

Teilredundanzen und ihre natürlichen Verallgemeinerungen

Ronald Jurisch und Georg Kampmann

Zusammenfassung

Unser Beitrag beschäftigt sich mit einer konzeptionellen Einbettung der bekannten Teilredundanzen der linearen Ausgleichsrechnung in das Werkzeug der Plücker-Koordinaten. Verblüffende Wechselwirkungen sowohl im Hinblick auf die statistische Testtheorie wie auch im Hinblick auf die Interpretation geometrischer Besonderheiten von Beobachtungs-(Design-)Konstellationen werden damit sichtbar.

Summary

Introducing Pluecker-Co-ordinates for the analysis of partial redundancies within linear regression adjustment reveals amazing relationships between statistical parameters and geometrical features of observations geometry.

Wir beginnen mit unseren Darlegungen bei dem bekannten linearen Modell der Ausgleichsrechnung, der vermittelnden Ausgleichung nach der Methode der kleinsten Quadrate

$$\begin{aligned} Ax &= l + v & D(l) &= \sigma^2 P^{-1} \\ \text{Zielfunktion: } v^T P v &\rightarrow \text{Min} \end{aligned} \quad (1)$$

worin n = Anzahl der (reduzierten) Beobachtungen, u = Anzahl der Unbekannten, $n > u$, $A = (n, u)$ -Koeffizientenmatrix mit $\text{rg } A = u$, $x = (u, 1)$ -Vektor der (reduzierten) Unbekannten, $l = (n, 1)$ -Vektor der (reduzierten) Beobachtungen, $v = (n, 1)$ -Vektor der Verbesserungen (= Residuen), $D(l) = (n, n)$ -Varianz-Kovarianz-Matrix der Beobachtungen a priori, σ^2 = Varianz der Gewichtseinheit, P^{-1} = symmetrische, positiv definite (n, n) -Varianz-Kovarianz-Matrix der Beobachtungen a priori, die gleich der Varianz-Kovarianz-Matrix der Verbesserungen a priori ist.

Für die weiteren Betrachtungen sei die a priori Varianz-Kovarianzmatrix der Beobachtungen als identisch zur Einheitsmatrix angesehen.

Ausgangsobjekte im Sinne einer geometrischen Analyse des Modells (1) sind diejenigen orthogonalen Projektionsmatrizen H bzw. $I-H$, die auf den Unterraum U (aufgespannt von den Spalten von A) bzw. dessen orthogonales Komplement U^\perp projizieren:

$$H = A(A^T A)^{-1} A^T, \quad R = I - H. \quad (2)$$

Von besonderem Interesse sind dabei die entsprechenden Hauptdiagonalelemente:

$$h_{i,i} \text{ bzw. } r_{i,i} = 1 - h_{i,i} \quad i = 1, \dots, n. \quad (3)$$

Der Grund liegt dabei in der Tatsache, dass große Werte von $h_{i,i}$ (nahe bei 1) auf so genannte Hebelpunkte (high-leverage-points) hinweisen »sollen«.

Analog könnte man auch nach kleinen Werten von $r_{i,i}$ suchen (nahe bei 0). Interessant ist jedoch die Tatsache, dass in der Literatur fast durchgehend der erste Weg bevorzugt wird. In Chatterjee and Hadi (1988) wird dies besonders deutlich.

Die $r_{i,i}$ verdienen dort noch nicht einmal eine eigenständige Bezeichnung. Nur in der geodätischen Literatur zur Auswertetechnik scheinen die $r_{i,i}$ (Teilredundanzen) sich besonderer Beachtung zu erfreuen.

Doch auch hier werden die Hebelpunkte durch große Werte von $h_{i,i}$ meist charakterisiert. Den Grund hierfür findet man in der klassischen Regression (Geradengleichung), da hier die Ursachen für große Werte von $h_{i,i}$ geometrisch anschaulich auftreten. Abszissenwerte, die weit von der Masse entfernt sind, erzeugen große Werte für $h_{i,i}$, sind also Hebelpunkte:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & a \end{pmatrix}^T \\ \Rightarrow h_{4,4} &= \frac{(a-3)^2 + (a-2)^2 + (a-1)^2}{6 + (a-1)^2 + (a-2)^2 + (a-3)^2} \\ \Rightarrow h_{4,4} &\rightarrow 1 \quad \text{für } a \rightarrow \infty \end{aligned} \quad (4)$$

In Chatterjee and Hadi (1988) wird versucht, diese Tatsache auch auf allgemeinere Fälle zu verallgemeinern (S. 25, »Conditions for large Values h_{ii} «).

Folgende Darstellung wird dort für $h_{i,i}$ angegeben:

$$h_{i,i} = a_i^T a_i \sum_{k=1}^u \alpha_k^{-1} \cos^2 \Theta_{i,k} \quad (5)$$

$a_i \hat{=}$ Zeilen von A $i = 1, \dots, n$
 $\alpha_k \hat{=}$ Eigenwerte von $A^T A$ $k = 1, \dots, u$
 $\Theta_{i,k} \hat{=}$ Winkel zwischen a_i und dem normalisierten Eigenvektor ϑ_k zu α_k .

Damit ergeben sich folgende Bedingungen für große Werte von $h_{i,i}$ (siehe auch Cook and Weisberg (1982) S. 3, in Chatterjee and Hadi (1988) zitiert):

- $a_i^T a_i$ ist groß in Bezug auf die restlichen Zeilen in A (siehe (4))
- a_i hat die Richtung eines Eigenvektors, der zu einem kleinen Eigenwert gehört.

Der Fall b) ist jedoch mit Vorsicht zu betrachten. Die $h_{i,i}$ sind geometrisch invariant gegenüber einem Basiswechsel in U , die Eigenwerte von $A^T A$ jedoch nicht!

Wir erläutern diesen Fakt an einem transparenten Beispiel:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ a & a & a & a & b \end{pmatrix}^T \quad (6)$$

$\Rightarrow h_{5,5} = 1$ unabhängig von a und b ($a \neq b$).

Die Ursache für $h_{5,5} = 1$ liegt also nicht darin begründet, dass $b \gg a$ gilt. Es müsste also der Fall b) vorliegen. Für $a = 1, b = 2$ erhält man:

$$A^T A = \begin{pmatrix} 5 & 6 \\ 6 & 8 \end{pmatrix} \quad \alpha_1 = 0.315 \quad \alpha_2 = 12.685$$

$$\vartheta_1 = \begin{pmatrix} -0.788 \\ 0.615 \end{pmatrix} \quad \vartheta_2 = \begin{pmatrix} 0.615 \\ 0.788 \end{pmatrix}$$

$$\Theta_{5,1} = 78.58^\circ$$

Allerdings kann nicht behauptet werden, dass α_5 und ϑ_1 fast parallel sind. Geht man jedoch zur Normalform von A (spezieller Basiswechsel) über (siehe Jurisch et. al. (1999)):

$$\bar{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

so erhält man

$$\bar{A}^T \bar{A} = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Rightarrow \bar{\alpha}_1 = 4 \quad \bar{\alpha}_2 = 1$$

$$\bar{\vartheta}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \bar{\vartheta}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\Rightarrow \bar{\vartheta}_2 \parallel \bar{a}_5 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

also doch der Fall b).

Das Beispiel (6) wurde schon des Öfteren in der Literatur untersucht (Chatterjee and Hadi (1988) S. 98, Behnken and Draper (1972), Draper and Smith (1981)).

Da hier jedoch kein Hebelpunkt im Sinne von a) ($b \gg a$) vorliegen muss, wurde für Beobachtungen dieser Art der Begriff der einflussreichen Beobachtung (influential observation) eingeführt. Dass dennoch beide Fälle a) und b) auf die gleiche Ursache zurückzuführen sind, zeigt sich erst durch Einführung der Plücker-Koordinaten (Jurisch et. al. (1999), Jurisch und Kampmann (2001)). Es gilt:

$$h_{i,i} = \frac{\sum d_i^2}{\sum d^2} \quad r_{i,i} = \frac{\sum d_{(i)}^2}{\sum d^2} \quad (7)$$

$$D = \sum d^2 = \det(A^T A).$$

Hierin sei d_i eine beliebige Plücker-Koordinate, die die i -te Beobachtung enthält, $d_{(i)}$ eine solche, die die i -te Beobachtung nicht enthält. Summiert wird über alle Möglichkeiten der Auswahl der entsprechenden Plücker-Koordinaten.

In (4) erhält man:

$$d_{1,2} = 1, d_{1,3} = 2, d_{1,4} = a-1, d_{2,3} = 1, d_{2,4} = a-2, d_{3,4} = a-3.$$

Also sind alle Plücker-Koordinaten, die die 4-te Beobachtung nicht enthalten, unabhängig von a :

$$r_{4,4} = \frac{d_{1,2}^2 + d_{1,3}^2 + d_{2,3}^2}{D}$$

$$= \frac{6}{6 + (a-1)^2 + (a-2)^2 + (a-3)^2} \xrightarrow{n \rightarrow \infty} 0.$$

In (6) wird sofort ersichtlich, dass $d_{i,j} = 0$ für $i, j \neq 5$ gilt:

$$r_{5,5} = 0.$$

Aus (7) wird weiterhin ersichtlich, dass bei der Suche nach Hebelpunkten eigentlich $r_{i,i}$ zu bevorzugen ist, da bekanntlich eine Quadratsumme genau dann klein ist, falls alle Summanden klein sind. Eine entsprechend einfache Erklärung für das Großwerden von $h_{i,i}$ (nahe bei 1) gibt es offensichtlich nicht.

Unmittelbar aus (7) ergibt sich eine weitere Darstellung:

$$r_{i,i} = \frac{\det A^T A_{(i)}}{\det A^T A}. \quad (8)$$

Hierbei wird $A^T A_{(i)}$ ohne die i -te Zeile von A berechnet.

Gilt $r_{i,i} = 0$, so entsteht durch Streichen der i -ten Beobachtung ein Rangabfall in A (die Normalgleichungsmatrix wird singular). Numerisch liegt dieser Fall aber auch schon für kleine Werte von $r_{i,i}$ (nahe bei 0) vor. $\det(A^T A)$ hängt hierbei mit dem Volumen des Konfidenz-ellipsoids der Unbekannten zusammen (siehe Chatterjee and Hadi (1988) S. 117 ff., S. 136 ff.), so dass (8) den Wechsel dieses Volumens beschreibt, falls die i -te Beobachtung gestrichen wird (omission-approach). Es stellt eine bekannte Tatsache dar, dass bei Vorliegen von Gruppen von Hebelpunkten sogenannte Maskierungseffekte (masking effects) auftreten, so dass die Hebelpunkte nicht mehr direkt aus den entsprechenden Teilredundanzen verifizierbar sind.

Durch (7) und (8) ist jedoch auch für diesen Fall eine erste Verallgemeinerung der klassischen Teilredundanzen

auf ganz natürliche Weise gegeben. Sei dazu I eine Indexmenge. Dann definiert man die gemeinsame Teilredundanz der Beobachtungen, die zu I gehören, durch:

$$r_I = \frac{\sum d_{(I)}^2}{\sum d^2} = \frac{\det A^T A_{(I)}}{\det A^T A} \quad (9)$$

Hierin ist $d_{(I)}$ eine Plücker-Koordinate, die keine der Beobachtungen, die zu I gehören, enthält. Analog wird $A^T A_{(I)}$ erklärt. Dies soll nun an zwei Beispielen erläutert werden:

Beispiel 1:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ a & a & a & a & b & b \end{pmatrix}^T$$

$$\Rightarrow r_{i,i} = \frac{3}{4} \quad \text{für } i = 1, 2, 3, 4 \quad r_{5,5} = r_{6,6} = 0.5$$

(Maskierungseffekt)

jedoch: $I = \{5, 6\} \Rightarrow r_I = 0$

$$I = \{i, j\}; i, j \neq 5, 6 \Rightarrow r_I = \frac{1}{2}$$

$$I = \{i, j\}; i \neq 5, 6; j = 5, 6 \Rightarrow r_I = \frac{3}{8}$$

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ a & a + \varepsilon & a + 2\varepsilon & a + 3\varepsilon & b & b + \varepsilon \end{pmatrix}^T$$

(für $\varepsilon = 0$ erhält man obiges Beispiel)

es gilt: $r_{5,5}, r_{6,6} \rightarrow \frac{1}{2}$ für $b \rightarrow \infty$ (Gruppe von Hebelpunkten)

jedoch: $I = \{5, 6\} \rightarrow r_I = \frac{20\varepsilon^2}{D} \rightarrow 0$ für $b \rightarrow \infty$

$$I = \{1, 2\} \rightarrow r_I = \frac{20\varepsilon^2 + 4(b-a)^2 - 16\varepsilon(b-a)}{D} \rightarrow \frac{1}{2}$$

für $b \rightarrow \infty$

$$D = 41\varepsilon^2 + 8(b-a)^2 - 18\varepsilon(b-a)$$

Die gemeinsamen Teilredundanzen decken die Hebelpunkt-Wirkung einer Gruppe von Beobachtungen deutlich auf!

Beispiel 2:

Hierbei handelt es sich um ein Beispiel aus [10]. Betrachtet wird hierin eine Geraden-Ausgleichung mit neun Beobachtungen:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 21 & 22 & 23 & 41 & 42 & 43 \end{pmatrix}^T$$

	1	2	3	4	5	6	7	8	9
$\Rightarrow r_i$	0.706	0.723	0.739	0.888	0.889	0.888	0.739	0.723	0.706

Die Teilredundanzen sind verhältnismäßig groß und unterscheiden sich nur in geringem Maße (alle Beobachtungen kontrollieren sich scheinbar gut). Andererseits liegen offenbar aber drei Gruppen von jeweils drei Beobachtungen vor.

Geht man nun zur Betrachtung der (verallgemeinerten) gemeinsamen Teilredundanzen über, so erhält man zunächst für Zweier-Gruppen z. B.:

$$I = \{1, 2\} \text{ bzw. } \{8, 9\} \Rightarrow r_I = 0.428$$

$$I = \{4, 5\} \text{ bzw. } \{5, 6\} \Rightarrow r_I = 0.777$$

Die Unterschiede zwischen verschiedenen Gruppen werden deutlicher. Für Gruppen von jeweils drei Beobachtungen wird klar erkennbar, dass die ersten drei bzw. letzten drei Beobachtungen Hebelpunkt-Wirkung aufweisen:

$$I = \{1, 2, 3\} \text{ bzw. } \{7, 8, 9\} \Rightarrow r_I = 0.167$$

$$I = \{4, 5, 6\} \Rightarrow r_I = 0.666$$

$$I = \{1, 2, 4\} \Rightarrow r_I = 0.352$$

$$I = \{5, 6, 9\} \Rightarrow r_I = 0.522$$

Dass bei der Einführung der gemeinsamen Teilredundanzen die Nebendiagonalelemente der Projektionsmatrizen benutzt werden, wird durch folgende Darstellung deutlich. Es gilt:

$$r_I = \det R_I \quad (10)$$

Hierin ist R_I diejenige Submatrix von R , die den Indizes aus I entspricht:

$$R_I = (r_{i,j})_{i,j \in I}$$

Für $I = \{i, j\}$ erhält man z. B.:

$$r_I = \det \begin{pmatrix} r_{i,i} & r_{i,j} \\ r_{i,j} & r_{j,j} \end{pmatrix} = r_{i,i}r_{j,j} - r_{i,j}^2$$

d. h. $r_I = 0$ bedeutet $\frac{r_{i,j}^2}{r_{i,i}r_{j,j}} = 1!$

Die Darstellung (10) ist auch von praktischer Bedeutung. Man kann die gemeinsamen Teilredundanzen sehr effektiv ohne den Umweg der Plücker-Koordinaten berechnen. Auf einen weiteren wichtigen Aspekt muss an dieser Stelle hingewiesen werden. Beobachtungsgruppen mit $r_I = 0$ sind gerade die in Jurisch et. al. (1999) definierten latenten Restriktionen m -ter Ordnung ($m: \hat{=}$ Kardinalzahl von I). Diese stellen also »reine« Hebelpunkte dar. Jetzt ist jedoch auch die lange Zeit offene Frage nach der Verallgemeinerung dieser latenten Restriktionen geklärt (kleine Werte von r_I).

Nun soll eine weitere Verallgemeinerung der klassischen Teilredundanz vorgenommen werden. Dazu bezeichne \bar{A} die mit dem Beobachtungsvektor I geränderte Matrix:

$$\bar{A} = (A | I).$$

Damit lassen sich die folgenden Projektionsmatrizen erklären:

$$\bar{H} = \bar{A}(\bar{A}^T \bar{A})^{-1} \bar{A}^T \quad \text{bzw.} \quad \bar{R} = I - \bar{H} \quad (11)$$

Die Darstellung ist allerdings nicht ganz so neu. Man findet sie bereits in Chatterjee and Hadi (1988) S. 104 ff.. Es wird allerdings wieder mit \bar{H} anstatt \bar{R} gearbeitet. Es gilt nun:

$$\bar{H} = H + \frac{\hat{v}\hat{v}^T}{\hat{v}^T\hat{v}}, \quad (12)$$

d. h. \bar{H} ist eine so genannte Rang-1-Modifikation von H ($\hat{v} \hat{=}$ geschätzte Verbesserungen). Entsprechend gilt für die Hauptdiagonal-Elemente $\bar{h}_{i,i}$:

$$\bar{h}_{i,i} = h_{i,i} + \frac{\hat{v}_i^2}{\hat{v}^T\hat{v}}. \quad (13)$$

Gesucht wird nun wiederum nach großen Werten von $\bar{h}_{i,i}$ (nahe bei 1). Aus (13) leitet sich der Hauptkritikpunkt an den $\bar{h}_{i,i}$ ab. Sie können nicht unterscheiden zwischen Hebelpunkten ($h_{i,i}$ nahe bei 1) und Beobachtungen, die große standardisierte Verbesserungen (Ausreißer) haben. Wir schlagen nun vor, statt (13) die so genannten erweiterten Teilredundanzen $\bar{r}_{i,i}$ zu benutzen:

$$1 - \bar{h}_{i,i} = \bar{r}_{i,i} = r_{i,i} - \frac{\hat{v}_i^2}{\hat{v}^T\hat{v}}. \quad (14)$$

Aus (14) erhält man unmittelbar:

$$0 \leq r_{i,i} - \bar{r}_{i,i} = \frac{\hat{v}_i^2}{\hat{v}^T\hat{v}}. \quad (15)$$

Die standardisierten Verbesserungen sind also nach oben durch die rein geometrisch definierten Teilredundanzen $r_{i,i}$ beschränkt. Dadurch lässt sich ganz klar aufzeigen, wodurch ein Ausreißertest, der diese standardisierten Verbesserungen benutzt, versagen muss. Sei dazu nur ein grober Fehler in der i -ten Beobachtung vorhanden und es gelte $r_{i,i} < \frac{1}{n}$ (Hebelpunkt). Dann gilt dies auch für die entsprechende standardisierte Verbesserung ($\frac{\hat{v}_i^2}{\hat{v}^T\hat{v}} < \frac{1}{n}$).

Da die Summe der standardisierten Verbesserungen jedoch 1 beträgt, muss es ein $j \neq i$ geben, so dass $\frac{\hat{v}_j^2}{\hat{v}^T\hat{v}} > \frac{1}{n}$ ist, der Fehler also nicht detektiert wird.

Wir wollen nun auf die eigentliche Bedeutung der $\bar{r}_{i,i}$ näher eingehen. Dazu stellen wir zunächst den Zusammenhang mit den Plücker-Koordinaten \bar{d} der Matrix \bar{A} her.

Es gilt in Analogie zu (7):

$$\bar{r}_{i,i} = \frac{\sum \bar{d}_{(i)}^2}{\sum \bar{d}^2}. \quad (16)$$

Die Ursache für die Kleinheit von $\bar{r}_{i,i}$ besteht gerade darin, dass alle Plücker-Koordinaten $\bar{d}_{(i)}$, die die i -te Beobachtung nicht enthalten, klein sind. Dies wiederum

kann nun eine geometrische bzw. eine statistische Ursache haben. Wir zerlegen dazu den Beobachtungsvektor $l = \bar{l} + \epsilon$ in die »wahren« Werte \bar{l} (Erwartungswert) und den Fehlervektor ϵ . Dann gilt:

$$\bar{d}(l) = \bar{d}(\epsilon), \quad (17)$$

d. h. die Plücker-Koordinaten \bar{d} ändern sich nicht, falls man l durch den Fehlervektor ϵ ersetzt. Die \bar{d} können also als lokale Fehlerindikatoren (sie enthalten nur diejenigen Fehler in den Beobachtungen, die zur Bildung von \bar{d} herangezogen werden) interpretiert werden. Ist insbesondere nur eine dieser Beobachtungen fehlerbehaftet, so gilt:

$$\epsilon_i = \frac{\bar{d}_{1,\dots,i_u,i}}{d_{1,\dots,i_u}} \quad (18)$$

$$(\epsilon_i = \dots = \epsilon_{i_u} = 0).$$

Aus dem Entwicklungssatz (Laplace) für Determinanten erhält man nun folgende Ursachen für die Kleinheit von $\bar{r}_{i,i}$ in (16):

1) Alle Beobachtungen, die den Index i nicht enthalten, weisen kleine Fehler auf (im Grenzfall fehlerfrei):

$$\epsilon_j \gg \epsilon_i \quad j \neq i \text{ (grober Fehler)}$$

2) Alle Plücker-Koordinaten $d(i)$, die die i -te Beobachtung nicht enthalten, sind klein (im Grenzfall gleich 0):

$$r_{i,i} \ll 1 \text{ bzw. } r_{i,i} = 0 \text{ (Hebelpunkt).}$$

Im Fall 2) ist die eigentliche Problematik bei der Grobfehlersuche enthalten. Hebelpunkte verdecken grobe Fehler (Maskierungseffekt), dies wird im Weiteren noch deutlicher aufgezeigt.

Wir untersuchen nun einen weiteren Aspekt im Zusammenhang mit den Teilredundanzen $r_{i,i}$ und $\bar{r}_{i,i}$. Es gilt:

$$\frac{\bar{r}_{i,i}}{r_{i,i}} = \frac{\det(\bar{A}_{(i)}^T \bar{A}_{(i)})}{\det(A^T A)} = \frac{\hat{v}^T \hat{v}_{(i)}}{\hat{v}^T \hat{v}}. \quad (19)$$

Das Verhältnis der Teilredundanzen ist also gleich dem Verhältnis der Quadratsummen der Verbesserungen, wobei $\hat{v}^T \hat{v}_{(i)}$ die Schätzung ohne die i -te Beobachtung bedeutet.

Ist nun $\bar{r}_{i,i}$ klein, aber $r_{i,i}$ groß (gut kontrollierte Beobachtung), so ergibt sich durch Streichen der i -ten Beobachtung eine erhebliche Verkleinerung der Quadratsumme der Verbesserungen.

Dies lässt sich natürlich auch anhand der Schätzungen der Varianz der Gewichtseinheit verdeutlichen:

$$\frac{\hat{\sigma}_{(i)}^2}{\hat{\sigma}^2} = \frac{n-u}{n-u-1} \cdot \frac{\bar{r}_{i,i}}{r_{i,i}} \quad (20)$$

($\hat{\sigma}_{(i)}^2$) bezeichnet die Schätzung der Varianz der Gewichtseinheit ohne die i-te Beobachtung)

Liegt jedoch $r_{i,i}$ in derselben Größenordnung wie $\bar{r}_{i,i}$, so ändert sich durch Streichung der i-ten Beobachtung die Quadratsumme der Verbesserungen nur wenig. Dies bedeutet jedoch, dass Hebelpunkte grobe Fehler maskieren. Zur Grobfehlersuche gibt es in der Literatur (Chatterjee and Hadi (1988)) eine Vielzahl verschiedener Diagnose-Tools. Wir wollen nun aufzeigen, dass sich die meisten davon als einfache Funktionen des Verhältnisses der Teilredundanzen in (19) darstellen lassen:

1) (Intern) studentisiertes Residuum

$$\bar{v}_i^2 = \frac{\hat{v}_i^2}{r_{i,i} \hat{\sigma}^2} = (n-u) \left(1 - \frac{\bar{r}_{i,i}}{r_{i,i}} \right) \quad (21)$$

2) (Extern) studentisiertes Residuum (jackknifed residual)

$$v_i^{*2} = \frac{\hat{v}_i^2}{r_{i,i} \hat{\sigma}_{(i)}^2} = (n-u-1) \left(\frac{r_{i,i}}{\bar{r}_{i,i}} - 1 \right) \quad (22)$$

3) Cooks Distance

$$C_i = \frac{1}{u} \frac{1-r_{i,i}}{r_{i,i}} \bar{v}_i^2 = \frac{n-u}{u} \cdot \frac{1-r_{i,i}}{r_{i,i}} \cdot \left(1 - \frac{\bar{r}_{i,i}}{r_{i,i}} \right) \quad (23)$$

4) Cooks Distance (verallgemeinert)

$$\begin{aligned} C_i^{*2} &= v_i^{*2} \frac{1-r_{i,i}}{r_{i,i}} \frac{n-u}{u} \\ &= \frac{(n-u)(n-u-1)}{u} \cdot \frac{1-r_{i,i}}{r_{i,i}} \cdot \left(\frac{r_{i,i}}{\bar{r}_{i,i}} - 1 \right) \end{aligned} \quad (24)$$

usw. Grobe Fehler sollen nun anhand großer Werte der Diagnosetools (21)–(24) detektiert werden. Dies ist offensichtlich gleichbedeutend mit der Suche nach kleinen Werten von $\frac{\bar{r}_{i,i}}{r_{i,i}}$ bzw. großen Werten von $\frac{r_{i,i}}{\bar{r}_{i,i}}$, also kleinen Werten von $\bar{r}_{i,i}$. Allerdings kann auch $r_{i,i}$ klein sein (Hebelpunkt), so dass die Ergebnisse verfälscht werden. Dies kann zum völligen Versagen des Ausreißertests führen, wie folgendes kleine (aber instruktive) Beispiel zeigt:

Beispiel 3:

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & a \end{pmatrix}$$

$$1^T = (2+\varepsilon \quad 3-\varepsilon \quad 4+\varepsilon \quad 5-\varepsilon \quad a+1+\delta)$$

Hierin soll durch $a \in \mathbb{R}$ ein Hebelpunkt simuliert werden ($a \gg 1$). Durch δ soll ein grober Fehler in der fünften Beobachtung ($\delta \gg \varepsilon$) gegeben sein.

Zunächst erhält man folgende Plücker-Koordinaten:

$$d_{1,2} = 1, d_{1,3} = 2, d_{1,4} = 3, d_{1,5} = a-1, d_{2,3} = 1, d_{2,4} = 2, d_{2,5} = a-2, d_{3,4} = 1, d_{3,5} = a-3, d_{4,5} = a-4$$

$$\bar{d}_{1,2,3} = 4\varepsilon, \bar{d}_{1,2,4} = 4\varepsilon, \bar{d}_{1,2,5} = \delta + 2\varepsilon \left(a - \frac{3}{2} \right), \bar{d}_{1,3,4} = -4\varepsilon,$$

$$\bar{d}_{1,3,5} = 2(\delta - \varepsilon), \bar{d}_{1,4,5} = 3\delta + 2\varepsilon \left(a - \frac{5}{2} \right), \bar{d}_{2,3,4} = -4\varepsilon,$$

$$\bar{d}_{2,3,5} = \delta - 2\varepsilon \left(a - \frac{5}{2} \right), \bar{d}_{2,4,5} = 2(\delta + \varepsilon), \bar{d}_{3,4,5} = \delta + 2\varepsilon \left(a - \frac{7}{2} \right)$$

Die entsprechenden Teilredundanzen $r_{i,i}$ bzw. $\bar{r}_{i,i}$ sind:

$$\begin{aligned} r_{1,1} &= \frac{3(a-3)^2 + 8}{D} \\ r_{2,2} &= \frac{3\left(a - \frac{8}{3}\right)^2 + \frac{56}{3}}{D} \\ r_{3,3} &= \frac{3\left(a - \frac{7}{3}\right)^2 + \frac{56}{3}}{D} \\ r_{4,4} &= \frac{3(a-2)^2 + 8}{D} \\ r_{5,5} &= \frac{20}{D} \end{aligned} \quad D = 4\left(a - \frac{5}{2}\right)^2 + 25$$

Für $a \rightarrow \infty$ gilt:

$$r_{i,i} \rightarrow \frac{3}{4} \quad i = 1, 2, 3, 4; \quad r_{5,5} \rightarrow 0 \quad (\text{Hebelpunkt})$$

Des Weiteren gilt:

$$\begin{aligned} \bar{r}_{1,1} &= \frac{2}{D} \cdot \left(\varepsilon^2 (4(a-3)^2 + 11) + 3\delta^2 + 2\delta\varepsilon \right) \\ \bar{r}_{2,2} &= \frac{2}{D} \cdot \left(\varepsilon^2 (4(a-3)^2 + 11) + 7\delta^2 + 2\delta\varepsilon(4a-13) \right) \\ \bar{r}_{3,3} &= \frac{2}{D} \cdot \left(\varepsilon^2 (4(a-2)^2 + 11) + 7\delta^2 + 2\delta\varepsilon(4a-7) \right) \\ \bar{r}_{4,4} &= \frac{2}{D} \cdot \left(\varepsilon^2 (4(a-2)^2 + 11) + 3\delta^2 - 2\delta\varepsilon \right) \\ \bar{r}_{5,5} &= \frac{2}{D} \cdot (32\varepsilon^2) \end{aligned} \quad (26)$$

$$\bar{D} = 4 \left(4\varepsilon^2 \left(\left(a - \frac{5}{2} \right)^2 + 5 \right) + 5\delta^2 + 4\delta\varepsilon \left(a - \frac{5}{2} \right) \right)$$

Für $a \rightarrow \infty$ gilt:

$$\bar{r}_{i,i} \rightarrow \frac{3}{4} \quad i = 1, 2, 3, 4; \quad \bar{r}_{5,5} \rightarrow 0$$

Das Verhältnis $\frac{\bar{r}_{5,5}}{r_{5,5}}$ nähert sich dann also einem unbestimmten Ausdruck der Form 0/0 (die Testgrößen (21)–(24) sind nicht mehr definiert). Untersucht man den Ausdruck $\frac{\bar{r}_{5,5}}{r_{5,5}}$ eingehender, so erhält man:

$$\lim_{n \rightarrow \infty} \frac{\bar{r}_{5,5}}{r_{5,5}} = \frac{4}{5}$$

und damit:

	$\frac{\bar{r}_{i,i}}{r_{i,i}}$	$\frac{\hat{v}_i^2}{\hat{v}^T \hat{v}}$	\bar{v}_i^2 (21)	v_i^2 (22)	C_i (23)	C_i^2 (24)
1	2/3	1/4	1	1	1/6	1/2
2	2/3	1/4	1	1	1/6	1/2
3	2/3	1/4	1	1	1/6	1/2
4	2/3	1/4	1	1	1/6	1/2
5	4/5	0	3/5	1/2	∞	∞

d. h. die standardisierten bzw. studentisierten Residuen decken den groben Fehler nicht auf.

Eine genauere Aussage zum Versagen dieser Tests erhält man, indem man das Verhältnis $\gamma = \frac{\delta}{\epsilon}$ einführt. Untersucht man die

Gleichheit $\frac{\bar{r}_{5,5}}{r_{5,5}} = \frac{\bar{r}_{i,i}}{r_{i,i}}$ in Abhängigkeit von γ , so erhält man:

$$\frac{\bar{r}_{5,5}}{r_{5,5}} = \frac{\bar{r}_{1,1}}{r_{1,1}} \Leftrightarrow \gamma_{1/2} = \frac{-1 \pm 2\sqrt{\frac{3(a-3)^2 + 8}{5}}}{3} \sim \pm 0.6a$$

$$\frac{\bar{r}_{5,5}}{r_{5,5}} = \frac{\bar{r}_{2,2}}{r_{2,2}} \Leftrightarrow \gamma_{1/2} = \frac{-4(a-13) \pm 6\sqrt{\frac{3\left(a-\frac{8}{3}\right)^2 + \frac{56}{3}}{5}}}{7} \sim 0.17a \text{ bzw. } -1.34a$$

$$\frac{\bar{r}_{5,5}}{r_{5,5}} = \frac{\bar{r}_{3,3}}{r_{3,3}} \Leftrightarrow \gamma_{1/2} = \frac{-4(a-7) \pm 6\sqrt{\frac{3\left(a-\frac{7}{3}\right)^2 + \frac{56}{3}}{5}}}{7} \sim 0.17a \text{ bzw. } -1.34a$$

$$\frac{\bar{r}_{5,5}}{r_{5,5}} = \frac{\bar{r}_{4,4}}{r_{4,4}} \Leftrightarrow \gamma_{1/2} = \frac{1 \pm 2\sqrt{\frac{3(a-2)^2 + 8}{5}}}{3} \sim \pm 0.6a$$

Das Fehlerverhältnis $\frac{\delta}{\epsilon}$ muss also hinreichend groß sein, damit die Fehler durch (21) bzw. (22) aufgedeckt werden.

Aus dem bisher Dargelegten wird deutlich, dass die Ausreißersuche zum einen die Aufdeckung von Hebeln im erweiterten Design (kleine Werte von $\bar{r}_{i,i}$) und von Hebeln im ursprünglichen Sinne (kleine Werte von $r_{i,i}$) beinhalten muss. Liegen insbesondere Ausreißer in starken Hebeln vor, so kann dies zum Versagen entsprechender Diagnosetools führen.

Eine andere Ursache für ein solches Versagen liegt in den Maskierungseffekten bei Vorliegen von mehreren groben Fehlern (Gruppen von Hebeln). Wir betrachten dazu das Beispiel in Koch (1997) Beispiel 2: Hier liegen grobe Fehler

Fall A) in den Beobachtungen 4, 5, 6

$$I^T = (2, 3, 4, 10, 10, 10, 42, 43, 44) \\ \epsilon_4 = -12, \epsilon_5 = -13, \epsilon_6 = -14$$

Fall B) in den Beobachtungen 7, 8, 9

$$I^T = (2, 3, 4, 22, 23, 24, 30, 30, 30) \\ \epsilon_7 = -12, \epsilon_8 = -13, \epsilon_9 = -14$$

Untersucht werden diese Fälle mit einem M-Schätzer bzw. modifizierten M-Schätzer nach Huber (Koch (1997) S. 290 ff.).

Im Fall A deckt der M-Schätzer die Fehler auf, der modifizierte M-Schätzer »verlegt« die Fehler in die Beobachtungen 7, 8 und 9. Im Fall B deckt der modifizierte M-Schätzer die richtigen Fehler auf, der M-Schätzer verlegt sie jedoch in die Beobachtungen 4, 5 und 6.

In den nachfolgenden Tabellen wird dies anhand der Testgrößen (21)–(24) verifiziert.

Die studentisierten Residuen decken die Fehler in den entsprechenden Spitzenbeträgen auf (unterstrichen), Cook's Distance bzw. Cook's Distance (verallgemeinert) decken jedoch nur den größten Fehler ϵ_6 auf. (Fall A)

Die studentisierten Residuen verlegen den Fehler in die Beobachtungen 5 und 6, Cook's Distance entdeckt nur den groben Fehler ϵ_9 . Beachtet man weiterhin, dass Cook's Distance den zusätzlichen Faktor $\frac{1-r_{i,i}}{r_{i,i}}$ enthält (also Hebeln stärker gewichtet), so erscheint dasselbe Bild wie bei den M-Schätzern. Anhand der Teilredundanzen sieht man deutlich die Maskierungseffekte (weder $r_{i,i}$ noch $\bar{r}_{i,i}$ ist wirklich klein!). (Fall B)

Will man nun diese Maskierungseffekte aufheben, muss man in Analogie zu (9) bzw. (10) die Teilredundanzen $\bar{r}_{i,i}$ auf Gruppen von Beobachtungen verallgemeinern. Sei dazu I eine Indexmenge von m Beobachtungen.

Dann erhält man:

$$\bar{r}_I = \frac{\det \bar{A}_{(I)}^T \bar{A}_{(I)}}{\det \bar{A}^T \bar{A}} = \frac{\sum \bar{d}_{(I)}^2}{\sum \bar{d}^2} = \det \bar{R}_I \quad (27)$$

Im Beispiel 2 aus Koch (1997) gilt nun:

$$\bar{r}_I = 0 \quad \text{für } I = \{4, 5, 6\} \quad \text{im Fall A} \\ \text{bzw.} \\ \bar{r}_I = 0 \quad \text{für } I = \{7, 8, 9\} \quad \text{im Fall B.}$$

Interessant ist weiterhin, dass sich die in der Literatur aufgeführten Diagnosetools für diesen Fall (Multiple

Fall A:

	$r_{i,i}$	$\bar{r}_{i,i}$	$\frac{\bar{r}_{i,i}}{r_{i,i}}$	\bar{v}_i^2 (21)	v_i^{*2} (22)	C_i (23)	C_i^{*2} (24)
1	0.7056	0.6508	0.9223	0.5437	0.5052	0.1134	0.7377
2	0.7226	0.6678	0.9242	0.5309	0.4924	0.1019	0.6616
3	0.7388	0.6840	0.9258	0.5192	0.4807	0.0918	0.5949
4	0.8885	0.7156	0.8054	<u>1.3622</u>	<u>1.4497</u>	0.0855	0.6368
5	0.8889	0.6680	0.7515	<u>1.740</u>	<u>1.9841</u>	0.1088	<u>0.8680</u>
6	0.8885	0.6137	0.6907	<u>2.165</u>	<u>2.6867</u>	<u>0.1359</u>	<u>1.1801</u>
7	0.7388	0.6832	0.9248	0.5268	0.4883	0.0931	0.6043
8	0.7226	0.6670	0.9231	0.5386	0.5001	0.1034	0.6721
9	0.7056	0.6499	0.9211	0.6536	0.5142	0.1053	0.7508

Fall B:

	$\bar{r}_{i,i}$	$\frac{\bar{r}_{i,i}}{r_{i,i}}$	\bar{v}_i^2 (21)	v_i^{*2} (22)	C_i (23)	C_i^{*2} (24)
1	0.6322	0.8974	0.7183	0.6860	0.1498	1.0017
2	0.6679	0.9243	0.5299	0.4914	0.1017	0.6603
3	0.6993	0.9465	0.3743	0.3389	0.0662	0.4194
4	0.7013	0.7893	<u>1.5185</u>	<u>1.6016</u>	0.0953	0.7035
5	0.6701	0.7539	<u>1.7230</u>	<u>1.9591</u>	0.1077	0.8571
6	0.6357	0.7155	<u>1.9917</u>	<u>2.3860</u>	0.1250	1.0481
7	0.7129	0.9649	0.2454	0.2118	0.0434	0.2621
8	0.6680	0.9244	0.5289	0.4904	0.1015	0.6589
9	0.6115	0.8666	0.9335	0.9233	<u>0.1947</u>	<u>1.3482</u>

Case) wiederum in einfachster Weise durch die verallgemeinerten Teilredundanzen r_i bzw. \bar{r}_i darstellen lassen:

$$1) \bar{v}_i^2 = \frac{\hat{v}_i^T (\mathbf{E} - \mathbf{H}_i)^{-1} \hat{v}_i}{\hat{\sigma}^2} = (n - u) \left(1 - \frac{\bar{r}_i}{r_i} \right) \quad (28)$$

$$2) v_i^{*2} = (n - u - m) \left(\frac{r_i}{\bar{r}_i} - 1 \right) \quad (29)$$

$$3) C_i = \frac{\hat{v}_i^T (\mathbf{E} - \mathbf{H}_i)^{-1} \mathbf{H}_i (\mathbf{E} - \mathbf{H}_i)^{-1} \hat{v}_i}{u \hat{\sigma}^2} \quad (30)$$

$$= \frac{n - u}{u} \cdot \frac{1 - r_i}{r_i} \cdot \left(1 - \frac{\bar{r}_i}{r_i} \right)$$

$$4) C_i^{*2} = \frac{(n - u)(n - u - m)}{u} \cdot \frac{1 - r_i}{r_i} \cdot \left(\frac{r_i}{\bar{r}_i} - 1 \right). \quad (31)$$

Es sei zum Schluss angemerkt, dass diese grundlegenden neuen Betrachtungen auch für andere Ausprägungen der linearen Ausgleichsrechnung (wie die bedingte, vermittelnde Ausgleichung mit Restriktionen usw.) ihre Gültigkeit behalten. Bekanntlich lassen sich die Modelle ohne Ansehung der Zielfunktion, d. h. zielfunktionsinvariant ineinander überführen, siehe Kampmann (1992).

Literatur

Behnken, D.W., Draper, N.R.: Residuals and their Variance. Technometrics 11, No. 1, pp 101-111, 1972.
 Chatterjee, S., Hadi, A.S.: Sensitivity Analysis in Linear Regression. John Wiley and Sons, New York, 1988.
 Cook, R.D., Weisberg, S.: Residuals and Influence in Regression. Chapman and Hall, London, 1982.
 Jurisch, R., Kampmann, G.: Plücker-Koordinaten – Ein neues Hilfsmittel zur Geometrie-Analyse und Ausreißersuche, in: First International Symposium on Robust Statistics and Fuzzy Techniques in Geodesy and GIS, Swiss Federal Institut of Technology Zürich, IGP-Bericht Nr. 295, 2001.
 Jurisch, R., Kampmann, G., Linke, J.: Über die Analyse von Beobachtungen in der Ausgleichsrechnung. Teil I und II, ZfV 124, S. 350-357, und S. 388-395, 1999
 Kampmann, G.: Zur numerischen Überführung verschiedener linearer Modelle der Ausgleichsrechnung. ZfV 117, S. 278-287, 1992.
 Koch, K.R.: Parameterschätzung und Hypothesentests. Ferdinand Dümmlers Verlag, Bonn, 1997.

Anschrift der Autoren

Prof. Dr. rer. nat. Ronald Jurisch
 Prof. Dr.-Ing. Georg Kampmann

Hochschule Anhalt (FH)
 Fachbereich Vermessungswesen
 Postfach 2215
 06818 Dessau