Ausgleichende natürliche kubische Splines und die Schätzung des Glättungsparameters in SAS

M. Wodny, B. Jäger, K.-E. Biebler
Institut für Biometrie und Medizinische Informatik
W.-Rathenau-Str. 48
17487 Greifswald
wodny@biometrie.uni-greifswald.de

Zusammenfassung

Ausgleichende Splinefunktionen sind durch die Prozedur FIT(YX) im SAS verfügbar. Allerdings ist die angezeigte Ausgleichsfunktion nicht ohne weiteres für weitergehende Berechnungen zugänglich.

Vorgestellt wird eine IML Prozedur, die die in einem gewissen Sinne optimale ausgleichende natürliche kubische Splinefunktion s(x) zu gegebenen Wertepaaren (x_i,y_i), i=1,2,...,m, konstruiert. Als Ergebnis liegen die Splinekoeffizienten vor, so dass bei Bedarf auch andere Kurvenparameter, wie Ableitungen oder bestimmte Integrale, berechenbar sind.

Einführend werden die behandelten Funktionen definiert und eine Interpolationsaufgabe betrachtet. Diese ist eindeutig lösbar, wenn zwei Zusatzbedingungen gestellt werden. Natürliche kubische Splines sind durch spezielle Randbedingungen gekennzeichnet und erfüllen eine integrale Minimaleigenschaft. Dadurch ist es möglich, ausgleichende kubische Splines zu konstruieren. Zu diesem Zweck werden verschiedene Optimierungsaufgaben im Raum der zweimal stetig differenzierbaren Funktionen formuliert. Es ergibt sich, dass alle eindeutig lösbar sind und die entsprechende Lösung eine natürliche kubische Splinefunktion ist.

Nachdem die einzelnen Probleme in äquivalente quadratische Optimierungsaufgaben in einem endlichdimensionalen Vektorraum überführt wurden, erhält man die numerische Lösung in zwei Schritten. Zunächst werden die noch unbekannten Funktionswerte $s(x_i)$ über ein lineares Gleichungssystem ermittelt. In einem zweiten Schritt ist es dann möglich, den entsprechenden interpolierenden natürlichen kubischen Spline zu konstruieren.

Jeder Lösungsspline ist durch einen Parameter gekennzeichnet, der die Glättung bestimmt. Die Schätzung dieses Glättungsparameters erfolgt nach einer one-live-out (cross validation) Methode durch Minimierung eines eingeführten Fehlerfunktionals.

Keywords: Kurvenanpassung, kubische Splines, quadratische Optimierung.

1 Interpolierende kubische Splinefunktionen

Gegeben seien die m Wertepaare (x_i,y_i). Es soll davon ausgegangen werden, dass die x-Werte der Größe nach geordnet und paarweise verschieden sind.

Definition (kubische Splinefunktion)

Eine Funktion s(x): ú6ú heißt kubische Splinefunktion zu den Stützstellen bzw. Knoten $x_1<x_2<...<x_m$ genau dann, wenn

a) s(x) auf jedem Intervall $[x_i,x_{i+1}]$, i=1,2,...,m-1, auf $(-4,x_1]$ und $[x_m,+4)$ durch je ein Polynom vom Grad kleiner oder höchstens gleich 3 dargestellt wird und

b s(x) auf ganz ú zweimal stetig differenzierbar ist.

s(x) heißt eine **natürliche** kubische Splinefunktion, wenn s(x) auf den Intervallen $(-4,x_1]$ sowie $[x_m,+4)$ jeweils durch ein Polynom vom Grad kleiner oder gleich eins dargestellt wird.

Weil s(x) zwischen den einzelnen Stützstellen durch ein Polynom vom Grad kleiner oder gleich drei beschrieben werden kann, gilt für alle x aus $[x_i,x_{i+1}]$ der Ansatz $s(x)=A_i(x-x_i)^3 + B_i(x-x_i)^2 + C_i(x-x_i) + D_i$

(bzw. $s(x)=A_0(x_1-x)^3+B_0(x_1-x)^2+C_0(x_1-x)+D_0$ auf $(-4,x_1]$ und

$s(x)=A_m(x-x_m)^3+ B_m(x-x_m)^2+C_m(x-x_m)+D_m$ über $[x_m,+4)$).

Da s(x) die Interpolationsaufgabe lösen soll, folgt sofort

$$s(x_i)=y_i=D_i,\ i=1,2,...,m-1. \qquad (1.1)$$

Außerdem ist $s''(x_i)=2B_i$, also

$$B_i=s''(x_i)/2. \qquad (1.2)$$

Die zweite Ableitung s"(x) ist auf $[x_i,x_{i+1}]$ eine Gerade mit dem Anstieg $(s''(x_{i+1})-s''(x_i))/(x_{i+1}-x_i)$, so dass sich für die A_i, i=1,2,...,m-1, die Beziehungen

$$A_i=1/6(s''(x_{i+1})-s''(x_i))/(x_{i+1}-x_i) \qquad (1.3)$$

ergeben.

Mit (1.1), (1.2) und (1.3) erhält man ebenfalls eine Darstellung der C_i aus den Wertepaaren und den zweiten Ableitungen $s''(x_{i+1})$ und $s''(x_i)$:

$$C_i=(y_{i+1}-y_i)/(x_{i+1}-x_i) - 1/6(x_{i+1}-x_i)(2s''(x_i) + s''(x_{i+1})) . \qquad (1.4)$$

Aus diesen Überlegungen folgt, dass eine natürliche kubische Splinefunktion, die die Interpolationsbedingungen $s(x_i)=y_i$, i=1,2,...,m, erfüllt, eindeutig durch die Wertepaare und die zweiten Ableitungen $s''(x_i)$ in allen Stützstellen bestimmt ist.

Um diese Unbekannten zu ermitteln, muss man nur beachten, dass s'(x) stetig in den x_i, i=1,2,...,m, ist und deshalb $s'(x_i)=3A_{i-1}(x_i-x_{i-1})^2 + 2B_{i-1}(x_i-x_{i-1}) + C_{i-1} = C_i$, i=2,3,...,m-1, gilt. Nach einigen Umformungen erhält man mit den Abkürzungen **)** $x_i:=x_{i+1}-x_i$ und **)** $y_i:=y_{i+1}-y_i$, i=1,2,...,m-1,

$$\Delta x_{i-1}s''(x_{i-1}) + 2(\Delta x_{i-1} + \Delta x_i)s''(x_i) + \Delta x_i s''(x_{i+1}) = 6\left(\frac{\Delta y_1}{\Delta x_1} - \frac{\Delta y_{i-1}}{\Delta x_{i-1}}\right). \qquad (1.5)$$

Das sind zunächst nur m-2 Gleichungen (i=2,3,...,m-1) zur Bestimmung der m Unbekannten $s''(x_i)$. Im Allgemeinen werden die Randbedingungen $s''(x_1)=a$ und $s''(x_m)=b$ angegeben. Die beiden fehlenden unabhängigen Gleichungen gewinnt man aber auch durch Vorgabe der ersten Ableitungen $s'(x_1)=$**"** und $s'(x_m)=$**$** oder anderer Bedingungen.

Natürliche kubische Splines s(x) sind laut Definition über den offenen Intervallen $(-4,x_1]$ und $[x_m,+4)$ eine Gerade und die zweite Ableitung ist überall stetig (insbesondere in x_1 und x_m). Daraus folgen $s''(x_1)=0$ und $s''(x_m)=0$.

Das lineare Gleichungssystem (1.5) lautet dann in Matrixform unter Verwendung der Splinekoeffizienten B_i

$$
\begin{pmatrix}
2(\Delta x_1+\Delta x_2) & \Delta x_2 & 0 & \ldots & 0 \\
\Delta x_2 & 2(\Delta x_2+\Delta x_3) & \Delta x_3 & \ldots & 0 \\
0 & \Delta x_3 & 2(\Delta x_3+\Delta x_4) & \ldots & 0 \\
. & . & . & . & . \\
. & . & . & . & . \\
0 & 0 & 0 & \Delta x_{m-2} & 2(\Delta x_{m-2}+\Delta x_{m-1})
\end{pmatrix}
\circ
\begin{pmatrix}
B_2 \\ B_3 \\ B_4 \\ . \\ . \\ B_{m-1}
\end{pmatrix}
=
$$

$$
\begin{pmatrix}
3\left[\dfrac{\Delta y_2}{\Delta x_2}-\dfrac{\Delta y_1}{\Delta x_1}\right] \\
3\left[\dfrac{\Delta y_3}{\Delta x_3}-\dfrac{\Delta y_2}{\Delta x_2}\right] \\
3\left[\dfrac{\Delta y_4}{\Delta x_4}-\dfrac{\Delta y_3}{\Delta x_3}\right] \\
. \\
. \\
3\left[\dfrac{\Delta y_{m-1}}{\Delta x_{m-1}}-\dfrac{\Delta y_{m-2}}{\Delta x_{m-2}}\right]
\end{pmatrix}
; \quad B_1 = B_m = 0. \tag{1.6}
$$

Definition (Gesamtkrümmung)

Es sei f(x) eine auf [a,b] zweimal stetig differenzierbare Funktion. Dann wird $\int_a^b$ f"(x)^{2}dx als die Gesamtkrümmung von f(x) über diesem Intervall bezeichnet.

Satz (HOLLADAY)

Der interpolierende natürliche kubische Spline s(x), durch die Wertepaare (xi,yi), i=1,2,...,m, ist eindeutig bestimmt. Er besitzt minimale Gesamtkrümmung über jedem Intervall [a,b], ạx1 und xṃb, unter allen zweimal stetig differenzierbaren Funktionen f(x), die ebenfalls die Interpolationsbedingungen f(xi)=yi, i=1,2,...,m, erfüllen.

Es gilt $\int_a^b$ f"(x)^{2}dx > $\int_a^b$ s"(x)^{2}dx, wenn f(x)†s(x) ist.

2 Ausgleichende kubische Splinefunktionen

Nach diesen Vorbetrachtungen werden die Optimierungsaufgaben formuliert, durch deren Lösung eine Glättung der gegebenen Wertepaare erreicht werden soll.

Gegeben seien die Wertepaare (x_i,y_i), i=1,2,...,m, und ein **:**$0. Gesucht wird eine zweimal stetig differenzierbare Funktion f(x), für die die Zielfunktion $Z(f):=\mu \int^{x_m} f''(x)^2dx + \frac{1}{m}\sum^{m} (f(x_i) - y_i)^2$ minimal wird.	(OP1)

Die Funktion f(x) soll so konstruiert werden, dass die gewichtete Summe aus der Gesamtkrümmung und der Fehlerquadratsumme ein Minimum wird. Dabei ist **:,**ú$^+$ ein zwar beliebiger, aber fest vorgegebener Parameter, der bestimmt, mit welchem Gewicht die Gesamtkrümmung in die Zielfunktion Z(f) eingehen soll.

Gesucht wird eine Funktion f(x)**,**$C^2[x_1,x_m]$, für die $\int_{x_1}^{x_m} f''(x)^2dx$ minimal wird. Zur Minimumbildung werden alle zweimal stetig differenzierbaren Funktionen f(x) zugelassen, für die $\sum_{i=1}^{m} (f(x_i) - y_i)^2 \leq S$ gilt.	(OP2)

Dabei ist $S \geq 0$ eine beliebige aber fest vorgegebene reelle Zahl (vergl. REINSCH).

Gegeben seien die Wertepaare (x_i,y_i), i= 1, 2,...,m, und $T \geq 0$. Gesucht wird eine zweimal stetig differenzierbare Funktion f(x) mit der Nebenbedingung $\int^{x_m} f''(x)^2dx \leq T$, so dass $\sum_{i=1}^{m} (f(x_i) - y_i)^2$ ein Minimum wird.	(OP3)

Gesucht wird somit eine Funktion aus $C^2([x_1,x_m])$, die mit minimaler Fehlerquadratsumme durch die gegebenen Punkte (x_i,y_i), i=1,2,...,m, verläuft. Dabei soll ihre Gesamtkrümmung eine vorgegebene feste Konstante T nicht überschreiten.
Am Beispiel von (OP1) wird der Lösungsweg beschrieben. Das Optimierungsproblem wird zunächst nur in der Menge aller natürlichen kubischen Splines bezüglich des gegebenen Stützstellensystems $x_1<x_2<...<x_m$ betrachtet. Da s(x) eindeutig durch die Interpolationspunkte $(x_i,s(x_i)=D_i)$ und die B_i gegeben ist, ist es auch möglich, einen Ausdruck für die Gesamtkrümmung zu ermitteln, der von diesen Größen abhängig ist. Es gilt

$$\int_{x_1}^{x_m} s''(x)^2 dx = \sum_{i=1}^{m-1} \int_{x_i}^{x_{i+1}} s''(x)^2 dx = \sum_{i=1}^{m-1} \frac{1}{18 A_i} [6A_i(x - x_i) + 2B_i]^3 \Big|_{x_i}^{x_{i+1}}$$

$$= \frac{4}{3} \sum_{i=1}^{m-1} \Delta x_i (B_i^2 + B_i B_{i+1} + B_{i+1}^2)$$

$$= \frac{2}{3} \sum_{i=1}^{m-1} \Delta x_i (B_i^2 + (B_i + B_{i+1})^2 + B_{i+1}^2)$$

(OP1) kann dann in ein äquivalentes Problem im endlich dimensionalen Raum ú^m überführt werden:

$$\mu \frac{2}{3}\sum_{i=1}^{m-1}\Delta x_i (B_i^2 + (B_i + B_{i+1})^2 + B_{i+1}^2) + \frac{1}{m}\sum_{i=1}^{m} (D_i - y_i)^2 = \min!$$

Dabei soll das Minimum durch die passende Wahl der B_i, i=2,3,...,m-1, (B_1=B_m=0) und der noch unbekannten Funktionswerte D_i=s(x_i), i=1,2,...,m, realisiert werden.
Da s(x) eindeutig durch das Gleichungssystem (1.6) bestimmt ist, können diese Gleichungen als Nebenbedingungen in die Zielfunktion eingebaut werden. Durch Einführung der Lagrangeschen Multiplikatoren V_2, V_3, ... , V_{m-1} ergibt sich dann

$$L(B_2,\dots,B_{m-1},D_1,\dots,D_m,V_{m-1}) := \frac{2}{3}\mu\sum_{i=1}^{m-1}\Delta x_i (B_i^2 + (B_i + B_{i+1})^2 + B_{i+1}^2) +$$

$$\frac{1}{m}\sum_{i=1}^{m}(D_i - y_i)^2 +$$

$$\sum_{i=2}^{m-1} V_i\left[\Delta x_{i-1}B_{i-1} + 2(\Delta x_{i-1} + \Delta x_i)B_i + \Delta x_i B_{i+1} - 3\left(\frac{(D_{i+1} - D_i)}{\Delta x_i} - \frac{(D_i - D_{i-1})}{\Delta x_{i-1}}\right)\right] = \min! \qquad (2.1)$$

Notwendige Bedingungen für ein lokales Minimum von L sind die Beziehungen

$$\frac{\partial L}{\partial B_i} = \frac{2}{3}\mu\Delta x_{i-1}(4B_i + 2B_{i-1}) + \frac{2}{3}\mu\Delta x_i(4B_i + 2B_{i+1}) +$$

$$V_{i-1}\Delta x_{i-1} + 2V_i(\Delta x_{i-1} + \Delta x_i) + V_{i+1}\Delta x_i = 0$$

bzw.

$$\frac{\partial L}{\partial B_i} = \frac{4}{3}\mu[\Delta x_{i-1}B_{i-1} + 2(\Delta x_{i-1} + \Delta x_i)B_i + \Delta x_i B_{i+1}] +$$

$$V_{i-1}\Delta x_{i-1} + 2V_i(\Delta x_{i-1} + \Delta x_i) + V_{i+1}\Delta x_i = 0 \quad \text{für i= 2, 3,..., m-1,} \qquad (2.2)$$

$$\frac{\partial L}{\partial V_i} = \frac{2}{m}(D_i - y_i) - 3\frac{V_{i-1}}{\Delta x_{i-1}} + 3V_i\left[\frac{1}{\Delta x_{i-1}} + \frac{1}{\Delta x_i}\right] - 3\frac{V_{i+1}}{\Delta x_i} = 0,\ i = 1,2,\dots,m, \qquad (2.3)$$

und für i=2,3,..., m-1

$$\frac{\partial L}{\partial V_i} = \Delta x_{i-1}B_{i-1} + 2(\Delta x_{i-1} + \Delta x_i)B_i + \Delta x_i Bx_{i+1} - 3\left[\frac{(D_{i+1} - D_i)}{\Delta x_i} - \frac{(D_i - D_{i-1})}{\Delta x_{i-1}}\right] = 0. \qquad (2.4)$$

(Man beachte, dass B_1=B_m=0 gilt und dass zum Zwecke einer bequemeren Schreibweise in (2.3) Konstanten V_0=V_1=V_m=V_{m+1}=0 und) x_0=) x_m=1 eingeführt wurden.)
Die Beziehungen (2.2) und (2.3) sind zusammen mit (2.4) ein lineares Gleichungssystem zur Bestimmung der B_i, V_i und D_i. Es könnte versucht werden, dieses direkt zu lösen. Das ist aber nicht zu empfehlen, weil es in zwei separate Gleichungssysteme für die D_i und B_i gegliedert werden

kann. Zur Vereinfachung der Schreibweise seien die Vektoren $B:=(B_2,B_3,...,B_{m-1})^T$, $V:=(V_2,V_3,...,V_{m-1})^T$, $D:=(D_1,D_2,...,D_m)^T$, $Y:=(y_1,y_2,...,y_m)^T$ sowie die Matrizen

$$Z := \frac{1}{3}\begin{pmatrix} 2(\Delta x_1+\Delta x_2) & \Delta x_2 & 0 & ... & 0 \\ \Delta x_2 & 2(\Delta x_2+\Delta x_3) & \Delta x_3 & ... & 0 \\ 0 & \Delta x_3 & 2(\Delta x_3+\Delta x_4) & ... & 0 \\ . & . & . & ... & . \\ . & . & . & ... & . \\ . & . & . & ... & . \\ 0 & ... & \Delta x_{m-3} & 2(\Delta x_{m-3}+\Delta x_{m-2}) & \Delta x_{m-2} \\ 0 & ... & 0 & \Delta x_{m-2} & 2(\Delta x_{m-2}+\Delta x_{m-1}) \end{pmatrix}$$

und

$$Q := \begin{pmatrix} \frac{1}{\Delta x_1} & -\left(\frac{1}{\Delta x_1}+\frac{1}{\Delta x_2}\right) & \frac{1}{\Delta x_2} & 0 & ... & 0 \\ 0 & \frac{1}{\Delta x_2} & -\left(\frac{1}{\Delta x_2}+\frac{1}{\Delta x_3}\right) & \frac{1}{\Delta x_3} & ... & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & ... & \frac{1}{\Delta x_{m-2}} & -\left(\frac{1}{\Delta x_{m-2}}+\frac{1}{\Delta x_{m-1}}\right) & \frac{1}{\Delta x_{m-1}} \end{pmatrix}$$

eingeführt.

Das Gleichungssystem (2.4) hat dann die Form $Z \circ B=Q \circ D$, woraus sich für B ergibt:

$$B=Z^{-1}BQBD. \qquad (2.5)$$

(2.2) geht über in $(4/3)\mathbf{:}3ZBB = -3ZBV$. Damit gilt

$$V = (-4/3)\mathbf{:}B. \qquad (2.6)$$

Aus (2.3) folgt die Beziehung

$$\frac{2}{m}\,[D-Y] = 3Q^T \circ V. \qquad (2.7)$$

Es wird (2.6) und anschließend (2.5) in (2.7) eingesetzt, und man erhält

$\frac{2}{m}\,[D-Y] = -3Q^T(4/3)\mathbf{:}Z^{-1} \circ Q \circ D$. Daraus folgt

$$[I + 2\mathbf{:}mQ^T \circ Z^{-1} \circ Q] \circ D = Y. \qquad (2.8)$$

Nach kurzer Rechnung überzeugt man sich von der Gleichung

$$\int_{x_1}^{x_m} s''(x)^2 dx = 2B^T \circ Z \circ B = 2D^T \circ Q^T \circ Z^{-1} \circ Q \circ D \geq 0.$$

Die Matrix $E:=Q^T \circ Z^{-1} \circ Q$ ist somit positiv semidefinit. Daraus folgt, dass die Koeffizientenmatrix von (2.8) für alle $\mathbf{:} \geq 0$ positiv definit und damit invertierbar ist. Aus den D_i lassen sich mit Hilfe

von (1.6) die B_i, i=2,3,...,m-1, berechnen, so dass die gesuchte Splinefunktion s(x) vollständig konstruiert werden kann.
(Aus (2.7) und (2.6) folgt $D=Y-2m\mu Q^T \circ B$. Das wird in $Z \circ B=Q \circ D$ eingesetzt. Daraus folgt $[Z+2m\mu Q \circ Q^T] \circ B=Q \circ Y$. Die Matrix $[Z+2m\mu Q \circ Q^T]$ ist ebenfalls positiv definit, so dass $B=[Z+2m\mu Q \circ Q^T]^{-1} \circ Q \circ Y$ gilt und
$D=Y - 2m\mu Q^T \circ [Z+2m\mu Q \circ Q^T]^{-1} \circ Q \circ Y$ ist. Das bedeutet
$I - 2m\mu Q^T \circ [Z+2m\mu Q \circ Q^T]^{-1} \circ Q = [I + 2 : mQ^T \circ Z^{-1} \circ Q]^{-1}$.)
Dass die so ermittelte natürliche kubische Splinefunktion s(x) auch (OP1) löst, folgt aus dem Satz von HOLLADAY. Wird nämlich angenommen, dass die Zielfunktion von (OP1) durch ein f(x) minimiert wird, welches kein Spline ist, konstruiere man zu den Wertepaaren $(x_i,f(x_i))$, i=1,2,...,m, den eindeutigen interpolierenden natürlichen kubischen Spline. Dieser besitzt die gleiche Fehlerquadratsumme bezüglich der y_i hat aber eine kleinere Gesamtkrümmung, so dass f(x) nicht (OP1) lösen kann.
In analoger Weise werden die anderen Aufgabenstellungen in äquivalente Optimierungsaufgaben im ú^m überführt. (OP2) geht über in

$$L^*(B_2,...,B_{m-1},D_1,...,D_m,V_2,...,V_{m-1},p_2,h) := \frac{2}{3}\sum_{i=1}^{m-1}\Delta x_i[B_i^2+(B_i+B_{i+1})^2+B_{i+1}^2]+$$

$$p_2\left[\sum_{i=1}^{m}(D_i-y_i)^2+h^2-S\right]+$$

$$\sum_{i=2}^{m-1}V_i\left[\Delta x_{i-1}B_{i-1}+2(\Delta x_{i-1}+\Delta x_i)B_i+\Delta x_i B_{i+1}-3\left(\frac{(D_{i+1}-D_i)}{\Delta x_i}-\frac{(D_i-D_{i-1})}{\Delta x_{i-1}}\right)\right]=\min!$$

Das Problem (OP3) liefert

$$L^{**}(B_2,...,B_{m-1},D_1,...,D_m,V_2,...,V_{m-1},p_3,h) := \sum_{i=1}^{m}(D_i-y_i)^2+$$

$$p_3\left[\frac{2}{3}\sum_{i=1}^{m-1}\Delta x_i[B_i^2+(B_i+B_{i+1})^2+B_{i+1}^2]+h^2-T\right]+$$

$$\sum_{i=2}^{m-1}V_i\left[\Delta x_{i-1}B_{i-1}+2(\Delta x_{i-1}+\Delta x_i)B_i+\Delta x_i B_{i+1}-3\left(\frac{(D_{i+1}-D_i)}{\Delta x_i}-\frac{(D_i-D_{i-1})}{\Delta x_{i-1}}\right)\right]=\min!$$

Dabei sind die Nebenbedingungen jeweils über einen Lagrangeschen Multiplikator (p_2 bzw. p_3) und die Schlupfvariable h in das Zielfunktional eingebaut worden.
Die entsprechenden notwendigen Bedingungen $ML^*/MB_i=0$, $ML^*/MV_i=0$, i=2,3,...,m-1, und $ML^*/MD_i=0$, i=1,2,...,m, führen zu den Beziehungen
$B=Z^{-1} \circ Q \circ D$, $V=-4/3B$ und $2p_2[D-Y]=3Q^T \circ V$. Daraus ergibt sich das für $p_2>0$ eindeutig lösbare lineare Gleichungssystem

$$[I+2/p_2 Q^T \circ Z^{-1} \circ Q] \circ D=Y. \tag{2.9}$$

Die ebenfalls für ein lokales Minimum notwendigen Bedingungen $ML^{**}/MB_i=0$, $ML^{**}/MV_i=0$, i=2,3,...,m-1, und $ML^{**}/MD_i=0$, i=1,2,...,m, ergeben die Gleichungen
$V=-(4/3)p_3B$, $Z \circ B=Q \circ D$ und $2[D-Y]=3Q^T \circ V$. Daraus folgt
$K^{-2} \circ [D - Y] = -2pQ^T \circ Z^{-1} \circ Q \circ D$, also

$$[I + 2p_3Q^T \circ Z^{-1} \circ Q] \circ D = Y \tag{2.10}$$

Aus den ebenfalls notwendigen Bedingungen $ML^*/Mh=2hp_2=0$ und $ML^{**}/Mh=2hp_3=0$ folgt, dass die jeweils eindeutigen Lösungen von (OP2) bzw. (OP3) die vorgegebene Schranke auch annehmen, wenn der Lagrangesche Multiplikator nicht verschwindet. In diesem Sinne können die Fehlerquadratsumme in (OP2) und die Gesamtkrümmung in (OP3) ebenfalls als Glättungsparameter angesehen werden. Die enge Verwandtschaft der zunächst sehr verschiedenen Optimierungsprobleme zeigt sich auch in den die Vektoren der D_i definierenden linearen Gleichungssystemen (2.8), (2.9) und (2.10).
Ist einer der Lagrangeschen Multiplikatoren oder **:** bekannt, können die entsprechenden Lösungssplines konstruiert werden.

3 Schätzung der Glättungsparameter

Wie bereits ausgeführt, entspricht der Parameter **:** in (OP1) einem Glättungsparameter, der den Ausgleich der gegebenen Wertepaare steuert. Für **:**=0 ergibt sich der interpolierende natürliche kubische Spline und für **:** gegen Unendlich erhält man im Grenzprozess die Ausgleichsgerade bezüglich der (x_i,y_i). Beim Problem (OP2) übernimmt diese Funktion die Schranke S. Für S=0 erhält man ebenfalls den interpolierenden Spline als Lösung, und wenn S zwischen Null und S_{max}, der Fehlerquadratsumme der Ausgleichsgeraden, gewählt wird, ist die eindeutige Lösung ein natürlicher kubischer Spline mit genau dieser Fehlerquadratsumme bezüglich der y_i, i=1,2,...,m. In (OP3) wird die Glättung durch die Gesamtkrümmung gesteuert. Liegt T zwischen Null und der Gesamtkrümmung des interpolierenden Splines, dann ist die eindeutige Lösung von (OP3) eine natürliche kubische Splinefunktion, die die Schranke T auch annimmt.
Ein großes Problem stellt die richtige Wahl der entsprechenden Glättungsparameter dar. Je nach dem, wie das Verhältnis von Fehlerquadratsumme und Gesamtkrümmung für einen gegebenen Spline ist, treten sichtbare Glättungseffekte bei recht unterschiedlichen Parametern auf. Die Vorgabe ist in jedem Fall von einer gewissen Willkür.
Die bekannte cross-validation-Methode zur Schätzung der Glättungsparameter wird am Beispiel für **:** in (OP1) erläutert. Bei fixiertem **:** wird jeweils der q-te Punkt (x_q,y_q), q=1,2,…m, aus den Daten entfernt und zu den verbleibenden m-1 die eindeutige Lösung $s_q^{\mu}(x)$ von (OP1) berechnet. Der ausgesparte y-Wert besitzt die quadratische Abweichung $(y_q - s_q^{\mu}(x_q))^2$ zum konstruierten Spline. Als Schätzer des Glättungsparameters fungiert dann das Minimum der Funktion

$$FE_{(OP1)}=FE_{(OP1)}(\mu):=\sum_{q=1}^{m}(y_q - s_q^{\mu}(x_q))^2 . \tag{3.1}$$

In analoger Weise können für die anderen Optimierungsaufgaben die Fehlerfunktionen

$$FE_{(OP2)}=FE_{(OP2)}(S):=\sum_{q=1}^{m}(y_q - s_q^{S}(x_q))^2 \tag{3.2}$$

und

$$FE_{(OP3)}=FE_{(OP3)}(T):=\sum_{q=1}^{m}(y_q - s_q^{T}(x_q))^2 \tag{3.3}$$

definiert werden. Man beachte, dass alle Fehlerfunktionen mehrere lokale Minima besitzen können.

Als Beispiel werden die 10 Wertepaare (9.3, 15.9), (10.2, 20.7), (11.4, 18.6), (11.8, 20.4), (13.1, 19.0), (13.4, 18.0), (14.0, 22.5), (14.2, 21.5), (15.3, 23.9) und (17.2, 26.2) betrachtet. Abbildung 1 zeigt einen Teil der Fehlerfunktion $FE_{(OP1)}(:)$ und Abbildung 2 die entsprechende Lösung von (OP1) für $:=:_{min}$=0.0277325 von (3.1).

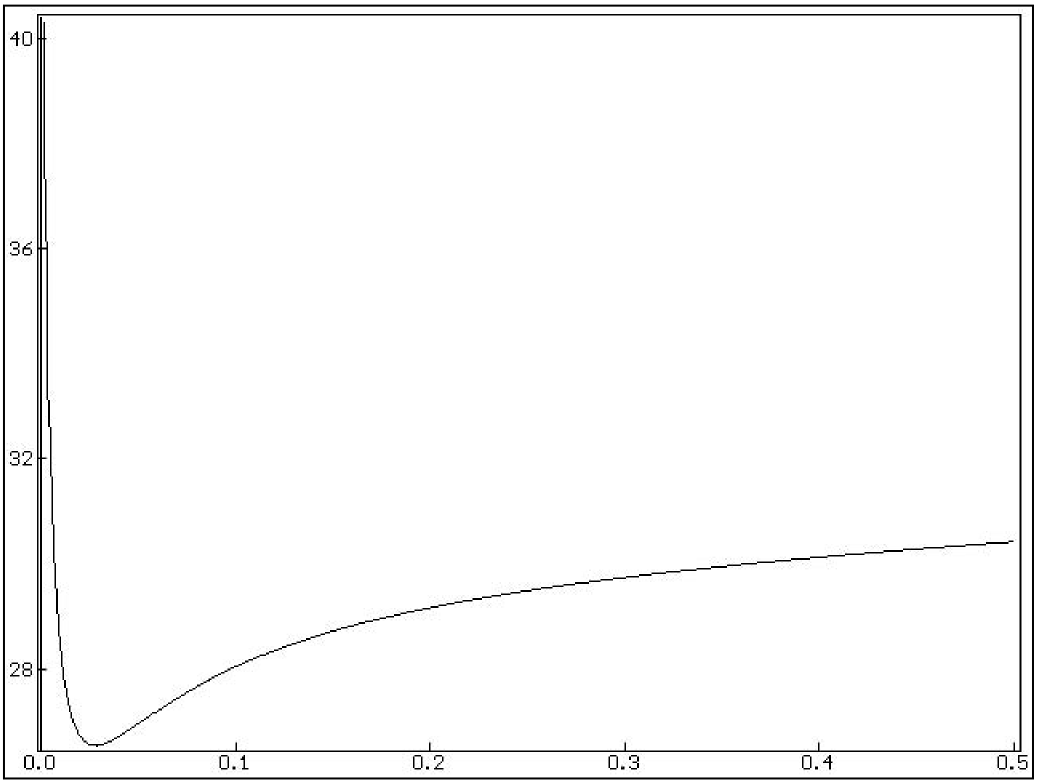

Abbildung 1: Ausschnitt der Fehlerfunktion $FE_{(OP1)}(\mu)$

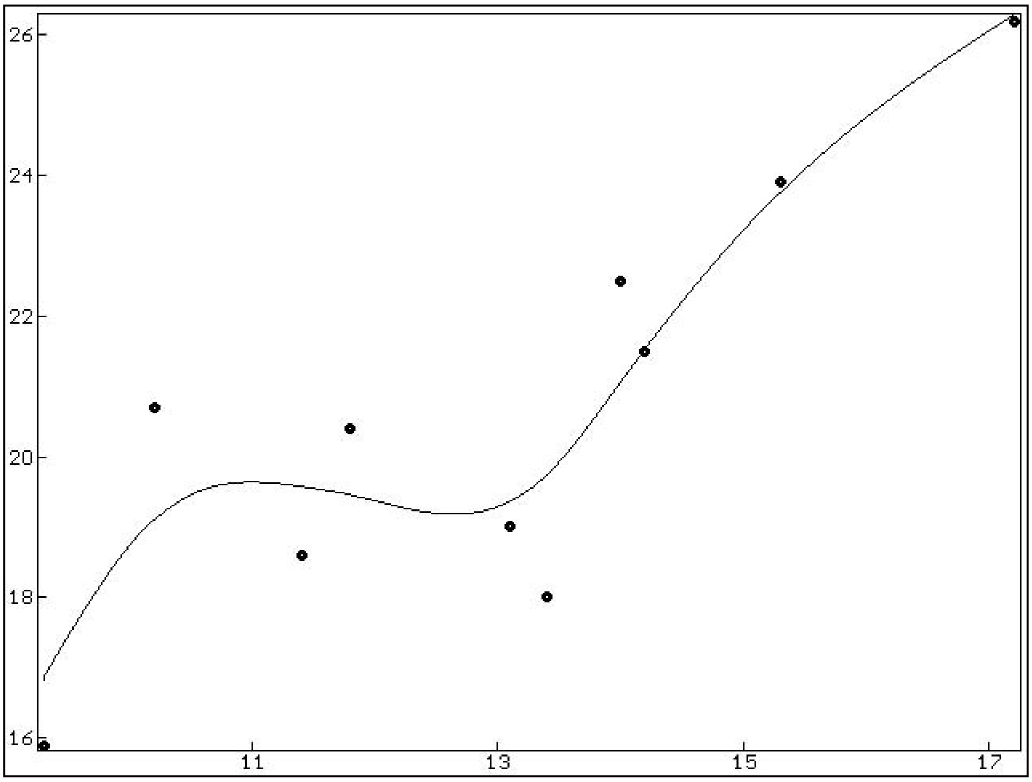

Abbildung 2: Die Lösung von (OP1) zum Minimum μ_{min}=0.0277 von $FE_{(OP1)}(\mu)$

Wird S aus (OP2) als Glättungsparameter benutzt und (3.2) minimiert, ist im Allgemeinen die Rechenzeit größer als bei (3.1). Dem gegenüber stehen zwei Vorteile. Der Parameter S befindet sich im endlichen Intervall $[0,S_{max}]$ und es ergeben sich Möglichkeiten zur Bestimmung eines Startwertes von S. Im vorliegenden Programm wird dieser berechnet, indem durch ein gleitendes Verfahren jeweils zu fünf aufeinander folgenden Punkten (x_j,y_j), (x_{j+1},y_{j+1}),..., (x_{j+4},y_{j+4}), j=1,2,...m-4, des Ausgleichspolynoms $P_j(x)$ vom Grad 3 berechnet wird. Es wird festgelegt $S_{Start} = \sum_{j=1}^{m-4}(y_{j+2} - P_j(x_{j+2}))^2$.

Abbildung 3 zeigt für das obige Beispiel $FE_{(OP2)}(S)$ im Bereich von Null bis S_{max}. Es existieren zwei lokale Minima S_{Min1}=7.529 und S_{Min2}=19.919. Die entsprechenden eindeutigen Lösungen von (OP2) sind in Abbildung 4 dargestellt.

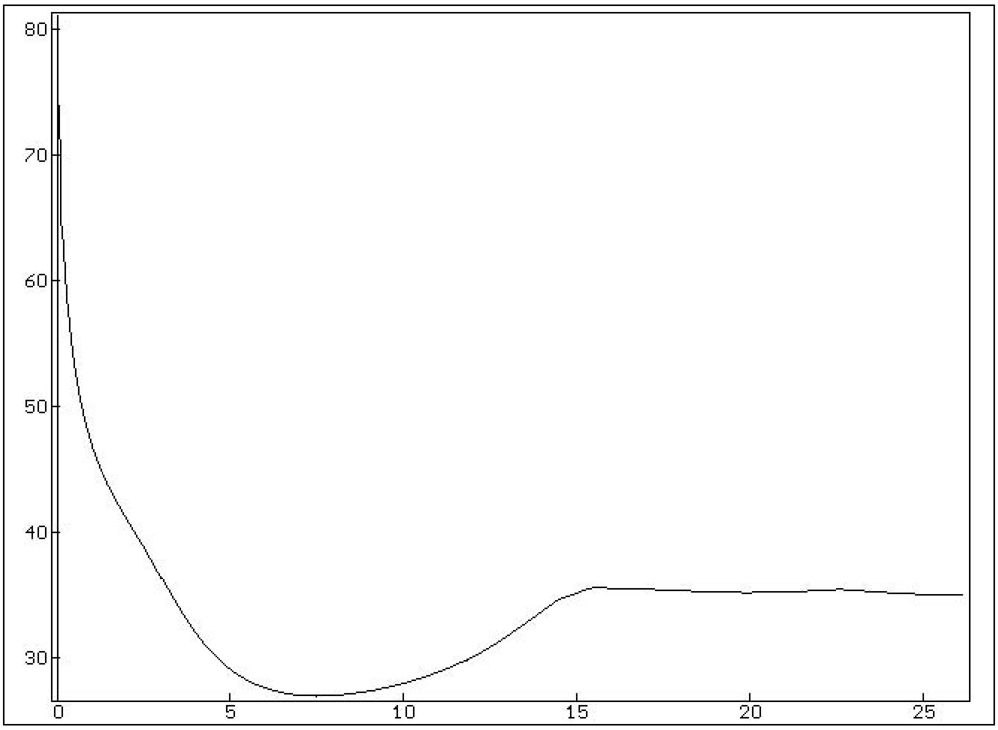

Abbildung 3: $FE_{(OP2)}(S)$ zum Beispieldatensatz

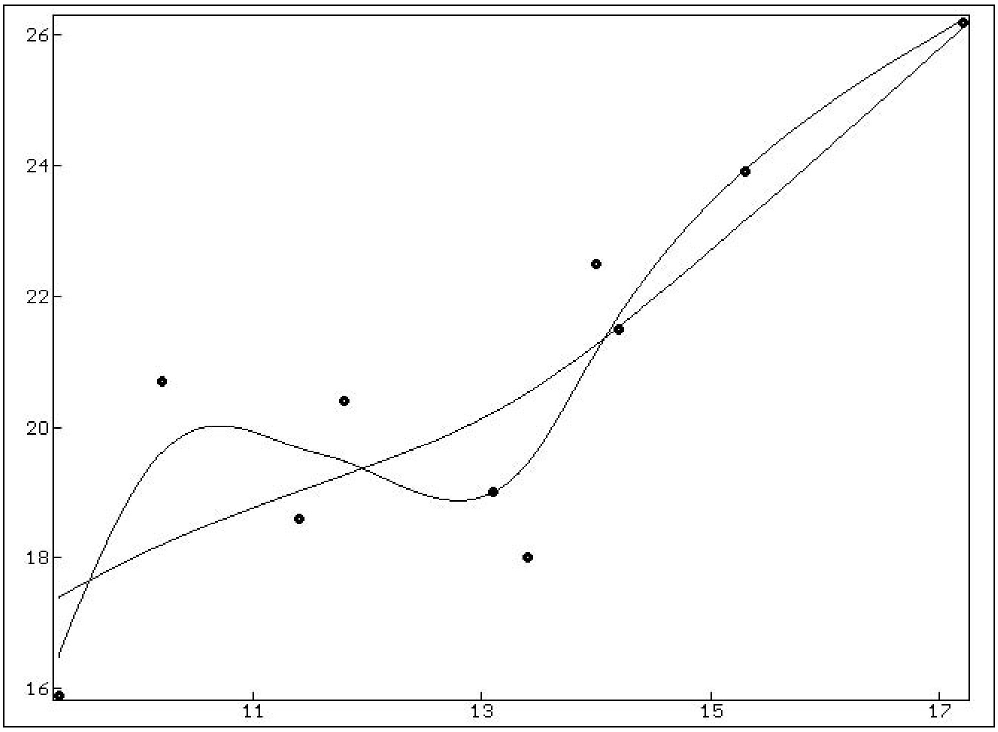

Abbildung 4: Die Lösungen von (OP2) zu S_{min1}=7.529 und S_{min2}=19.919

HASTIE und TIBSHIRANI zeigten 1990, dass sich das zu minimierende Fehlerfunktional von (OP1) als

$$FE_{(OP1)}(\mathbf{:}):=\sum_{q=1}^{m} (y_q - s_q^{\mu}(x_q))^2=\sum_{q=1}^{m}\left(\frac{y_q-s^{\mu}(x_q)}{1-h_{\mu}[q,q]}\right)^2$$

darstellen lässt. Dabei ist $h_{\mu}[q,q]$ das q-te Diagonalelement der Matrix
$H:= [I + 2\mathbf{:}mQ^TBZ^{-1}BQ]^{-1}$, welche zu μ den ausgleichenden Spline $s^{\mu}(x)$ bezüglich ***aller*** Punkte bestimmt (vergleiche (2.8)).
Die generalized cross validation (GCV) ersetzt $h_{\mu}[q,q]$ durch den Mittelwert
$\bar{h}_{\mu}:=\frac{1}{m}\sum_{q=1}^{m}h_{\mu}[q,q]$, so dass sich letztlich

$$FE_{(OP1)}(\mathbf{:}):=\frac{1}{(1-\bar{h}_{\mu})^2}\sum_{q=1}^{m}(y_q-s^{\mu}(x_q))^2 \qquad (3.4)$$

ergibt.
FIT(YX) benutzt bei Wahl von GCV das μ, welches (3.4) minimiert. Man muss allerdings beachten, dass der ausgegebene Glättungsparameter (Smoothing Parameter) $\lambda=2\mu/m$ ist. Die Splinefunktion minimiert dann das Funktional

$$\frac{\lambda}{2}\int_{x_1}^{x_m} f''(x)^2dx + \frac{1}{m}\sum_{i=1}^{m} (f(x_i) - y_i)^2. \qquad (3.5)$$

Analoges gilt für die in IML verfügbare Funktion SPLINEC.
Bei Eingabe des Glättungsparameters λ und der Spezifikation type='zero' wird die eindeutige Lösung von $\frac{\lambda}{2}\int_{x_1}^{x_m} f''(x)^2dx + \sum_{i=1}^{m} (f(x_i) - y_i)^2 = \min!$ berechnet.

Damit ist es möglich, mit den direkten Mitteln von SAS die Größe λ zu schätzen und die entsprechende Lösung von (3.5) über die Splinekoeffizienten zur Verfügung zu stellen.

1. Verwende FIT(YX) mit GCV. Ergebnis ist der smoothing Parameter λ.
2. SPLINEC(fitted, coeff, endval, data, mλ, , 'zero',) liefert die entsprechenden Spline-koeffizienten coeff zur Berechnung von Funktionswerten mittels SPLINEV.

Literatur

1. Greville, T.N.E. (1969). *Theory and Applications of Spline Functions*. ACADEMIC PRESS New York - San Francisco – London
2. Hastie, T.J.; Tibshirani, R.J. (1990). *Generalized additive models.* Monographs on Statistics and Applied Probability 43. London etc.: Chapman and Hall. [ISBN 0-412-34390-8]
3. Holladay, (1957). *A smoothest curve approximation*. Math. Tables Aids Comput. 11, 233-243
4. Reinsch, C.H (1967). *Smoothing by Spline Functions*. Numer. Math. 10, 177-183.
5. Wodny, M. (1998) *Ausgewählte Probleme der Kurvenanpassung.* Biometrie und Medizinische Informatik - Greifswalder Seminarberichte GinkgoPark Mediengesellschaft. [ISBN 3-932463-00-5]
6. SAS Institute Inc. (2001). *Programming with SAS/IML Software*. Cary, NC 27513