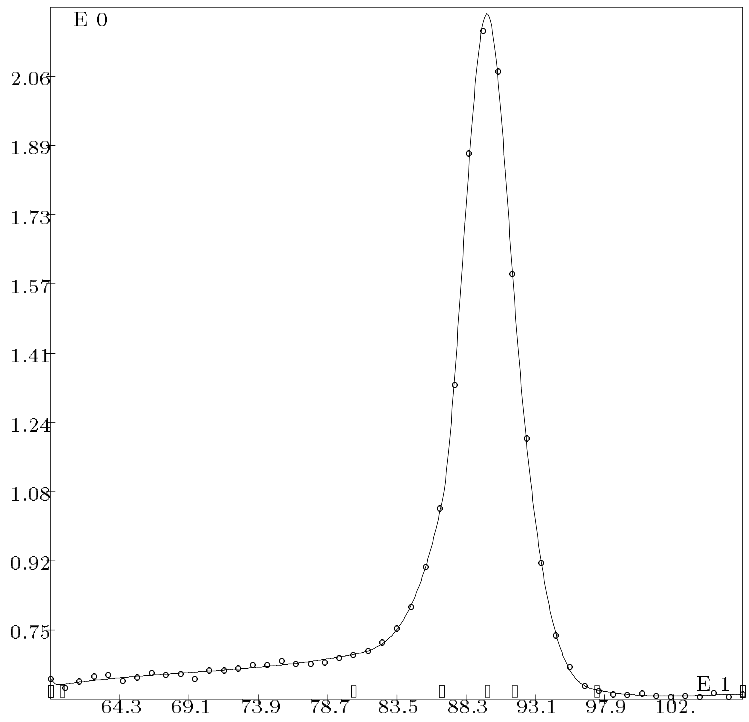


Abbildung 2: Titanium Heat Data, $k = 4$, $n = 11$, $l = 7$, Kaufman-Model

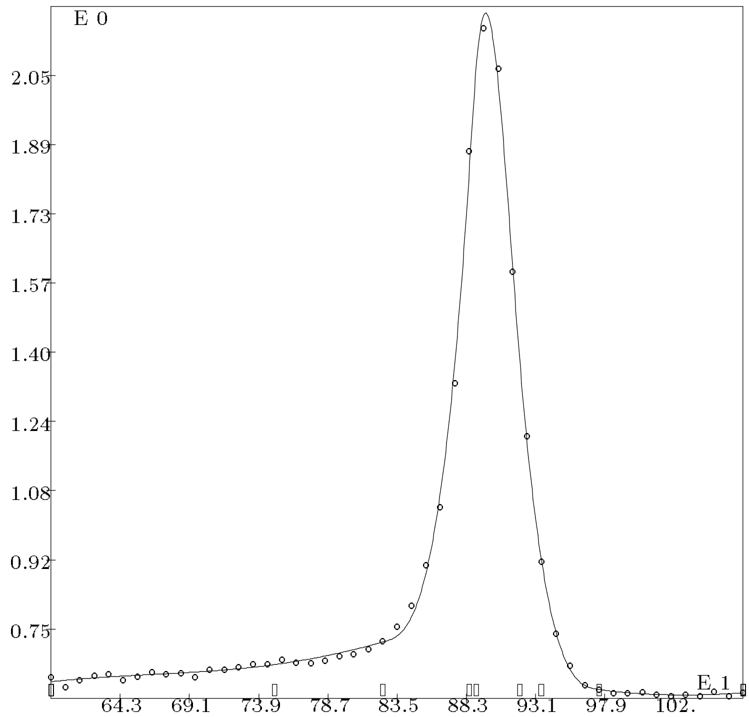


Abbildung 3: Titanium Heat Data, $k = 4$, $n = 11$, $l = 7$, good startpoint

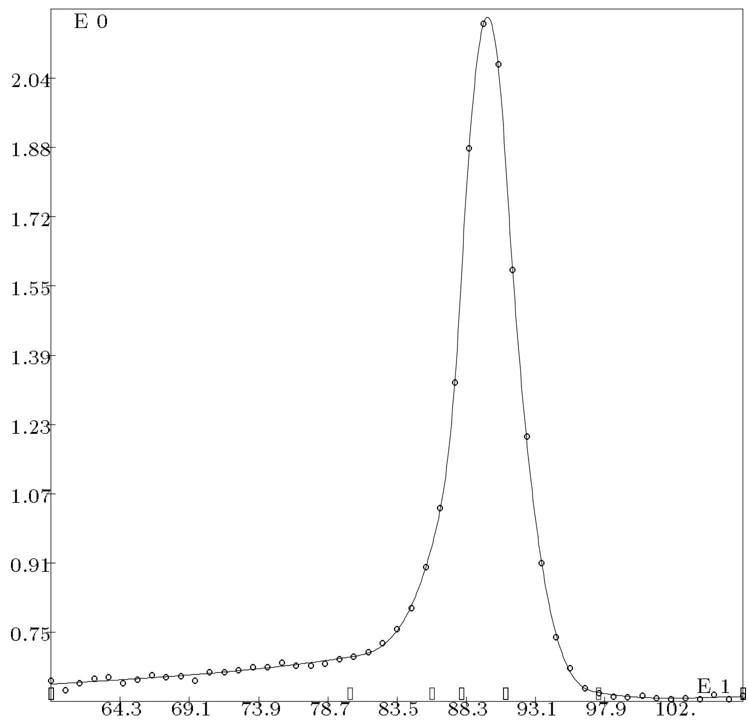


Abbildung 4: Titanium Heat Data, $k = 4$, $n = 11$, $l = 7$, Kaufman-Model

6.7.2 Titan-Daten II

Ein Motiv für die Entwicklung des vorgestellten Algorithmus war die bessere Plazierung der freien Knoten, um anschließend gegebenfalls die Anzahl der Knoten zu reduzieren. Diese Knotenreduktion möchten wir mit dem folgenden Beispiel demonstrieren.

Wir approximieren erneut die Titan-Daten. Diesmal wählen wir allerdings $n = 17$ B-Splines der Ordnung $k = 4$. Wir starten die Approximation wieder mit äquidistanten inneren Knoten ($l = 13$).

	Startpunkt	MATLAB-CONSTR	Golub/Pereyra-Modell
t_5	629.285714	626.746755	595.017590
t_6	663.571428	662.988340	595.020012
t_7	697.857142	697.976877	694.798680
t_8	732.142857	732.369775	740.838863
t_9	766.428571	766.966956	826.246672
t_{10}	800.714285	803.464577	863.392120
t_{11}	835.000000	866.069225	863.459128
t_{12}	869.285714	866.169225	863.597505
t_{13}	903.571428	897.564015	897.564233
t_{14}	937.857142	917.369575	917.109542
t_{15}	972.142857	967.743774	971.160283
t_{16}	1006.428571	1005.407904	994.053308
t_{17}	1040.714285	1040.315564	1066.665752
$\|\|\mathbf{r}\|\|$	$3.237391E-01$	$5.468452E-02$	$5.235772E-02$
Schritte		924 Func. calls	39
Zeit		26.912 MFlops	1.65 sec
$\|\mathbf{G}^T\mathbf{J}\mathbf{s}\|$			$9.25177E-11$
$\|\|\mathbf{J}^T\mathbf{G}\|\|$			$4.89054E-08$
Return Code		erfolgreich	3

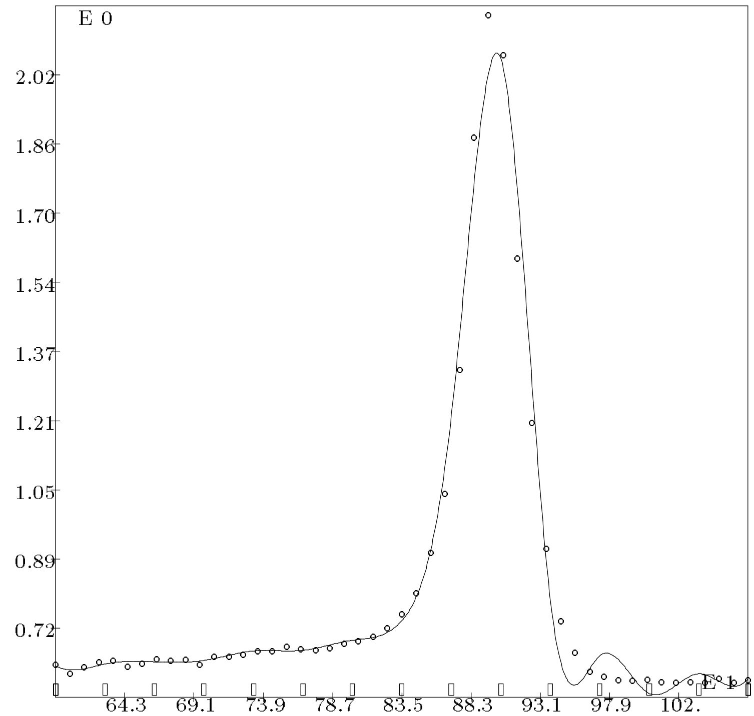


Abbildung 5: Titanium Heat Data, $k = 4$, $n = 17$, equidistant knots

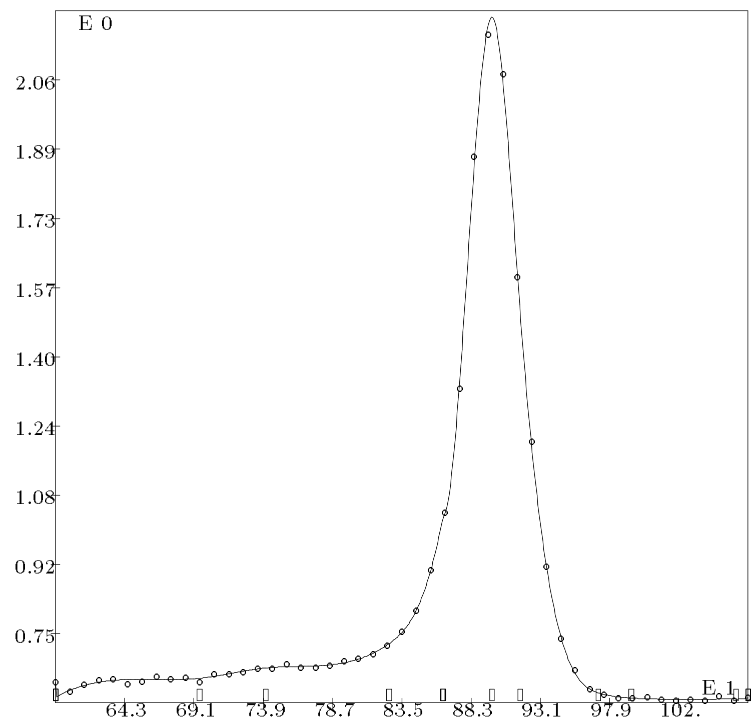


Abbildung 6: Titanium Heat Data, $k = 4$, $n = 17$, $l = 13$, Golub/Pereyra-Model

Wir modifizieren jetzt die Knotenfolge, welche wir als Ergebnis der Optimierung mit dem Golub/Pereyra-Modell erhalten haben.

Knoten	Modifikation
t_5, t_6	entfernen
t_{10}, t_{11}, t_{12}	ein einfacher Knoten $\frac{t_{10}+t_{11}+t_{12}}{3}$
t_{15}, t_{17}	entfernen

Statt $n = 17$ erhalten wir $n = 11$ Knoten. Auf diese neue Knotenfolge wenden wir erneut das Minimierungsverfahren an.

	Startpunkt	Golub/Pereyra-Modell
t_5	694.798680	604.372097
t_6	740.838863	805.055833
t_7	826.246672	866.128905
t_8	863.482918	866.128905
t_9	897.564233	897.867279
t_{10}	917.109542	916.774162
t_{11}	994.053308	973.773496
$\|\|\mathbf{r}\|\|$	$2.863939E-01$	$5.651648E-02$
Schritte		18
Zeit		0.50 sec
$\|\|\mathbf{G}^T\mathbf{J}\mathbf{s}\|\|$		$2.86444E-11$
$\|\|\mathbf{J}^T\mathbf{G}\|\|$		$4.44755E-08$
Return Code		3

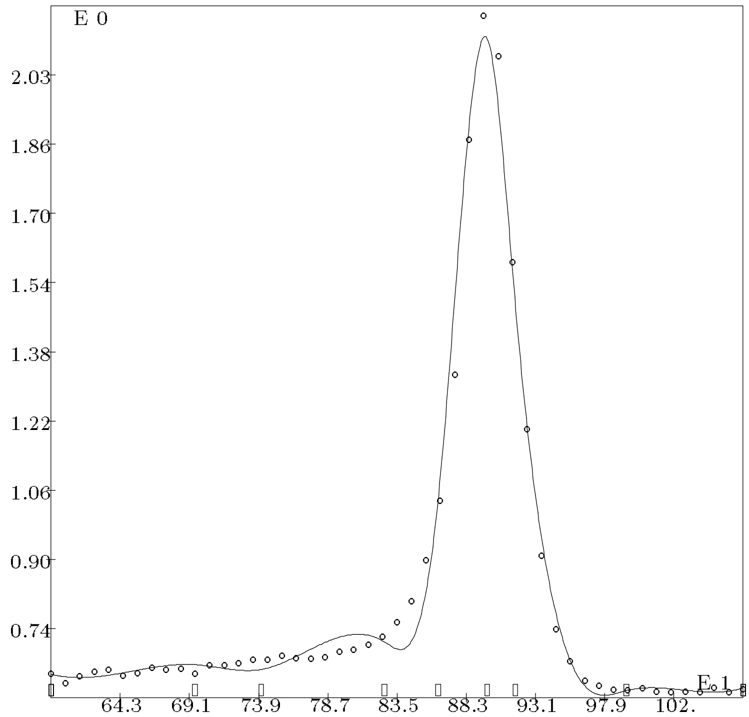


Abbildung 7: Titanium Heat Data, $k = 4$, $n = 11$, Number of knots reduced

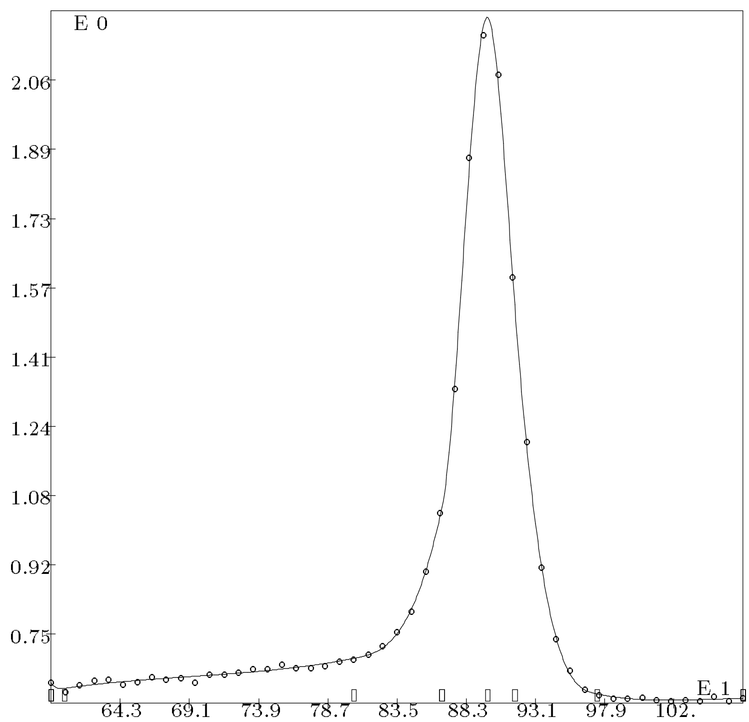


Abbildung 8: Titanium Heat Data, $k = 4$, $n = 11$, $l = 7$, Golub/Pereyra-Model

6.7.3 Spektraldaten

Wir betrachten Meßwerte, welche durch die Spektralfunktion

$$f(x) \quad = \quad 7 \times \exp\left(-\frac{(x-5)^2}{0.25}\right) - 5 \times \exp\left(-(x+3)^2\right) + 0.1 \times x$$

beschrieben werden. Mittels Pseudozufallszahlen ϵ_i, $-0.05 \leq \epsilon_i \leq 0.05$, erzeugen wir die Werte

$$y_i \quad = \quad f(x_i) + \epsilon_i$$

an $m = 90$ äquidistanten Punkten im Intervall $[-10, 10]$. Für den verwendeten Meßwertsatz gilt $E(\boldsymbol{\epsilon}) = 8.3E - 04$ und $D^2(\boldsymbol{\epsilon}) = 3.0E - 03$.

Bemerkung 33
Wenn man den physikalischen Hintergrund der Daten kennt, ist natürlich das ebenfalls separable Parameterschätzproblem mit der Ansatzfunktion

$$r(x, \tilde{\mathbf{t}}, \boldsymbol{\alpha}) \quad = \quad \sum_{j=1}^{k} \alpha_j \exp\left(-\left(\frac{x - \tilde{t}_j}{\tilde{t}_{j+k}}\right)^2\right) + \alpha_{k+1} + \alpha_{k+2}x + \alpha_{k+3}x^2$$

der Problemstellung besser angepaßt. Wir stellen uns jedoch auf den Standpunkt, daß nichts über den physikalischen Hintergrund bekannt ist.

Diese Daten wollen wir durch $n = 25$ B-Splines der Ordnung $k = 5$ approximieren. Wir wählen $l = 20$ äquidistante innere Splineknoten als Ausgangspunkt der Iteration. Wir verzichten diesmal auf die numerische Angabe der berechneten Knoten.

	Startpunkt	Kaufman-Modell	Golub/Pereyra-Modell
$\|\|\mathbf{r}\|\|$	$3.796930E+00$	$2.143237E-01$	$1.445334E+00$
Schritte		34	50
Zeit		5.01 sec	8.94 sec
$\|\mathbf{G}^T\mathbf{J}\mathbf{s}\|$		$8.77864E-11$	$1.06799E-03$
$\|\|\mathbf{J}^T\mathbf{G}\|\|$		$4.22577E-08$	$4.53193E+00$
Return Code		3	5

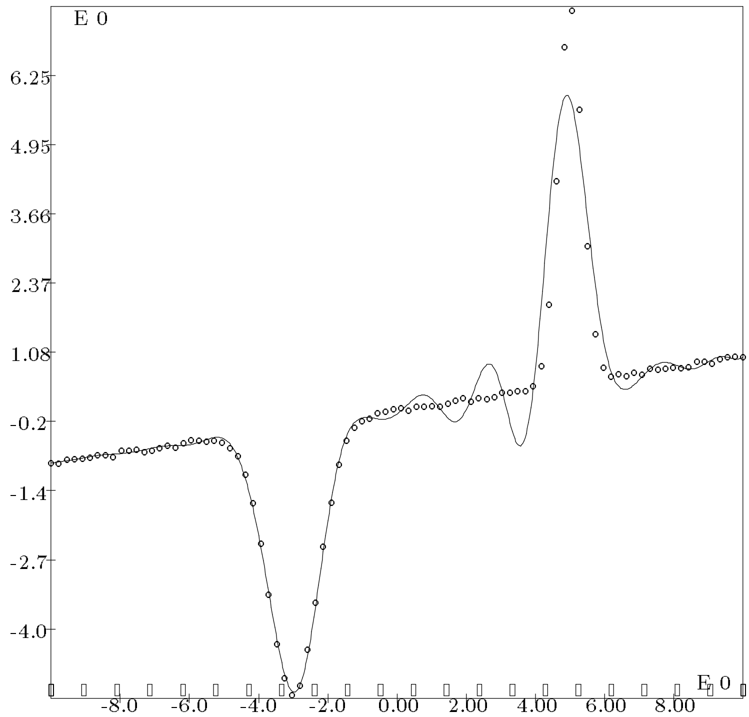


Abbildung 9: Spectral Data, $k = 5$, $n = 25$, equidistant knots

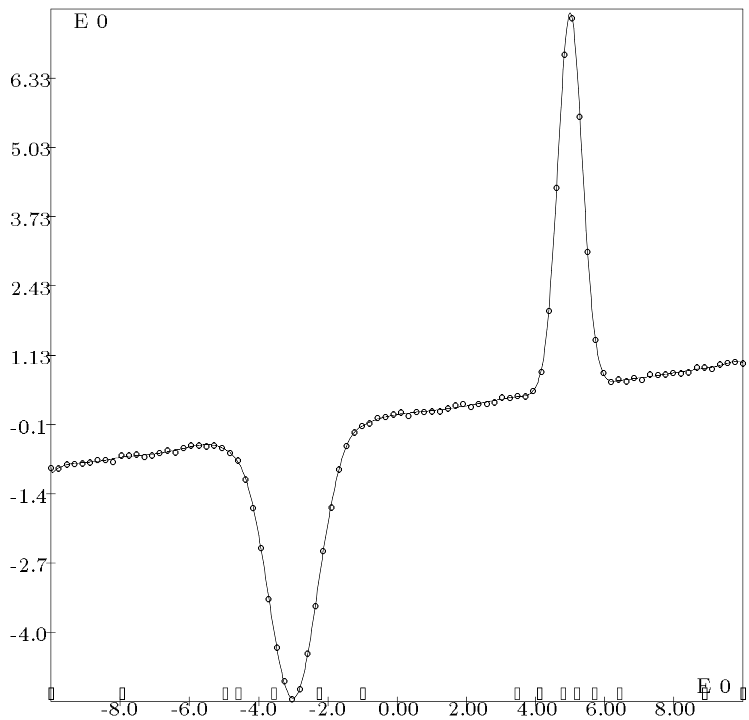


Abbildung 10: Spectral Data, $k = 5$, $n = 25$, $l = 20$, Kaufman-Model

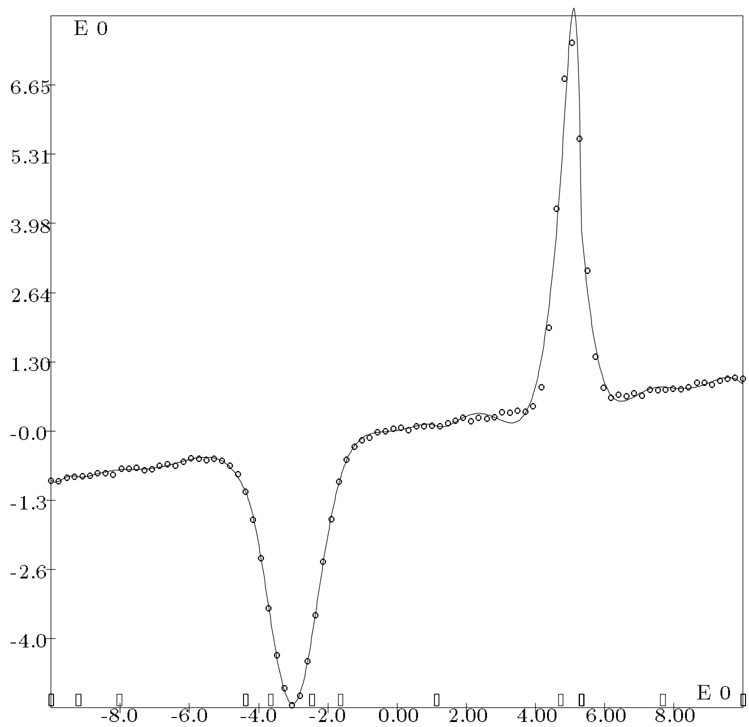


Abbildung 11: Spectral Data, $k = 5$, $n = 25$, $l = 20$, Golub/Pereyra-Model

7 Splineglättung mit freien Knoten

Bei der praktischen Anwendung von Approximationsschemata ist oftmals ein Kompromiß zwischen der Güte der Approximation und der Glattheit gefragt. Neben dem *Approximationsterm*

$$\varphi(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) \quad := \quad \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2 \tag{7.1}$$

betrachten wir jetzt zusätzlich einen sogenannten *Glättungsterm*

$$\bar{\rho}(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) \quad := \quad \left\| s^{(r)} \right\|_{L_2}^2 = \int_a^b \left[s^{(r)}(x) \right]^2 dx \tag{7.2}$$

mit festem $r < k$ und das zugehörige Minimierungsproblem

$$\varphi(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) + \mu \bar{\rho}(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) \quad = \quad \left\| \mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha} \right\|^2 + \mu \int_a^b \left[s^{(r)}(x) \right]^2 dx \longrightarrow \min. \tag{7.3}$$

Der *Glättungsparameter* $\mu > 0$ liefert im Grenzfall $\mu \to +0$ die Approximation mit kleinster Seminorm φ und im Grenzfall $\mu \to +\infty$ die beste polynomiale Approximation der Ordnung r. Die Bestimmung des Glättungsparameters μ kann in den Optimierungsprozeß einbezogen werden (siehe [Wah90] für die entsprechenden Verfahren und den statistischen Hintergrund). Wir werden dagegen in dieser Arbeit den Glättungsparameter interaktiv durch den Nutzer vorgeben lassen.

Wir werden zunächst den Glättungsterm (7.2) in Abhängigkeit von den Splinekoeffizienten $\boldsymbol{\alpha}$ und den Knoten $\mathbf{t}$ ausdrücken. Anschließend wird dieser exakte Glättungsterm durch eine einfacher zu berechnende Approximation ersetzt. Die Darstellung in den beiden folgenden Abschnitten sowie die Algorithmen zur Berechnung eines glättenden Splines folgen [Kun89] und [SK93].

7.1 Darstellung des Glättungsterms

Für die Knotenfolge $\mathbf{t}$ gelte

$$t_{j+k-r} \quad > \quad t_j \qquad j = r+1, \ldots, n.$$

Dann existiert $s^{(r)}(x)$ und besitzt die Darstellung

$$\begin{aligned}
s^{(r)}(x) \quad &= \quad \sum_{j=r+1}^{n} B_{j,k-r}(x) \alpha_j^{(r)} \\
\text{mit} \quad \boldsymbol{\alpha}^{(r)} \quad &:= \quad \mathbf{D}_r \boldsymbol{\alpha} \\
\mathbf{D}_r \quad &:= \quad \mathbf{H}_r \mathbf{L}_r \ldots \mathbf{H}_1 \mathbf{L}_1 \in \mathbb{R}^{n-r,n} \\
\mathbf{H}_\nu \quad &:= \quad (k-\nu)\, diag \left(\frac{1}{t_{k+j-\nu} - t_j} \right)_{j=1+\nu,\ldots,n} \in \mathbb{R}^{n-\nu,n-\nu} \\
\mathbf{L}_\nu \quad &:= \quad \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n-\nu,n-\nu+1} \quad \nu = 1, \ldots, r.
\end{aligned}$$

Setzt man dies in den Glättungsterm $\bar{\rho}$ ein, so erhält man

$$
\begin{aligned}
\bar{\rho} &= \int_a^b \left[\sum_{i=r+1}^{n} B_{i,k-r}(x)\alpha_i^{(r)} \right] \left[\sum_{j=r+1}^{n} B_{j,k-r}(x)\alpha_j^{(r)} \right] dx \\
&= \sum_{i,j=r+1}^{n} \alpha_i^{(r)}\alpha_j^{(r)} \int_a^b B_{i,k-r}(x)B_{j,k-r}(x)\,dx \\
&= \boldsymbol{\alpha}^{(r)^T} \bar{\mathbf{M}}_r \boldsymbol{\alpha}^{(r)} \\
\text{mit} \quad \bar{\mathbf{M}}_r &= \left(\bar{m}_{ij}^{(r)}\right) \in \mathbb{R}^{n-r,n-r} \\
\bar{m}_{ij}^{(r)} &:= \int_a^b B_{i,k-r}(x)B_{j,k-r}(x)\,dx \quad i,j = r+1,\ldots,n.
\end{aligned}
$$

Bemerkung 34

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

Wir führen die Cholesky-Faktorisierung der symmetrischen, positiv definiten Matrix $\bar{\mathbf{M}}_r$ durch.

$$
\begin{aligned}
\bar{\mathbf{M}}_r &= \bar{\mathbf{F}}_r^T \bar{\mathbf{F}}_r \\
\bar{\mathbf{F}}_r \in \mathbb{R}^{n-r,n-r} & \quad \text{reguläre obere Dreiecksmatrix der Bandbreite } k-r
\end{aligned}
$$

Damit erhält man für den Glättungsterm

$$
\begin{aligned}
\bar{\rho} &= \boldsymbol{\alpha}^{(r)^T} \bar{\mathbf{M}}_r \boldsymbol{\alpha}^{(r)} \\
&= \boldsymbol{\alpha}^T \mathbf{D}_r^T \bar{\mathbf{F}}_r^T \bar{\mathbf{F}}_r \mathbf{D}_r \boldsymbol{\alpha} \\
&= \left\| \bar{\mathbf{S}}_r \boldsymbol{\alpha} \right\|^2
\end{aligned}
$$

mit der Matrix

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

$\bar{\mathbf{S}}_r$ ist eine obere Dreiecksmatrix und besitzt die Bandbreite k.
Die Berechnung der Elemente $\bar{m}_{ij}$ der Gramschen Matrix ist für $k-r > 2$ sehr teuer. Sie kann etwa, wie von [Sch81] empfohlen, über Gaußsche Quadraturformeln erfolgen. Wir verwenden eine billigere Approximation des Glättungstermes, welche sich in den Anwendungen (siehe [SK93]) als günstig erwiesen hat.

7.2 Approximation des Glättungsterms

Wir ersetzen die L_2-Norm in der Definition des Glättungsterms durch ihr diskretes Äquivalent im Splineraum $\mathcal{S}_{k-r,\mathbf{t}}$.

$$
\tilde{\rho}(\tilde{\mathbf{t}}, \boldsymbol{\alpha}) \quad := \quad \left\| s^{(r)} \right\|_{l_2}^2 = \sum_{j=r+1}^{n} \left(\alpha_j^{(r)}\right)^2 \frac{t_{j+k-r} - t_j}{k-r} \tag{7.4}
$$

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:

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

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_2-Norm durch die diskrete Norm l_p wird ebenfalls in [LM88] und [ADLM90] vorgeschlagen.
Für diese Approximation erhalten wir

$$\begin{aligned}\tilde{\rho} &= \boldsymbol{\alpha}^{(r)^T}\tilde{\mathbf{F}}_r^T\tilde{\mathbf{F}}_r\boldsymbol{\alpha}^{(r)} \\ &= \boldsymbol{\alpha}^T\mathbf{D}_r^T\tilde{\mathbf{F}}_r^T\tilde{\mathbf{F}}_r\mathbf{D}_r\boldsymbol{\alpha}\end{aligned}$$

mit $\tilde{\mathbf{F}}_r := diag\left(\sqrt{\frac{t_{j+k-r}-t_j}{k-r}}\right)_{j=r+1,\ldots,n}$, und schließlich

$$\tilde{\rho} = \left\|\tilde{\mathbf{S}}_r\boldsymbol{\alpha}\right\|^2$$

mit der oberen Dreiecksmatrix

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

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.

7.3 Formulierung des kompletten Funktionals

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

$$\begin{aligned}\varphi(\tilde{\mathbf{t}},\boldsymbol{\alpha}) + \mu\rho(\tilde{\mathbf{t}},\boldsymbol{\alpha}) &= \left\|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{t}})\boldsymbol{\alpha}\right\|^2 + \mu\left\|\mathbf{S}_r\boldsymbol{\alpha}\right]^2 &(7.5)\\ &= \left\|\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix} - \begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}\right\|^2 \longrightarrow \min. &(7.6)\end{aligned}$$

Zusammen mit den Anordnungsbedingungen $\mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}$ ergibt sich das separable Quadratmittelproblem

$$\min\left\{\psi(\tilde{\mathbf{t}},\boldsymbol{\alpha}) := \frac{1}{2}\left\|\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix} - \begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}(\tilde{\mathbf{t}})\end{bmatrix}\boldsymbol{\alpha}\right\|^2 : \mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}\,,\, \tilde{\mathbf{t}} \in \mathbb{R}^l, \boldsymbol{\alpha} \in \mathbb{R}^n\right\}. \tag{7.7}$$

Wir setzen im folgenden voraus, daß

$$m \geq r. \tag{7.8}$$

Dann folgt aus $\mathbf{B}\boldsymbol{\alpha} = 0$ und $\mathbf{S} = \left\{\begin{matrix}\bar{\mathbf{S}}_r\\ \tilde{\mathbf{S}}_r\end{matrix}\right\} = 0$, daß $\boldsymbol{\alpha} = 0$, d.h. die Matrix $\begin{bmatrix}\mathbf{B}\\ \sqrt{\mu}\mathbf{S}\end{bmatrix}$ hat für jedes $\mu > 0$ Vollrang und (7.5) besitzt eine eindeutige Lösung $\boldsymbol{\alpha}_{opt}(\tilde{\mathbf{t}})$.

Bemerkung 35
Die Bedingung (7.8) ersetzt gewissermaßen die Schoenberg-Whitney-Bedingung. Die, im Minimierungsprozeß schwer zu realisierende, Forderung nach einer Vollrangmatrix $\mathbf{B}$ *entfällt durch die Regularisierung (7.8) und wird durch die, in unseren Anwendungen immer erfüllte, Bedingung* $m \geq r$ *ersetzt.*

Unter der Voraussetzung (7.8) können wir zum reduzierten, restringierten Problem

$$\begin{aligned}\min\Bigg\{\phi(\tilde{\mathbf{t}}) &:= \frac{1}{2}\left\|\left\{\begin{bmatrix}\mathbf{I}_m\\ \mathbf{I}_{n-r}\end{bmatrix} - \begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}(\tilde{\mathbf{t}})\end{bmatrix}\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}(\tilde{\mathbf{t}})\end{bmatrix}^+\right\}\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}\right\|^2 \\ &= \frac{1}{2}\left\|\mathbf{G}(\tilde{\mathbf{t}})\right\|^2 \,:\, \mathbf{C}\tilde{\mathbf{t}} \geq \mathbf{h}\,,\, \tilde{\mathbf{t}} \in \mathbb{R}^l\Bigg\}\end{aligned} \tag{7.9}$$

mit der differenzierbaren Abbildung $\mathbf{G} : \mathbb{R}^l \longrightarrow \mathbb{R}^{m+n-r}$

$$\mathbf{G}(\tilde{\mathbf{t}}) := \left\{\begin{bmatrix}\mathbf{I}_m\\ \mathbf{I}_{n-r}\end{bmatrix} - \begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}(\tilde{\mathbf{t}})\end{bmatrix}\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}(\tilde{\mathbf{t}})\end{bmatrix}^+\right\}\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix} \tag{7.10}$$

übergehen.
Dieses Problem kann mit dem gedämpften Verallgemeinerten Gauß-Newton-Verfahren aus Abschnitt 6 gelöst werden, indem man dort $\mathbf{B}(\tilde{\mathbf{t}})$ durch $\begin{bmatrix}\mathbf{B}(\tilde{\mathbf{t}})\\ \sqrt{\mu}\mathbf{S}(\tilde{\mathbf{t}})\end{bmatrix}$ und $\mathbf{y}$ durch $\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}$ ersetzt.
Die Berechnung der Jacobi-Matrix im Kaufman-Modell ist wegen der komplizierten Gestalt von $\mathbf{S}(\tilde{\mathbf{t}})$ schwierig. Da die Testrechnungen im Fall der Quadratmittelapproximation gezeigt haben, daß Golub/Pereyra-Modell und Kaufman-Modell nahezu gleichwertig sind, verwenden wir die einfacher zu berechnende Approximation

$$\mathbf{G}'(\tilde{\mathbf{t}})\mathbf{e}^k \approx \frac{\mathbf{G}(\tilde{\mathbf{t}} + h_k\mathbf{e}^k) - \mathbf{G}(\tilde{\mathbf{t}})}{h_k}$$

nach dem Golub/Pereyra-Modell.
Diese Approximation benötigt $l+1$ Funktionsberechnungen von $\mathbf{G}$. Es ist also lediglich die Berechnung eines glättenden Splines sicherzustellen. Dazu verwenden wir ein existierendes Verfahren aus [Kun89]. Bei diesem Verfahren wird zuerst $\mathbf{B}$ mittels zeilenweiser GIVENS-Drehungen in eine obere Dreiecksmatrix $\mathbf{R}_B$ transformiert. Im nächsten Schritt wird die erweiterte Matrix $\begin{bmatrix}\mathbf{R}_B\\ \sqrt{\mu}\mathbf{S}\end{bmatrix}$ auf obere Dreiecksform gebracht. Dazu wird die Matrix mit Permutationen auf Standardform gebracht und mittels GIVENS-Drehungen faktorisiert. Die Bandbreite k bleibt bei diesem Prozeß erhalten (siehe auch [Eld84]).

7.4 Testbeispiele

Ein Verfahren zur Berechnung eines glättenden Splines mit freien Knoten wurde unter Verwendung der Approximation des Glättungsterms mittels des Golub/Pereyra-Modells implementiert. Die Aussagen zur Wahl der Abbruchkonstanten des Verallgemeinerten Gauß-Newton-Verfahrens aus Abschnitt 6 gelten weiterhin. Bei der Berechnung eines glättenden Splines ist jedoch darauf zu achten, daß $t_j < t_{j+k-r}$, $j = r+1, \ldots, n$ gilt. Wir wählen deshalb für die Konstante ϵ, welche den minimalen Abstand zwischen den freien Knoten angibt, den Wert $\epsilon = 0.1$.

Für kleine Werte des Glättungsparameters $\mu > 0$ erhält man annähernd die gleichen Ergebnisse wie in Abschnitt 6.7. Der Aufwand ist dabei natürlich wegen der größeren Dimension von $\mathbf{G}$ entsprechend größer.

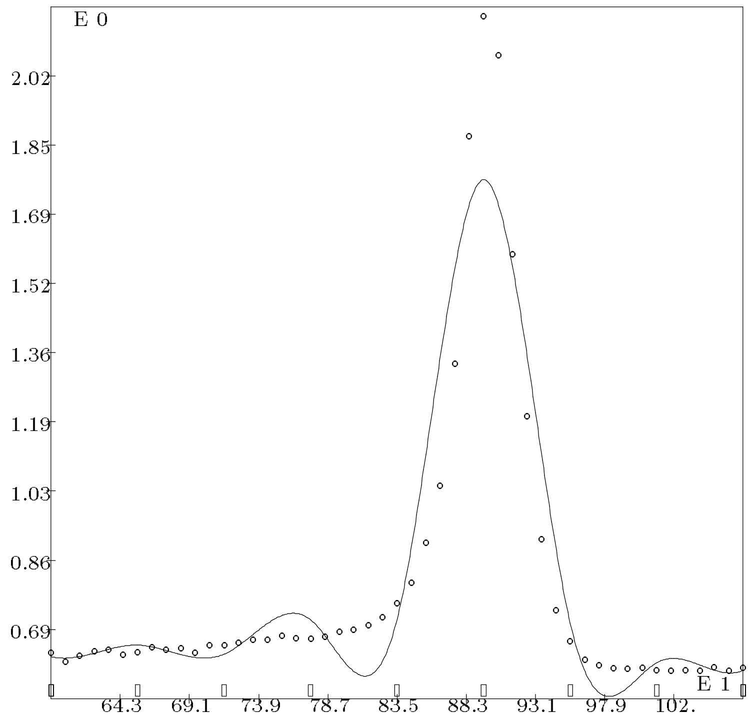


Abbildung 12: Titanium Heat Data, $r = 2$, $\mu = 1E3$, equidistant knots

7.4.1 Titan-Daten III

Wir wenden uns ein letztes Mal den Titan-Daten zu. Erneut wollen wir die $m = 49$ Meßpunkte durch $n = 11$ B-Splines der Ordnung $k = 4$ approximieren. Für den Glättungsparameter μ wählen wir $\mu = 1E3$. Wir beginnen den Minimierungsprozeß wieder mit äquidistanten inneren Knoten. Man vergleiche die folgenden Ergebnisse mit den Ergebnissen aus Abschnitt 6!

	Startpunkt	Golub/Pereyra-Modell
t_5	655	627.239439
t_6	715	657.590926
t_7	775	801.398888
t_8	835	869.748225
t_9	895	897.358543
t_{10}	955	928.206791
t_{11}	1015	994.289363
$\|\mathbf{r}\| = \|\mathbf{y} - \mathbf{B}\boldsymbol{\alpha}\|$	$8.151829E-01$	$4.318360E-01$
$\left(\|\mathbf{y} - \mathbf{B}\boldsymbol{\alpha}\|^2 + \mu\,\|\mathbf{S}\boldsymbol{\alpha}\|^2\right)^{1/2}$	$9.219330E-01$	$7.045600E-01$
Schritte		63
Zeit		2.15 sec
$\|\mathbf{G}^T\mathbf{J}\mathbf{s}\|$		$5.85493E-11$
$\|\mathbf{J}^T\mathbf{G}\|$		$3.45285E-08$
Return Code		3

Zum Vergleich ist an der ermittelten Optimalstelle der Spline ohne Glättungsterm ($\mu = 0.0$) berechnet worden.

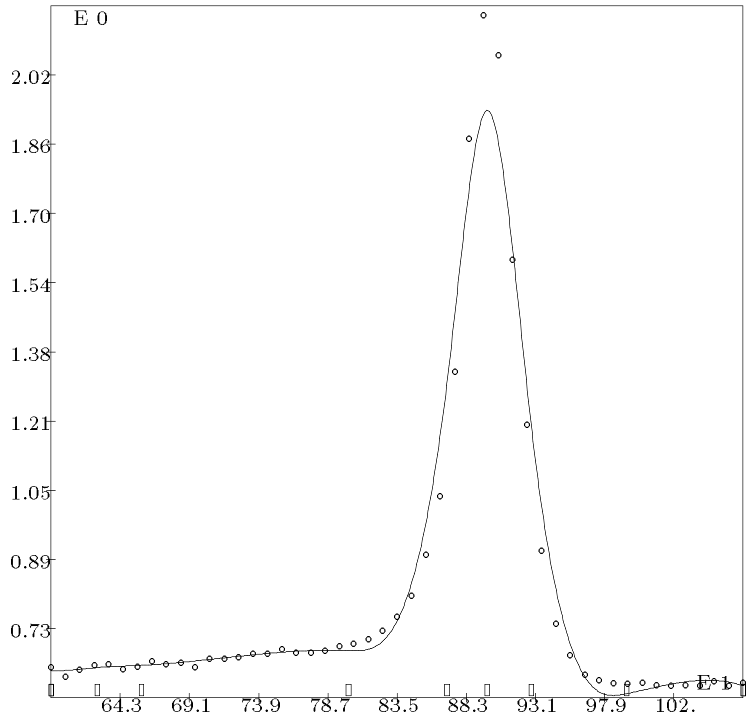


Abbildung 13: Titanium Heat Data, $l = 7$, $r = 2$, $\mu = 1E3$, Golub/Pereyra-Model

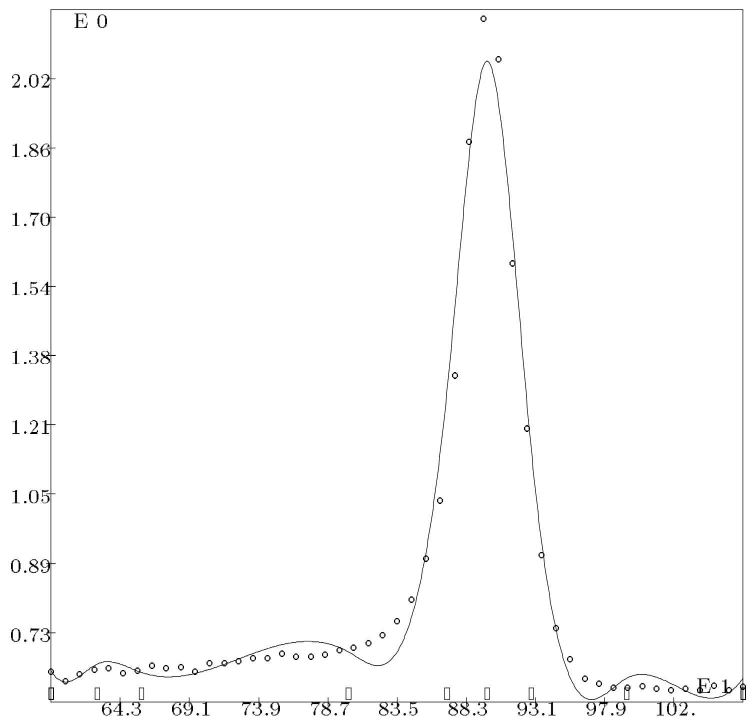


Abbildung 14: Titanium Heat Data, $\mu = 0.0$

7.4.2 RPN-14-Daten

Zum Abschluß wollen wir die RPN 14 Daten von Fritsch/Carlson [FC80] approximieren. Diese Daten stammen von einem radioaktiven Prozeß und zeigen eine für physikalisch-chemische Prozesse typische Sättigungserscheinung.

i	x_i	y_i
1	7.99	$0.00000E+00$
2	8.09	$2.76429E-05$
3	8.19	$4.37498E-02$
4	8.70	$1.69183E-01$
5	9.20	$4.69428E-01$
6	10.00	$9.43740E-01$
7	12.00	$9.98636E-01$
8	15.00	$9.99919E-01$
9	20.00	$9.99994E-01$

Die $m = 9$ Meßpunkte wollen wir durch quadratische B-Splines ($k = 3$) approximieren. Für die $l = 2$ inneren Splineknoten wählen wir $t_4 = 9$ und $t_5 = 12$. Es sei $r = 1$ und $\mu = 1E-02$.

	Startpunkt	Golub/Pereyra-Modell
t_4	9.000000	9.464841
t_5	12.000000	10.341722
$\|\|\mathbf{r}\|\| = \|\|\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\|\|$	$1.725484E-01$	$2.752452E-02$
$\left(\|\|\mathbf{y}-\mathbf{B}\boldsymbol{\alpha}\|\|^2 + \mu\,\|\|\mathbf{S}\boldsymbol{\alpha}\|\|^2\right)^{1/2}$	$1.917395E-01$	$9.099074E-02$
Schritte		8
Zeit		0.02 sec
$\|\|\mathbf{G}^T\mathbf{J}\mathbf{s}\|\|$		$9.42421E-12$
$\|\|\mathbf{J}^T\mathbf{G}\|\|$		$3.70994E-07$
Return Code		3

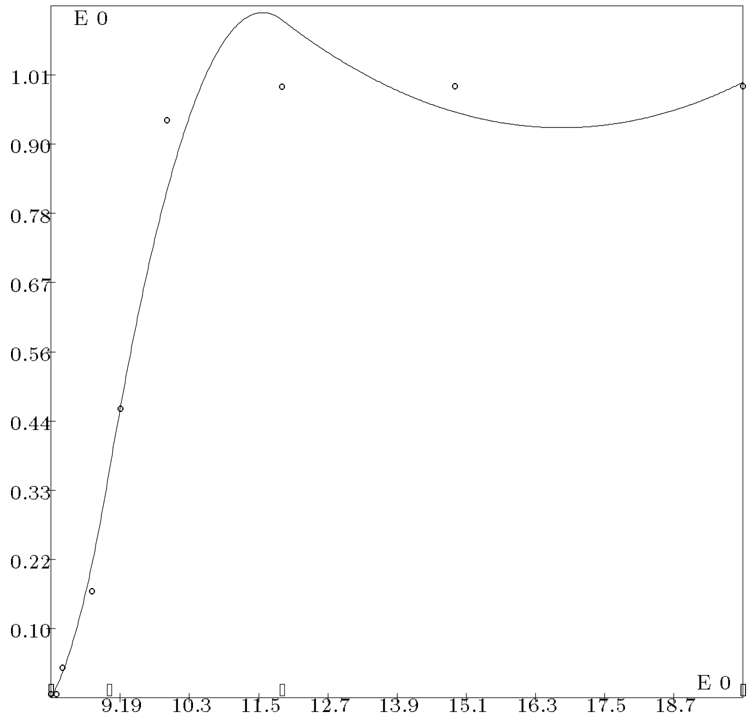


Abbildung 15: RPN14Data, $k = 3$, $n = 5$, $r = 1$, $\mu = 1E - 2$, start point

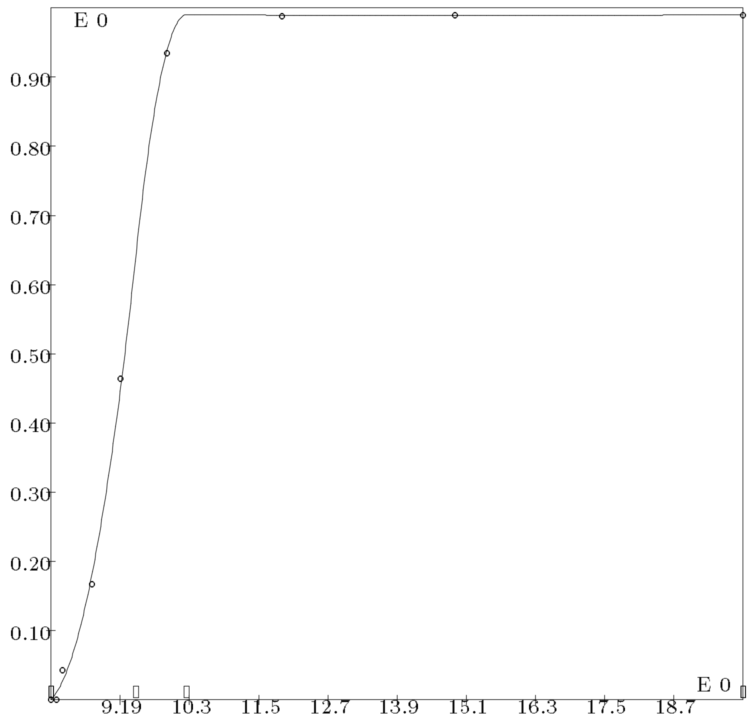


Abbildung 16: RPN14Data, $l = 2$, $r = 1$, $\mu = 1E - 2$, Golub/Pereyra-Model

Literatur

[ADLM90] E. Arge, M. Dæhlen, T. Lyche, and K. Mørken. Constrained spline approximation of functions and data based on constrained knot removal. In J. C. Mason and M. G. Cox, editors, *Algorithms for Approximation II*, pages 4–20. Chapman and Hall, 1990.

[Bjö90] Å. Björck. Least squares methods. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis, volume I, Solution of equations in R^n*, pages 589–652. North Holland, 1990.

[Boc87] H. G. Bock. *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen (Habilitationschrift).* Bonner Mathematische Schriften Nr.183, Universität Bonn, 1987.

[CMSW79] A. K. Cline, C. B. Moler, G. W. Stewart, and J. H. Wilkenson. An estimate for the condition of a matrix. *SIAM J. Numer. Anal.*, 16:368–375, 1979.

[Cox81] M. G. Cox. The least squares solution of overdetermined linear equations having band or augmented band structure. *IMA Journal Numerical Analysis*, 1:3–22, 1981.

[Cox82] M. G. Cox. Practical spline approximation. In P. R. Turner, editor, *Topics in Numerical Analysis*, pages 79–112. Springer Verlag, 1982.

[Cox89] M. G. Cox. Linear algebra support modules for approximation and other software. Technical Report DITC 142/89, National Physical Laboratory, Teddington, 1989.

[dB76] C. de Boor. Splines as linear combinations of B-splines. In G. G. Lorentz, C. K. Chui, and L. L. Schumaker, editors, *Approximation Theory II*, pages 1–47. Academic Press, New York, 1976.

[dB78] C. de Boor. *A practical guide to splines.* Springer-Verlag, New York, Heidelberg, Berlin, 1978.

[dBH85] C. de Boor and K. Hoellig. B-Splines without divided differences. TSR 2896, University of Wisconsin at Madison, 1985.

[DBMS79] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart. *LINPACK. Users guide.* SIAM Publications, 1979.

[dBP77] C. de Boor and A. Pinkus. Backward error analysis for totally positive linear systems. *Numer. Math.*, 27:485–490, 1977.

[DH91] P. Deuflhard and A. Hohmann. *Numerische Mathematik, Eine algorithmisch orientierte Einführung.* Walter de Gruyter, Berlin. New York, 1991.

[DS83] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, Inc., 1983.

[Eld84] L. Elden. An algorithm for the regularization of ill-conditioned banded least squares problems. *SIAM J. Sci. Stat. Comput.*, 5:237–254, 1984.

[FC80] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.*, 17(2):238–246, April 1980.

[Fle87] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, 1987.

[Gay84] D. M. Gay. A trust-region approach to linearly constrained optimization. In D. F. Griffiths, editor, *Proceedings 1983 Dundee Conference on Numerical Analysis*, pages 72–105, Berlin, 1984. Lecture Notes in Mathematics 1066, Springer.

[GMW81] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization.* Academic Press Inc., 1981.

[GP73] G. H. Golub and V. Pereyra. The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10:413–432, 1973.

[GP76] G. H. Golub and V. Pereyra. Differentiation of pseudoinverses, separable nonlinear least squares problems and other tales. In M. Nashed, editor, *Generalized Inverses and Applications*, pages 303–324. Academic Press, New York, 1976.

[GvL89] G. H. Golub and C. F. van Loan. *Matrix Computations.* The John Hopkins University Press, Oxford, Baltimore, 1989.

[HF79] J. N. Holt and R. Fletcher. An algorithm for constrained nonlinear least squares. *J. Inst. Math. Appl.*, 23:449–463, 1979.

[Kau75] L. Kaufman. A variable projection method for solving separable nonlinear least squares problems. *BIT*, 15(4):49–57, 1975.

[KP78] L. Kaufman and V. Pereyra. A method for separable nonlinear least squares problems with separably nonlinear equality constraints. *SIAM J. Numer. Anal.*, 15:12–20, 1978.

[Kro74] F. T. Krogh. Efficient implementation of a variable projection algorithm for nonlinear least squares problems. *Comm. ACM*, 17(3):167–169, March 1974.

[KS88] A. Kiełbasinski and H. Schwetlick. *Numerische Lineare Algebra.* Deutscher Verlag der Wissenschaften, Berlin, 1988.

[Kun89] V. Kunert. Splineglättung mit Ungleichheitsbedingungen an Ableitungen. Master's thesis, MLU Halle-Wittenberg, 1989.

[LH74] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems.* Prentice Hall, Inc., 1974.

[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.*, 8:185–208, 1988.

[Par85] T. A. Parks. *Reducible Nonlinear Programming Problems.* PhD thesis, Houston Univ., Dept. of Mathematics, Houston, 1985.

[RW80] A. Ruhe and P. Å. Wedin. Algorithms for separable nonlinear least squares problems. *SIAM Review*, 22(3):318–337, July 1980.

[Sch79] H. Schwetlick. *Numerische Lösung nichtlinearer Gleichungen.* Deutscher Verlag der Wissenschaften, 1979.

[Sch81] L. L. Schumaker. *Spline Functions: Basic Theory.* John Wiley and Sons, New York, 1981.

[SK93] H. Schwetlick and V. Kunert. Spline smoothing under constraints on derivatives. *BIT*, 1993. to appear.

[Spe93] P. Spellucci. *Numerische Verfahren der nichtlinearen Optimierung.* Birkhäuser, 1993.

[Wah90] G. Wahba. *Spline Models for Observational Data.* SIAM, Philadelphia, Pennsylvania, 1990.

[WH85] S. J. Wright and J. N. Holt. Algorithms for nonlinear least squares with linear inequality constraints. *SIAM J. Sci. Statist. Comp.*, 6:1033–1048, 1985.

[WL69] H. Wold and E. Lyttkens. Nonlinear iterative partial least squares (NIPALS) estimation procedures. *Bull. ISI*, 43:29–51, 1969.

[Wri92] S. J. Wright. An interior point algorithm for linearly constrained optimization. *SIAM J. Optimization*, 1992. to appear.

Erklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur angefertigt habe.

Halle/Saale, den ....