Ausgleich mit einer Geraden bzw. einer Ebene bei gemischt achsenparallelen und orthogonalen Abstandsquadraten

Helmut Späth

Gegeben seien in der Ebene oder im Raum gemessene Datenpunkte, durch die eine ausgleichende Gerade bzw. eine Ebene zu legen sei. Bei der üblichen Methode der kleinsten Quadrate wird die Summe der zu y- bzw. z-Achse parallelen Abstände von den Punkten zur Geraden bzw. zur Ebene hin minimiert. Wir betrachten den Fall, dass von einem Teil der Punkte aus alternativ orthogonale Abstandsquadrate verwendet werden, was insbesondere dann sinnvoll ist, falls nicht nur y- bzw. z-Ordinaten, sondern auch die zugehörigen Abszissen x bzw. x und y Fehler haben (können). Ein numerisches Verfahren wird vorgeschlagen, und numerische Beispiele werden angegeben.

1 Problemstellung und Verfahren für eine Gerade

Die gegebenen Datenpunkte seien $(x_i,\ y_i),\quad i \in M = \{1, \ldots,\ m\}$.
Die zu bestimmende Ausgleichsgerade kann in der üblichen Form

$$y = ax + b \tag{1}$$

oder (gleichwertig) in parametrischer Darstellung

$$x = t, y = at + b,\ -\infty \le t \le \infty \tag{2}$$

angesetzt werden. (y-achsenparallele Geraden x = konstant sind in beiden Fällen ausgeschlossen). Wir nehmen an, dass für $i \in C$ nur die y_i (und nicht die x_i) fehlerbehaftet sind und für $i \in D$ sowohl die y_i als auch die x_i Fehler haben (können). Dabei sind C und D Teilmengen von M mit $C \cup D = M$ und $C \cap D = \emptyset$, d.h. jeden Punkt i kann nur einer der beiden Teilmengen, die sich zu M ergänzen, zugeordnet sein. $D = \emptyset$, d.h. üblichen Ausgleich, oder $C = \emptyset$, d.h. nur orthogonale Abstandsgrundrate, ist zugelassen.
Dann ist

$$(ax_i + b - y_i)^2,\ \ i \in C \tag{3}$$

das y-achsenparallele und ist

$$\min_t [at + b - y_i)^2 + (t - x_i)^2],\ \ i \in D \tag{4}$$

das orthogonale Abstandsquadrat vom Punkt $(x_i,\ y_i)$ zur Geradenform (1) bzw. (2). Die Wurzeln dieser Maßzahlen, also die Abstände selbst, sind in Fig. 1 visualisiert.
Ist t der Vektor mit den Komponenten t_i, $i \in D$ so ist die gemischte Abstandsquadratsumme

$$S(a, b, \boldsymbol{t}) = \sum_{i \in C} (ax_i + b - y_i)^2 + \sum_{i \in D} (at_i + b - y_i)^2 + (t_i - x_i)^2 \tag{5}$$

eine sinnvolle Größe zur Minimierung bzgl. a, b und $\boldsymbol{t}$. Der Wert t_i im zweiten Summanden entspricht für festes i demjenigen Wert von t, der (4) minimiert. Dieser ergibt sich für festes i aus der notwendigen Bedingung

$$\frac{1}{2}\frac{\partial S}{\partial t_i} = 0,\ \ i \in D, \boldsymbol{t} = (t_1, \ldots, t_m) \tag{6}$$

für ein Minimum, nämlich

$$a(at_i + b - x_i) + (t_i - x_i) = 0, \tag{7}$$

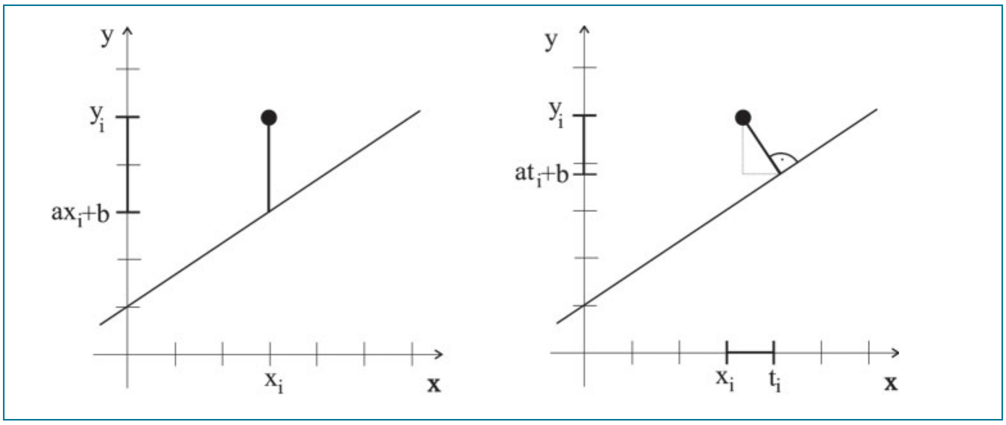

Fig. 1

zu

$$t_i = \frac{x_i - a(b - y_i)}{1 + a^2}, \; i \in D.$$

Die zwei weiteren notwendigen Bedingungen

$$\frac{1}{2}\frac{\partial S}{\partial a} = \frac{1}{2}\frac{\partial S}{\partial b} = 0 \tag{8}$$

liefern das für a und b lineare Gleichungssystem

$$\left(\sum_{i\in C} x_i^2 + \sum_{i\in D} t_i^2\right) a + \left(\sum_{i\in C} x_i + \sum_{i\in D} t_i\right) b = \sum_{i\in C} x_i y_i + \sum_{i\in D} t_i y_i \tag{9}$$

$$\left(\sum_{i\in C} x_i + \sum_{i\in D} t_i\right) a + mb = \sum_{i=1}^{m} y_i,$$

dessen eindeutige Auflösung nach a und b stets möglich ist, da unter Beachtung von $m = |C| + |D|$ die Koeffizientenmatrix von (9) bis auf unrealistische Spezialfälle als Summe von zwei positiv definiten Matrizen betrachtet werden kann, die dann ebenfalls positiv definit ist. Würde man (7) in (9) einsetzen, so erhielte man ein kompliziertes nichtlineares Gleichungssystem für a und b. Einfacher erscheint folgendes, auch anderweitig [1] schon erfolgreich eingesetztes Iterationsverfahren, das bei jedem Schritt die Funktion S i.a. verkleinert, d.h. ein Abstiegsverfahren ist:

Schritt 1:
Bestimme Anfangswerte $a^{(0)}$ und $b^{(0)}$ für a und b (z.B. kann man solche erhalten, indem man, was einfach ist, die Ergebnisse für **nur** achsenparallele Abstandsquadrate benutzt). Setze $l = 0$ (Iterationsindex).
Schritt 2: Bestimme Werte $t_i^{(l)} = t_i$ nach (7) mit $a = a^{(l)}$ und $b = b^{(l)}$.
Schritt 3: Setze in (9) $t_i = t_i^{(l)}$ ein und bestimme durch Auflösung des linearen Gleichungssystems nach a und b neue Näherungen $a^{(l+1)} = a$ und $b^{(l+1)} = b$. Falls kein Abstieg mehr vorliegt, oder eine maximale Anzahl *lmax* von Iterationen überschritten ist, STOP; andernfalls setze $l := l + 1$, berechne für den Vergleich $S = S(a^{(l+1)},\ b^{(l+1)},\ \boldsymbol{t}^{(l)})$ und gehe zurück zu Schritt 2.
Bemerkung: Man könnte auch mit Startwerten $\boldsymbol{t}^{(0)}$ (beliebig) beginnen und Schritt 3 mit Schritt 2 vertauschen.

2 Numerisches Beispiel für eine Gerade

Wir gehen von $m = 12$ Punkten mit den Koordinaten

x	1	2	3	4	5	7	8	9	10	12
y	2	4	1	3	6	5	5	8	6	9

aus und betrachten vier verschiedene Fälle laut Tab. 1

Tab 1

Fall	Mengen C, D	Anzahl der Iterationen	a	b	S
1	$C = M,\ D = \emptyset$	1	.5881	1.3127	15.087
2	$C = \{1,3,5,7,9\}$, $D = \{2,4,6,8,10\}$	5	.5984	1.1770	13.681
3	$C = \{1,2,3,4,5\}$, $D = \{6,7,8,9,10\}$	5	.6027	1.2492	13.781
4	$C = \emptyset,\ D = M$	7	.6465	.9561	10.931

Dabei ist Fall 1 der mit nur y-achsenparallelen und ist Fall 4 der mit nur orthogonalen Abstandsquadraten; die Fälle 2 und 3 haben gemischte Werte. Die Fig. 2 bis 5 enthalten visualisiert jeweils die gegebenen Punkte, die Ausgleichsgeraden für die Fälle 1 bis 4 samt den erhaltenen Abständen. Wie man an Tab. 1 und Fig. 2 bis 5 sieht, sind die Werte für a und b in diesem Beispiel nicht sehr unterschiedlich. Jedoch gilt hier und allgemein, dass mit wachsender Anzahl von orthogonalen Abstandsquadraten die erreichte Fehlerquadratsumme fällt.

3 Problemstellung und Verfahren für eine Ebene

Jetzt sind dreidimensionale Datenpunkte $(x_i,\ y_i,\ z_i)$, $i \in M$ mit Teilmengen C und D von M wie vorher gegeben. Ähnlich wie die Gerade in (1) und (2) kann man eine Ebene in der Form

$$z = ax + by + c \tag{10}$$

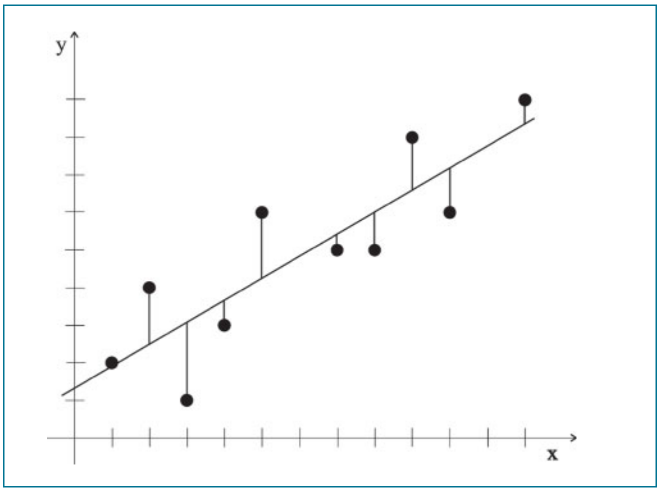

Fig. 2

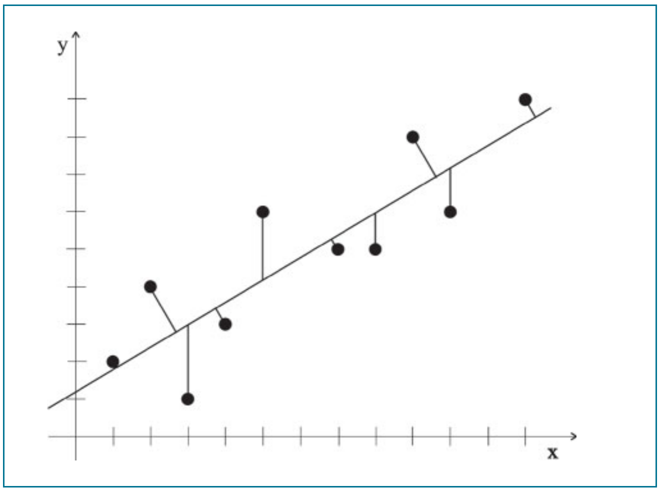

Fig. 3

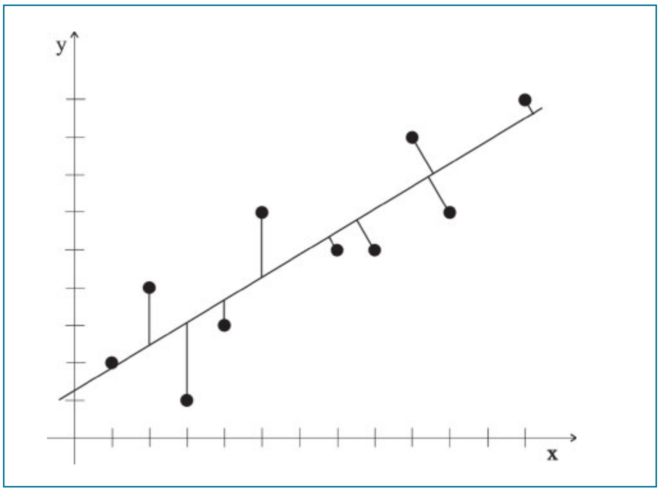

Fig. 4

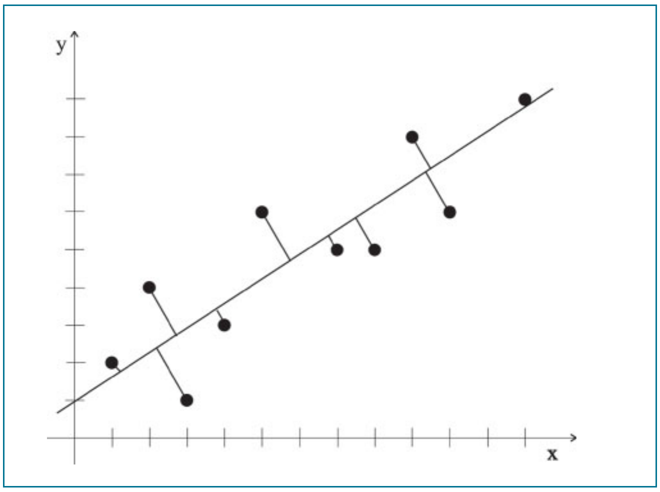

Fig. 5

bzw. parametisch als

$$x = u,\ y = v,\ z = au + bv + c, -\infty \le u,\ v \le \infty \qquad (11)$$

ansetzen, wenn man den Koeffizienten bei z als von Null verschieden voraussetzt. Für die zu minimierende Zielfunktion analog zu (5) ergibt sich jetzt

$$S(a,b,c,\boldsymbol{u},\boldsymbol{v}) = \sum_{i\in C}(ax_i + by_i + c - z_i)^2$$

$$+\sum_{i\in D}(au_i + bv_i + c - z_i)^2 + (u_i - x_i)^2 + (v_i - y_i)^2 \qquad (12)$$

Die Vektoren $\boldsymbol{u}$ und $\boldsymbol{v}$ enthalten die Werte $(u_i,\ v_i)$, $i \in D$. Die notwendigen Bedingungen für ein Minimum

$$\frac{1}{2}\frac{\partial S}{\partial u_i} = \frac{1}{2}\frac{\partial S}{\partial v_i} = 0,\ i \in D$$

für die Flächenparameter u_i und v_i ergeben

$$(1 + a^2)u_i + abv_i = x_i - a(c - z_i),$$

$$abu_i + (1 + b^2)v_i = y_i - b(c - z_i). \qquad (13)$$

Da die Determinante der Koeffizientenmatrix von (13) gleich $1 + a^2 + b^2$ und damit stets von Null verschieden ist, ist dieses lineare Gleichungssystem stets eindeutig lösbar. Die weiteren notwendigen Bedingungen

$$\frac{1}{2}\frac{\partial S}{\partial a} = \frac{1}{2}\frac{\partial S}{\partial b} = \frac{1}{2}\frac{\partial S}{\partial c} = 0$$

für ein Minimum ergeben

$$\left(\sum_{i\in C} x_i^2 + \sum_{i\in D} u_i^2\right)a + \left(\sum_{i\in C} x_i y_i + \sum_{i\in D} u_i v_i\right)b$$

$$+\left(\sum_{i\in C} x_i + \sum_{i\in D} u_i\right)c = \sum_{i\in C} x_i z_i + \sum_{i\in D} v_i z_i,$$

$$\left(\sum_{i\in C} x_i y_i + \sum_{i\in D} u_i v_i\right)a + \left(\sum_{i\in C} y_i^2 + \sum_{i\in D} v_i^2\right)b \qquad (14)$$

$$+\left(\sum_{i\in C} y_i + \sum_{i\in D} v_i\right)c = \sum_{i\in C} y_i z_i + \sum_{i\in D} v_i z_i,$$

$$\left(\sum_{i\in C} x_i + \sum_{i\in D} u_i\right)a + \left(\sum_{i\in C} y_i + \sum_{i\in D} v_i\right)b + mc = \sum_{i=1}^{m} z_i.$$

Dieses lineare Gleichungssystem für a, b und c ist ebenfalls i. a. stets eindeutig lösbar. Der zu dem in Abschnitt 2 analoge Algorithmus lautet dann:

Schritt 1′:
Bestimmte Anfangswerte $a^{(0)}$, $b^{(0)}$ und $c^{(0)}$ ähnlich wie in Schritt 1. Setze $l = 0$.

Schritt 2′:
Bestimmte Werte $u_i^{(l)} = u_i$ und $v_i^{(l)} = v_i$ durch Lösung von (13) für alle $i \in D$.

Schritt 3′:
Setze $v_i = v_i^{(l)}$ und $v_i^{(l)}$, $i \in D$ in (14) ein, löse nach $(a,\ b,\ c) = (a^{(l+1)},\ b^{(l+1)},\ c^{(l+1)})$ auf und fahre wie in Schritt 3 fort.

4 Numerisches Beispiel für eine Ebene

Wir gehen von $m = 12$ Punkten im Raum mit den Koordinaten

x	1	3	0	− 4	7	1	− 2	− 5	1	5	− 6	0
y	4	1	6	− 1	1	2	− 5	− 1	5	0	− 1	4
z	− 4	4	9	1	8	0	11	0	− 6	8	− 9	− 5

aus, wobei ausgehend von x und y die Werte für z durch Vorgabe von $a = 1$, $b = -2$, $c = 3$ bestimmt wurden. Für den Fall 1 $(C = M,\ \ d = \theta)$ Fall 2 $(C = \{1,3,5,7,9,11\},\ D = \{2,4,6,8,10,12\})$ und Fall 3 $(C = \theta,\ D = M)$ wurden nach einer Iteration die Werte a, b und c mit $S = 0$ erwartungsgemäß reproduziert. Jetzt änderten wir obige Werte durch willkürliche Addition von 0, + 1 oder − 1 bei allen drei Koordinatenwerten ab in

x	1	2	0	− 3	7	2	− 2	− 6	1	4	− 6	− 1
y	5	1	5	− 1	2	2	− 6	− 1	6	0	− 2	4
z	− 4	3	− 9	2	8	− 1	11	1	− 6	9	− 1	− 6

Dann ergab sich für die Fälle 1 bis 3 folgende Tabelle

Fall	Anzahl der Iterationen	a	b	c	S
1	1	1.1162	– 1.8779	3.0238	29.760
2	9	1.2589	– 1.8603	2.8063	13.302
3	31	1.1933	– 1.9855	3.1646	4.909

5 Ausblick

Das für die Gerade $y = ax + b$ (Polynom von Grad 1) angegebene Verfahren kann für Polynome

$$y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \tag{15}$$

vom Grad $n > 1$ verallgemeinert werden. Dabei treten allerdings erhebliche numerische Probleme auf. Das orthogonale Abstandsquadrat (4) lautet hier

$$\min_t [(a_n t^n + \ldots + a_1 t + a_0 - y_i)^2 + (t - x_i)^2], \; i \in D \tag{16}$$

Die Ableitung dieses Ausdrucks nach t ist

$$\frac{1}{2}[n a_n t^{n-1} + (n-1) a_{n-1} t^{n-2} + \ldots + 2 a_2 t + a_1]^*$$

$$[a_n t^n + \ldots + a_1 t + a_0 - y_i] + \frac{1}{2}(t - x_i). \tag{17}$$

Da der Grad hiervon $2n - 1$ und somit ungerade ist, existiert wenigstens eine reelle Nullstelle. Falls es mehrere reelle Nullstellen geben sollte, dann ist für t_i diejenige reelle Nullstelle mit dem kleinsten Wert für (16) zu wählen. Treten die gesuchten Parameter nichtlinear auf, so wird es schwierig. Bei einem Kreis mit Mittelpunkt (a, b) und Radius r z.B., der durch

$$(x - a)^2 + (y - b)^2 = r^2$$

beschrieben wird, ist die explizite Darstellung

$$y = b \pm \sqrt{r^2 - (x - a)^2}$$

und die parametrische Darstellung

$$x = a + r, \; y = b + rsint.$$

Die (5) entsprechende Zielfunktion hat dann die sehr unangenehme Eigenschaft, dass ihre partiellen Ableitungen nach a, b und r nichtlinear in diesen Größen sind.

Literatur

[1] [1] H. Späth: Zur numerischen Berechnung der Trägheitsgeraden und der Trägheitsebene. AVN, 7/2004, 273–275.

Anschrift des Verfassers:
Prof. Dr. Helmuth *Späth*, Fakultät V, Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, Postfach 25 03, D-26111 Oldenburg, Germany. e-mail: spaeth@mathematik.uni-oldenburg.de