BAYERISCHE AKADEMIE DER WISSENSCHAFTEN
MATHEMATISCH-NATURWISSENSCHAFTLICHE KLASSE

SITZUNGSBERICHTE

JAHRGANG

1958

MÜNCHEN 1958

VERLAG DER BAYERISCHEN AKADEMIE DER WISSENSCHAFTEN

In Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

Über approximative Nomographie. I

Von **Georg Aumann** in München

Vorgelegt am 12. Dezember 1958

Im folgenden wird die allgemeine Aufgabe der Nomographie neu formuliert im Sinne eines Approximationsproblems, womit die Nomographie als „approximative Nomographie" in die allgemeine Variationsrechnung eingegliedert erscheint.

Als ein erstes Beispiel zu dieser neuartigen Fragestellung wird die Aufgabe behandelt, eine reelle Matrix (a_{ik}), $i = 1, \ldots, p$; $k = 1, \ldots, q$, durch eine Matrix der Form $(x_i + y_k)$ zu approximieren, d. h. reelle Zahlen $x_1, \ldots, x_p, y_1, \ldots, y_q$ so zu bestimmen, daß der maximale Fehler $s = \max_{i,k} |a_{ik} - (x_i + y_k)|$ zu einem Minimum wird. Die Aufgabe erweist sich als ein Problem der Linearprogrammierung. Es wird ein Verfahren angegeben, wie man in endlich vielen Schritten zu allen Lösungen der Aufgabe kommt. Für numerische Zwecke aber ist dieses Verfahren bei größeren Zeilen- und Spaltenanzahlen unbequem. Dieser Mangel wird durch den Umstand wettgemacht, daß ein infinitäres Verfahren, ein konvergenter iterativer Algorithmus (die „alternierende Symmetrisierung einer Matrix") zur Verfügung steht, womit sich das Minimum von s und gewisse ausgezeichnete Lösungen des Problems mit beliebiger Genauigkeit bestimmen lassen.

Über das entsprechende Problem im Kontinuum, eine stetige Funktion $f(x, y)$ durch eine Summe $a(x) + b(y)$ möglichst gut zu approximieren, soll in einer zweiten Note berichtet werden.

1. Kritik an der herkömmlichen Nomographie. Dem Zweck der Nomographie, zu einer vorgegebenen Funktion f ein graphisch-mechanisches Verfahren bereitzustellen, mit dessen Hilfe der Wert von f rasch und mit hinreichender Genauigkeit ermittelt werden kann, entspricht man in herkömmlicher Weise durch die Behandlung der folgenden beiden Aufgaben: (A) der

mehr theoretischen Frage, ob und wie die Funktion f in exakter Weise durch ein Nomogramm darstellbar ist, und (B) der mehr praktischen, wie im Falle einer solchen Darstellung das Verfahren einzurichten ist, um die geforderte Genauigkeit zu erreichen. Ein rigoroser Praktiker könnte dagegen folgendes einwenden: *Erstens* ergeben sich bei der praktischen Anwendung eines Verfahrens sowieso nur Werte, die mit Fehlern behaftet sind; es ist also falsche Vorsicht zu verlangen, daß das Verfahren, rein theoretisch betrachtet, die Funktion genau darstellt. Es kommt, abgesehen von einer gewissen Handlichkeit des Verfahrens, nur darauf an, daß es f hinreichend genau beschreibt. *Zweitens* kann die Frage nach der exakten Darstellung überhaupt nur dann gestellt werden, wenn die Funktion formelmäßig gegeben ist. Zwar könnte man, falls etwa die Funktion f in Gestalt einer Tabelle vorliegt, f durch einen hinreichend genauen Rechenausdruck $\bar{f}$ ersetzen; aber je nach Wahl von $\bar{f}$ wird man auf die Frage (A) verschiedene Antworten erhalten.

Eine auf die praktische Anwendung ausgerichtete Nomographie muß daher grundsätzlich anders verfahren: Man sucht nicht nach einem nomographischen Verfahren, welches die Funktion f theoretisch exakt darstellt, sondern man nimmt ein handliches Verfahren her und prüft, mit welcher Güte man damit unter Ausnutzung der im Verfahren enthaltenen Variationsmöglichkeiten (an Skalen, Kurven und Konstanten usw.) die Funktion f approximieren kann. Liegt der Approximationsfehler unter der geforderten Genauigkeit, so hat man eine für die Praxis verwertbare Beschreibung von f.

2. Läßt man sich von diesem Gedanken leiten, so nimmt die *allgemeine Aufgabe der Nomographie* die folgende Gestalt an:

Es sei M ein bestimmter nomographischer Mechanismus und $\Phi = \mathfrak{N}(M, D)$ die Gesamtheit aller jener Funktionen φ mit einem festen Definitionsbereich D, welche sich mittels M exakt darstellen lassen durch geeignete Wahl der in M verfügbaren Skalen, Kurven und Konstanten. Aufgabe ist, zu prüfen, wie gut sich eine Funktion f mit dem Definitionsbereich D durch ein $\varphi \in \Phi$ approximieren läßt. Die Güte der Approximation von f durch φ wird man beurteilen nach der Kleinheit eines aus dem

Fehler $f-\varphi$ errechenbaren „Approximationsmaßes" $A(f-\varphi)$. Z. B. kann man für A nehmen das Supremum des absoluten Fehlers $|f-\varphi|$ auf D (T-[Tschebyscheff-]*Approximation*), oder das mittlere Fehlerquadrat $\int_D |f-\varphi|^2 dm$ (G-[Gauß-]*Approximation*), u. a. m. Die Minimalisierung von $A(f-\varphi)$ stellt ein Variationsproblem dar. Bei dieser Auffassung der Nomographie sind gerade die Fälle, die die traditionelle Nomographie nicht diskutiert, die interessanten, nämlich jene, wo f durch ein φ weder exakt darstellbar noch beliebig genau approximierbar ist, wo also $\inf\{A(f-\varphi): \varphi \in \Phi\}$ positiv ist.

Nomographie von der eben beschriebenen Art nenne ich *approximative Nomographie*. Der Anschluß an klassische Problemstellungen ist leicht zu finden, wenn man die Abgrenzung der approximativen Nomographie etwas weiter steckt, als es oben geschehen ist: Wir werden nicht verlangen, daß die Gesamtheit Φ mit einem System $\mathfrak{N}(M, D)$ übereinstimmt, sondern nur voraussetzen, daß Φ eine gewisse Gesamtheit von Funktionen φ mit demselben Definitionsbereich D bedeutet. Bei dieser allgemeineren Auffassung gehört z. B. das klassische Tschebyscheffsche Problem zur approximativen Nomographie:

Φ ist die Gesamtheit T_n aller Polynome φ vom Grad $\leq n$,

$$\varphi(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

mit beliebigen Koeffizienten $a_0, \ldots, a_n$, betrachtet im Intervall $D = \{x: -1 \leq x \leq 1\}$; zu vorgegebener Funktion $f(x)$ ist jenes Polynom $\varphi \in T_n$ zu bestimmen, für welches

$$\sup\{|f(x)-\varphi(x)|: x \in D\}$$

minimal ausfällt.

3. Bei Funktionen mehrerer Veränderlichen aber treten neuartige Probleme auf. Z. B. kann man die Aufgabe stellen, die Kurven und Skalen einer Fluchtentafel mit einer x-, y- und z-Leiter so zu bestimmen, daß eine vorgegebene Funktion $z = f(x, y)$ durch diese Fluchtentafel möglichst gut approximiert wird.

Wir untersuchen im folgenden ein wesentlich einfacheres Problem, nämlich den Fall, für eine tabellierte Funktion $z = f(x, y)$, wo also die Argumente x und y nur endlich vieler Werte fähig sind, eine Fluchtentafel mit geradlinigen parallelen Leitern, wobei die z-Leiter überdies regelmäßig sein soll, zu konstruieren, die im Sinne einer T-Approximation eine beste ist. Die analytische Formulierung dieser Aufgabe führt zum folgenden „Matrixproblem".

4. Das Matrixproblem. Gegeben ist die reelle Matrix $A = (a_{ik})$, $i = 1, \ldots, p$; $k = 1, \ldots, q$. Gesucht sind reelle Zahlen x_i, y_k, welche $S := \max_{i,k} | a_{ik} - (x_i + y_k) |$ minimalisieren.

Die Existenz einer Lösung, etwa mit $x_1 = 0$, ist leicht einzusehen. Da nämlich $s := \inf S \leq \max_{i,k} | a_{ik} | =: a^*$, so darf man sich bei der Suche nach einem Minimum $(x_i; y_k)$ auf $| a_{1k} - y_k | \leq a^*$ beschränken, oder wenigstens auf

$$| y_k | \leq 2a^*, \tag{1}$$

und dann weiter auf

$$| x_i | \leq 3a^*. \tag{2}$$

Auf dem durch (1) und (2) gekennzeichneten Bereich nimmt die stetige Funktion $S(x_1, \ldots, y_q)$ ihr Minimum an.

5. Um die Gesamtheit der Lösungen des Matrixproblems zu bestimmen, geben wir ihm die folgende Gestalt:

Unter allen möglichen Zahlen s und Vektoren $x = (x_1, \ldots, x_p)$, $y = (y_1, \ldots, y_q)$, für welche

$$| a_{ik} - (x_i + y_k) | \leq s \qquad \text{für alle } i, k \tag{3}$$

gilt, solche zu bestimmen, wo s minimal ist. Wir trennen (3) in

$$x_i + y_k \leq a_{ik} + s \qquad \text{für alle } i, k, \tag{4}$$

$$-x_i - y_{k'} \leq -a_{ik'} + s \qquad \text{für alle } i, k'.^1 \tag{5}$$

---

[1] Hier erkennt man, daß wir es mit einem *Problem des Linearprogrammierens* zu tun haben, nämlich die lineare Funktion $f(x_1, \ldots, y_q, s) = s$ unter den Nebenbedingungen (4) und (5) zu minimalisieren. Wir wenden zur

Wenn es x, y, s gibt, die (4) und (5) erfüllen, dann muß

$$y_k - y_{k'} \leq a_{ik} - a_{ik'} + 2s \qquad \text{für alle } i, k, k' \tag{6}$$

gelten. Umgekehrt kann man aus dem Bestehen von (6) auf die Existenz eines x schließen, welches (4) und (5) befriedigt. Schreiben wir nämlich (6) in der Form

$$a_{ik'} - y_{k'} - s \leq a_{ik} - y_k + s,$$

so folgt daraus

$$\underline{x}_i := \max_{k'} (a_{ik'} - y_{k'} - s) \leq \min_{k} (a_{ik} - y_k + s) =: \bar{x}_i.$$

Wählen wir dann irgendwelche x_i mit

$$\underline{x}_i \leq x_i \leq \bar{x}_i \qquad \text{für alle } i,$$

so folgt

$$a_{ik'} - y_{k'} - s \leq x_i \leq a_{ik} - y_k + s \qquad \text{für alle } i, k, k',$$

was mit (4) und (5) übereinstimmt.

6. Unser Augenmerk hat sich also auf die Lösung von (6) zu richten bei gleichzeitiger Minimalisierung von s. Aus (6) folgt

$$y_k - y_{k'} \leq c_{kk'} + 2s \qquad \text{für alle } k, k', \tag{7}$$

wobei $c_{kk'} := \min_i (a_{ik} - a_{ik'})$ gesetzt ist. Umgekehrt folgt aus (7) wiederum (6). Schreiben wir in (7) anstelle von k,k' der Reihe nach

---

Lösung nicht wie üblich die *Simplexmethode* (G. B. Dantzig, Maximization of a linear function of variables subject to linear inequalities, Chapter XXI in Activity Analysis of Production and Allocation, ed. by T. C. Koopmans, New York 1951) an, sondern die *Eliminationsmethode* zur Auflösung von linearen Ungleichungen, in welchen nur das Zeichen $\leq$ vorkommt. Die Elimination, etwa der Unbekannten x', besteht darin, daß man alle Ungleichungen, welche x' wirklich enthalten, auf die Form $x' \leq g_i$ oder $-x' \leq h_j$ bringt und diese Ungleichungen ersetzt durch $0 \leq g_i + h_j$. Diese Bedingungen enthalten x' nicht mehr und bilden mit den x' nicht enthaltenden Ungleichungen des ursprünglichen Systems ein System mit einer Unbekannten weniger.

$$(8) \qquad (k_0, k_1), (k_1, k_2), \ldots, (k_{m-1}, k_m)$$

und addieren die betreffenden Ungleichungen, so folgt

$$(9) \qquad y_{k_0} - y_{k_m} \leq \sum_{\mu=1}^{\mu} c_{k_{\mu-1}, k_\mu} + 2ms,$$

was offensichtlich eine Verallgemeinerung von (7) darstellt. Wir nennen (8) eine *m-Kette* $K_m(k_0, k_m)$, welche von k_0 nach k_m geht; es ist dann klar, was gemeint ist, wenn wir (9) in der Form

$$(10) \qquad y_k - y_{k'} \leq \sum_{K_m(k, k')} c + 2ms \qquad \text{für jedes } K_m$$

schreiben. In (10) gehen wir zum Infimum über durch Einführung von

$$(11) \qquad d_{kk'}(s) := \inf \Big(\sum_{K_m(k, k')} c + 2ms \Big),$$

das Infinum gebildet über alle m-Ketten $K_m(k, k')$ von k nach k', $m = 1, 2, \ldots$, und erhalten

$$(12) \qquad y_k - y_{k'} \leq d_{kk'}(s) \qquad \text{für alle } k, k',$$

was umgekehrt wieder mit (6) äquivalent ist.

7. Aus dem Bestehen von (12) für ein gewisses y folgt

$$(13) \qquad 0 \leq d_{kk}(s) \text{ für alle } k, \text{ weiter}$$

$$(14) \qquad 0 \leq \min_k d_{kk}(s) =: M(s);$$

für $M(s)$ können wir auch

$$M(s) = \inf \Big(\sum_{C_m} c + 2ms \Big)$$

setzen, wobei C_m die Gesamtheit aller „geschlossenen" m-Ketten (8) mit $k_0 = k_m$ durchläuft. In (14) haben wir eine notwendige Bedingung für die Lösbarkeit der Ungleichungen (6) in y und s erhalten; sie wird sich auch als hinreichend erweisen.

Zunächst sehen wir, daß für $s > a^*$ wegen $c_{kk'} \leq 2a^*$ die Funktion $M(s)$ positiv und endlich ist, und daß für $s < -a^*$

$c_{kk'} \geq -2a^*$ sich $M(s) = -\infty$ ergibt. Ferner ist $M(s)$ nichtfallend, was man unmittelbar erkennt, und nach oben halbstetig; in der Tat, ist $L > M(s_0)$, so gibt es ein C_m mit $\sum_{C_m} c + 2ms_0 < L$; alsdann ist auch $\sum_{C_m} c + 2ms < L$ und damit auch $M(s) < L$ für alle s in einer gewissen Umgebung von s_0. Aus diesen Eigenschaften von M folgt die Existenz eines kleinsten s^* mit $0 \leq M(s^*)$.

8. Wir setzen nun im folgenden $M(s) \geq 0$ voraus. Dies hat weiter zur Folge, daß $s \geq 0$. Wäre nämlich $s < 0$, so folgte, wegen

$$c_{kk} = 0 \quad \text{und} \quad c_{kk'} + c_{k'k''} \leq c_{kk''}, \tag{15}$$

was leicht zu beweisen ist, $c_{kk'} + c_{k'k} + 2s \leq 0 + 2s < 0$, also $M(s) < 0$. Ferner gilt:[1]

$$|d_{kk'}| \leq 2a^* + 2s \tag{16}$$

$$d_{kk'} + d_{k'k''} \geq d_{kk''} \qquad \text{für alle } k, k', k''. \tag{17}$$

In der Tat: $d_{kk'} \leq c_{kk'} + 2s \leq 2a^* + 2s$. Wäre andererseits $d_{kk'} < -2a^* - 2s$, so gäbe es eine m-Kette K_m von k nach k' mit $\sum_{K_m} c + 2ms < -2a^* - 2s$. Schließen wir K_m durch Hinzunahme des Gliedes (k', k) zur geschlossenen Kette C_{m+1}, so folgt $\sum_{C_{m+1}} c + 2(m+1)s = \sum_{K_m} c + c_{kk'} + 2(m+1)s < -2a^* + c_{kk'} \leq 0$, also $M(s) < 0$, was unserer Voraussetzung widerspricht.

Zu positivem ε gibt es eine m-Kette K von k nach k' und eine m'-Kette K' von k' nach k'', so daß $d_{kk'} > \sum_{K} c + 2ms - \varepsilon$ und $d_{k'k''} > \sum_{K'} c + 2m's - \varepsilon$. Fügt man K und K' zu einer $(m + m')$-Kette K'' von k nach k'' zusammen, so folgt $d_{kk'} + d_{k'k''} > \sum_{K''} c + 2(m + m')s - 2\varepsilon \geq d_{kk''} - 2\varepsilon$, was mit $\varepsilon \to 0$ zur zweiten Behauptung führt.

---

[1] Wir lassen im folgenden bei $d_{kk'}(s)$ das Argument s weg.

9. Nun können wir die Auflösung von (12) schrittweise durchführen. Wir haben nach (12) für y_q die Ungleichungen

$$(18) \qquad -d_{kq} + y_k \leq y_q \leq d_{qk'} + y_{k'} \qquad \text{für } (k \,\&\, k') < q.$$

Sind bereits Werte $y_1, \ldots, y_{q-1}$ gefunden, für welche die linke Seite von (18) stets $\leq$ der rechten Seite ist, d. h. für die

$$(19) \qquad y_k - y_{k'} \leq d_{kq} + d_{qk'} \qquad \text{für } (k \,\&\, k') < q$$

gilt, so läßt sich y_q im Durchschnitt aller abgeschlossenen Intervalle $[-d_{kq} + y_k,\ d_{qk'} + y_{k'}]$, $(k \,\&\, k') < q$, der dann ein gewisses abgeschlossenes (ev. nur einpunktiges) Intervall darstellt, frei wählen. Erfüllen aber die $y_1, \ldots, y_{q-1}$ die Ungleichungen (12), so auch (19) wegen (17). Es genügt also, die Ungleichungen (12) nur für $(k \,\&\, k') < q$ zu betrachten. Damit ist der erste Reduktionsschritt getan, und es bleibt nach dem $(q-2)$-ten Schritt nur noch die Ungleichung

$$-d_{21} \leq y_1 - y_2 \leq d_{12}$$

zu lösen. Hier dürfen wir y_1 beliebig wählen, und dann y_2 frei im Intervall $y_1 - d_{12} \leq y_2 \leq y_1 + d_{21}$.
Damit ist bewiesen der

Satz. *Für die Auflösbarkeit von (3) nach x und y ist $M(s) \geq 0$ notwendig und hinreichend; das oben beschriebene Verfahren liefert alle Lösungen $(x; y)$ von (3). Wählt man insbesondere für s den Minimalwert s^*, d. h. den kleinsten Wert s^* mit $M(s^*) \geq 0$, so erhält man alle Lösungen des Matrixproblems.*

10. Als Ergänzung zum bisherigen Ergebnis zeigen wir noch:
Das in 4. bis 9. angegebene Verfahren zur Auflösung von (3) sowie die Bestimmung von s^* ist in endlich vielen Schritten durchführbar.

Beweis. 1. Die Bedingung $M(s) \geq 0$ besagt, daß für jede geschlossene m-Kette C_m

$$(20) \qquad \sum_{C_m} c + 2ms \geq 0.$$

Diese Bedingung ist aber für alle m richtig, soferne sie es für $m \leq q$ ist. Für $m > q$ kann man nämlich schreiben

$$\sum_{C_m} c + 2ms = \sum_{\varrho} \left(\sum_{C_\varrho} + 2 m_\varrho s \right),$$

worin die C_ϱ endlich viele einfache (d. h. nicht-zerlegbare) geschlossene m_ϱ-Ketten mit $m_\varrho \leq q$ und $\sum_{\varrho} m_\varrho = m$ bezeichnen. Es genügt daher (20) lediglich für alle einfach-geschlossenen m-Ketten zu postulieren. Wir bilden demgemäß

$$M_1(s) := \min \left(\sum_{C_m} c + 2ms \right),$$

worin C_m alle einfach-geschlossenen m-Ketten (mit $m \leq q$) durchläuft. $M_1(s)$ ist eine stetige, steigende, stückweise lineare Funktion, welche genau einmal, etwa für $s = s_1$, verschwindet:

$$M_1(s_1) = 0 \text{ und } M_1(s) \geq 0 \text{ für } s \geq s_1.$$

Wegen $M_1(s) \geq M(s)$ für alle s und $M(s) \geq 0$ für $s \geq s_1$ ist

$$s_1 = s^*. \tag{21}$$

s_1 läßt sich aber in endlich vielen Schritten berechnen. Als Nebenergebnis dieser Überlegungen vermerken wir

$$M(s^*) = 0. \tag{22}$$

2. Ist s mit $M(s) \geq 0$ einmal gewählt, so genügt es, in (11) nur einfache (d. h. keine geschlossenen Ketten enthaltenden) $K_m(k, k')$ zuzulassen, da wegen (20) die nicht-einfachen Ketten K_m für die Bildung des Infimums in (11) unwirksam sind. Wir dürfen daher

$$d_{kk'} = \min \left(\sum_{K_m} c + 2ms \right)$$

setzen, wo K_m alle einfachen m-Ketten von k nach k' durchläuft (womit von selbst $m \leq q$ ist). Somit sind auch die endlich vielen $d_{kk'}$ in endlich vielen Schritten berechenbar, und damit auch die allgemeine Lösung von (3).

11. Mit **10.** ist das Matrixproblem prinzipiell erledigt; das angegebene Verfahren hat endlichen Charakter und führt zur allgemeinen Lösung. Wenn es sich aber darum handelt, es für eine gegebene Matrix numerisch durchzuführen, so wird es bei größerer Spalten- und Zeilenzahl ziemlich unbequem. Hier springt nun ein infinitärer Prozeß ein, ein konvergenter iterativer Algorithmus, der zum Wert s^* und zu Lösungen $(x; y)$ des Matrixproblems führt, allerdings nicht zu allen Lösungen, sondern nur zu gewissen ausgezeichneten Lösungen. Zur Beschreibung des Verfahrens führen wir folgende Definitionen ein:

Eine Matrix $A = (a_{ik})$ heißt *v-symmetrisch* (variations-symmetrisch), wenn

$$\min_k a_{ik} = -\max_k a_{ik} \qquad \text{für } i = 1, \ldots, p \tag{23}$$

und

$$\min_i a_{ik} = -\max_i a_{ik} \qquad \text{für } k = 1, \ldots, q. \tag{24}$$

Die Matrizen (a_{ik}) und (b_{ik}) heißen *(zueinander) verwandt*, wenn es Zahlen $x_1, \ldots, x_p, y_1, \ldots, y_q$ gibt, so daß

$$a_{ik} = b_{ik} + x_i + y_k$$

für $i = 1, \ldots, p$ und $k = 1, \ldots, q$.

Approximiert man die Matrix (a_{ik}) durch eine Matrix $(x_i + y_k)$, so ist die zugehörige Fehlermatrix $(a_{ik} - (x_i + y_k))$ zu (a_{ik}) verwandt; unser Matrixproblem besteht also darin, bei gegebener Matrix (a_{ik}) unter allen dazu verwandten Matrizen $B = (b_{ik})$ eine mit kleinster *Norm* $|| B || = \max_{i,k} | b_{ik} |$ zu finden.

12. Hilfssatz. *Ist $A = (a_{ik})$ eine v-symmetrische Matrix und $B = (b_{ik})$ eine dazu verwandte, so gilt $|| B || \geq || A ||$.*

Beweis. Im Falle $A = 0$ ist die Behauptung trivial; es sei daher $a := \max_{i,k} a_{ik} = a_{i'k'} > 0$. Wegen der v-Symmetrie von A tritt in der Zeile i' auch der Wert $-a = a_{i'k''}$ auf, weiter dann in der Spalte k'' der Wert $a = a_{i''k''}$, usw. Wandert man auf

diese Weise in der Matrix herum, so ergibt sich ein geschlossener Weg

$$i_1 k_1,\ i_1 k_2,\ i_2 k_2,\ \ldots,\ i_n k_n,\ i_n k_1,$$

wozu abwechselnd die Werte a, $-a$ gehören. Setzen wir allgemein $b_{ik} = a_{ik} + x_i + y_k$, so haben wir die Gleichungen

$$b_{i_1 k_1} = a + x_{i_1} + y_{k_1},$$

$$b_{i_1 k_2} = -a + x_{i_1} + y_{k_2},$$

$$\ldots\ldots\ldots$$

$$b_{i_n k_1} = -a + x_{i_n} + y_{k_1}.$$

Bildet man hier die alternierende Summe, so wird

$$\sum \pm b_{ik} = 2na,$$

woraus folgt, daß mindestens eines dieser b_{ik} von einem Betrag $\geq a$ ist.

13. Aus dem vorausgehenden Hilfssatz folgt, daß jede zu (a_{ik}) verwandte v-symmetrische Matrix (b_{ik}) die Fehlermatrix einer Lösung $(x_i + y_k)$ des Matrixproblems und damit eine Lösung selbst ergibt. Wir werden daher versuchen, zu (a_{ik}) eine v-symmetrische Verwandte zu bestimmen.

Die Gleichungen für $x_1, \ldots, x_p, y_1, \ldots, y_q$, die $(a_{ik} + x_i + y_k)$ als v-symmetrische Matrix kennzeichnen, sind

$$(25)\quad y_k = -g_k(x),\ x_i = -h_i(y),\ i = 1, \ldots, p;\ k = 1, \ldots, q,$$

wobei gesetzt ist

(26)

$$g_k(x) = g_k(x_1, \ldots, x_p) = \tfrac{1}{2}\,(\max_i\,(a_{ik} + x_i) + \min_i\,(a_{ik} + x_i)),$$

(27)

$$h_i(y) = h_i(y_1, \ldots, y_q) = \tfrac{1}{2}\,(\max_k\,(a_{ik} + y_k) + \min_k\,(a_{ik} + y_k)).$$

Eine direkte Behandlung der Gleichungen (25) dürfte mit einigen Schwierigkeiten verbunden sein; ihre besondere Form aber

legt eine *Auflösung durch Iteration* nahe. Wir führen zu diesem Zweck die *alternierende Symmetrisierung einer Matrix* ein:

Unter der *Zeilensymmetrisierung* verstehen wir den Übergang von der Matrix $A = (a_{ik})$ zur Matrix $A' = (a'_{ik})$ gemäß

$$a'_{ik} = a_{ik} - z_i \text{ mit } z_i = \frac{1}{2}(\max_h a_{ih} + \min_h a_{ih}),$$

wofür wir auch kurz $A' = Z(A)$ schreiben;
die *Spaltensymmetrisierung* ist der Übergang von A zur Matrix $A'' = S(A)$ gemäß

$$a''_{ik} = a_{ik} - s_k \text{ mit } s_k = \frac{1}{2}(\max a_{jk} + \min_j a_{jk}).$$

Offensichtlich haben wir:

Ist A v-symmetrisch, so gilt $A = Z(A) = S(A)$, und umgekehrt.

Die alternierende Anwendung der Z- und S-Symmetrisierung ergibt die *Symmetrisierungsfolge zu A:*

$$A^{(0)} = A,\ A^{(2n+1)} = S(A^{(2n)}),\ A^{(2n+2)} = Z(A^{(2n+1)})$$

$n = 0, 1, \ldots$

Bemerkungen: 1. Daß wir die Symmetrisierungsfolge mit einer S-Symmetrisierung eröffnen, ist natürlich eine Willkür; tatsächlich kann bei anderer Eröffnung auch eine andere Folge entstehen; z. B. ergeben sich für

$$C_0 = \begin{pmatrix} 2 & -2 & 1 & 1 \\ -2 & 2 & 1 & 1 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & -1 & 3 \end{pmatrix}$$

die Matrizen

$$S(C_0) = \begin{pmatrix} 2 & -2 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & -2 & 2 \end{pmatrix},\ Z(C_0) = \begin{pmatrix} 2 & -2 & 1 & 1 \\ -2 & 2 & 1 & 1 \\ -1 & -1 & 2 & -2 \\ -1 & -1 & -2 & 2 \end{pmatrix},$$

welche v-symmetrisch sind und daher bei weiteren Symmetrisierungen ungeändert bleiben. Dies Beispiel lehrt auch, daß es zu einer Matrix mehrere v-symmetrische Verwandte geben kann.

2. Gemäß unserer Definition von $A^{(n)}$ gilt

$$(A^{(n)})^{(m)} = \begin{cases} A^{(n+m)} & \text{für } n \text{ gerade} \\ A^{(n+m-1)} & \text{für } n \text{ ungerade,} \end{cases}$$

weil allgemein $S(S(A)) = S(A)$ ist.

14. **Satz.** 1. *Die Symmetrisierungsfolge ist für jede Matrix A konvergent gegen eine zu A verwandte v-symmetrische Matrix A^*.*

2. *$Z = A - A^*$ ist eine beste T-Approximation von A durch eine Matrix der Form $(x_i + y_k)$.* 3. *Bei jeder solchen besten Approximation Z' von A ist der Fehler $|| A - Z' || = || A^* ||$.*

Beweis. a) Da bei einer Zeilen- (oder Spalten-)Symmetrisierung das Maximum des absoluten Betrages der Zeilen- (bzw. Spalten-)Glieder nicht vergrößert wird, so folgt

$$0 \leq || A^{(n+1)} || \leq || A^{(n)} ||,$$

und wir können setzen

$$\lim_n || A^{(n)} || = \alpha$$

mit einem $\alpha \geq 0$.

Ist $\alpha = 0$, so ist $\lim_n A^{(n)} = 0$ und wir sind fertig.

Es sei daher fortan $\alpha > 0$.

b) Wegen der Beschränktheit der $A^{(n)}$ gibt es eine konvergente Teilfolge $((A^{(n_\nu)}))$ mit $\lim_\nu A^{(n_\nu)} = B$. Offensichtlich ist $|| B || = \alpha$. Wegen $\lim_\nu (A^{(n_\nu)})^{(k)} = B^{(k)}$ gilt auch $|| B^{(k)} || = \alpha$. Liegt auf einer Stelle von B ein Wert vom Betrag $< \alpha$, so bleibt der Betrag des Wertes an dieser Stelle $< \alpha$ für alle $B^{(k)}$. Durch wiederholte Z-S-Symmetrisierungen an B kann also ein Wert $\pm \alpha$ verlorengehen, jedenfalls nicht geschaffen werden. Es gibt daher ein k_0, so daß alle $B^{(k)}$ für $k \geq k_0$ dieselben $(+\alpha)$- und

$(-\alpha)$-Stellen haben. Ohne Beschränkung dürfen wir $k_0 = 0$ setzen; sonst würden wir die $(A^{(n_\nu)})^{(k_0)}$ anstelle der $A^{(n_\nu)}$ weiter betrachten. Die Erhaltung der $(\pm\alpha)$-Stellen in B ist nur in der Weise möglich, daß in jeder Zeile (Spalte) von B, welche den Wert $(\pm)\,\alpha$ enthält, auch der Wert $(\mp)\,\alpha$ steht. Eine solche Zeile (Spalte) nenne ich kurz *α-Zeile (-Spalte)*. Es gibt in B mindestens 2 α-Zeilen und 2 α-Spalten. Zur übersichtlicheren Anordnung nehmen wir eine Umnumerierung der Zeilen und Spalten vor, nämlich so, daß die ersten p_1 Zeilen und die ersten q_1 Spalten gerade die α-Zeilen und -Spalten von B sind. Dann ist $|b_{ij}| \leq \alpha' < \alpha$ für $i > p_1$ und $j > q_1$.

c) Wir betrachten nun alle möglichen konvergenten Teilfolgen $((A^{(n'_\nu)}))$ der Folge $((A^{(n)}))$, welche in der Teilmatrix der ersten p_1 Zeilen und q_1 Spalten gegen die Werte von B streben, in den übrigen Zeilen und Spalten aber gegen Werte von einem Betrag $\leq \alpha'' \leq \alpha'$. Ein einfaches Auswahlverfahren führt zu einer solchen Folge $((A^{(n'_\nu)}))$, für die das betreffende α'' minimal ist, etwa gleich α_1. Ohne Beschränkung der Allgemeinheit können wir voraussetzen, daß all dies bereits für die Folge $((A^{(n_\nu)}))$ selbst zutrifft. Ist $\alpha_1 = 0$, so ist B v-symmetrisch. Wir werden zeigen, daß wir auch im Falle $\alpha_1 > 0$ B als v-symmetrisch wählen dürfen.

d) Bei der alternierenden Symmetrisierung von B bleibt ein Wert in den Restzeilen und -spalten von einem Betrag $< \alpha_1$ weiterhin ein solcher, während ein Wert vom Betrag α_1 unter Umständen in einen vom Betrag $< \alpha_1$ verwandelt werden kann. Gäbe es ein m, so daß $B^{(m)}$ in den Restzeilen und -spalten nur Werte vom Betrag $< \alpha_1$ hätte, so würde die Folge der $(A^{(n_\nu)})^{(m)}$ gegen die Matrix $B^{(m)}$ konvergieren, welche natürlich in der Teilmatrix der ersten p_1-Zeilen und q_1-Spalten mit B übereinstimmt, was mit der Minimaleigenschaft von α_1 in Widerspruch steht. Daher gibt es ein m_1, so daß in $B^{(m_1)}$, $B^{(m_1+1)}$, ... die $(\pm\alpha_1)$-Stellen fest sind, was in einer restlichen Zeile (Spalte) nur in der Weise möglich ist, daß darin die Werte α_1 und $-\alpha_1$ auftreten („*α_1-Zeile* [*-Spalte*]"). Wir ergänzen die Umnumerierung der Zeilen und Spalten in der Weise, daß die α_1-Zeilen die Nummern $p_1 + 1, \ldots, p_2$, die α_1-Spalten die Nummern $q_1 + 1, \ldots, q_2$ er-

halten, während in Zeilen und Spalten von $B^{(m_1)}$ mit größerer Nummer nur Werte vom Betrag $\leq \alpha''' < \alpha_1$ auftreten. Ohne Beschränkung dürfen wir wieder $m_1 = 0$ voraussetzen.

e) Fortsetzung des in c) und d) beschriebenen Auswahlverfahrens führt schließlich zu einer gegen eine Matrix B konvergenten Teilfolge $((A^{(n_\nu)}))$, zu Werten

$$\alpha = \alpha_0 > \alpha_1 > \ldots > \alpha_s \geq 0$$

und zu Indexnummern

$$0 = p_0 < p_1 \leq p_2 \leq \ldots \leq p_s = p,$$

$$0 = q_0 < q_1 \leq q_2 \leq \ldots \leq q_s = q$$

derart, daß in jeder Zeile i mit $i > p_\sigma$ wie auch in jeder Spalte j mit $j > q_\sigma$ der Matrix B nur Werte von einem Betrag $\leq \alpha_\sigma$ stehen, aber in jeder Zeile i mit $p_\sigma < i \leq p_{\sigma+1}$, wie in jeder Spalte j mit $q_\sigma < j \leq q_{\sigma+1}$ sowohl der Wert α_σ als auch $-\alpha_\sigma$ auftreten, $\sigma = 0, \ldots, s$. Es ist klar, daß bei diesen Eigenschaften die Matrix B *v-symmetrisch* ist. Wir haben daher neben $A^{(n_\nu)} \to B$ für $\nu \to +\infty$ auch $Z(A^{(n_\nu)}) \to Z(B) = B$ und $S(A^{(n_\nu)}) \to S(B) = B$, also auch $A^{(n_\nu+1)} \to B$ für $\nu \to +\infty$. Wir können daraus schließen, daß

$$(28) \qquad \| A^{(n_\nu+1)} - A^{(n_\nu)} \| \to 0 \qquad \text{für } \nu \to +\infty$$

ist, ferner dürfen wir voraussetzen, daß *alle* n_ν *gerade* sind.

f) Nun gilt aber allgemein

$$(29) \qquad \| A^{(n+1)} - A^{(n)} \| \leq \| A^{(n)} - A^{(n-1)} \|.$$

In der Tat, betrachten wir etwa den Fall, daß eine Matrix A aus einer anderen durch Zeilensymmetrisierung hervorgegangen ist, daß auf A eine Spaltensymmetrisierung angewandt wird mit der maximalen Abänderung ε (≥ 0) der Elemente und auf das Ergebnis A' noch eine Zeilensymmetrisierung mit der maximalen Abänderung ε' (≥ 0). Der kleinste bzw. größte Wert in einer gewissen Zeile von A' sei s' bzw. S'; bei einer Symmetrisierung

dieser Zeile werden die Elemente um den Wert $\frac{s' + S'}{2}$ geändert. Ist s bzw. S der kleinste bzw. größte Wert dieser Zeile in A, so gilt $s = -S$, ferner

$$S - \mathfrak{e} \leq S' \leq S + \mathfrak{e}$$
$$s - \mathfrak{e} \leq s' \leq s + \mathfrak{e}, \text{ also}$$
$$-\mathfrak{e} \leq \frac{s' + S'}{2} \leq \mathfrak{e}, \text{ somit } \mathfrak{e}' \leq \mathfrak{e}, \text{ w. z. z. w.}$$

Aus (28) und (29) folgt

$$(30) \qquad \| A^{(n+1)} - A^{(n)} \| \to 0 \text{ für } n \to +\infty.$$

g) Um nun zu unserer Behauptung $A^{(n)} \to B$ zu kommen, betrachten wir den Symmetrisierungsprozeß in Hinblick auf die Transformationen (26) und (27):

$$(31) \qquad x \to y \; : \; y_k = -g_k(x),$$

$$(32) \qquad y \to x \; : \; x_i = -h_i(y).$$

Es entspricht (31) der S-Operation, (32) der Z-Operation. Führt man (31) und (32) hintereinander aus, so erhält man die Abbildung $x \to x'$ gemäß

$$(33) \quad x_i' = t_i(x) = -h_i(-g_1(x), \ldots, -g_q(x)), \; i = 1, \ldots, p.$$

Von der Transformation t läßt sich folgendes sagen:

(34) Ist $\bar{u} = (u, u, \ldots, u)$, so gilt

$$t_i(x + \bar{u}) = t_i(x) + u;$$

denn Entsprechendes gilt auch für die Funktionen g_k und h_i.

Ausführlich bestimmt sich t_i folgendermaßen:

Sind i_1, i_2 (von k abhängige) Indizes, so daß

$$(35) \qquad a_{i_1 k} + x_{i_1} \leq a_{ik} + x_i \leq a_{i_2 k} + x_{i_2}$$

für $i = 1, \ldots, p$, und $k = 1, \ldots, q$, so ist

$$g_k(x) = \tfrac{1}{2}(a_{i_1 k} + a_{i_2 k}) + \tfrac{1}{2}(x_{i_1} + x_{i_2}).$$

Ist weiter mit (von i abhängigen) Indizes k_1 und k_2

$$a_{ik_1} - g_{k_1}(x) \leq a_{ik} - g_k(x) \leq a_{ik_2} - g_{k_2}(x) \tag{36}$$

für $i = 1, \ldots, p$ und $k = 1, \ldots, q$, so ist

$$t_i(x) = -\tfrac{1}{2}(a_{ik_1} + a_{ik_2}) + \tfrac{1}{2}(g_{k_1}(x) + g_{k_2}(x)).$$

Bei festen Indexfunktionen i_1, i_2, k_1, k_2 wird durch (35) und (36) ein abgeschlossener Teilbereich des Raumes R^p der x bestimmt; ich nenne einen solchen Teilbereich eine „*Röhre*", weil mit x, auch jeder Punkt $x + \tilde{u}$ in einem solchen Bereich enthalten ist. Da es nur endlich viele Röhren gibt, so gehört ein jeder Punkt x nur endlich vielen Röhren an und im übrigen mindestens einer. Auf einer Röhre T hat t die Darstellung

$$t_i(x) = c_{iT} + \frac{1}{4}(x_{j_1} + x_{j_2} + x_{j_3} + x_{j_4}), \tag{38}$$

wobei die j_ν nur von i und T abhängen.

h) Beginnt man das Iterationsverfahren

$$x_i^{(n+1)} = t_i(x^{(n)}),\ n = 0, 1, \ldots$$

mit $x^{(0)} = 0$, so hat man

$$S(A) = (a_{ik} - g_k(0)),\ ZS(A) = (a_{ik} - g_k(0) + t_i(0)),$$

allgemein

$$A^{(2n)} = (a_{ik} - g_k(x^{(n-1)}) + x_i^{(n)}).$$

Aus (e) folgt demgemäß, daß für eine passende Teilfolge $((n'))$ von natürlichen Zahlen

$$-g_k(x^{(n'-1)}) + x_i^{(n')} \to \beta_{ik} \text{ für } n' \to +\infty$$

gilt, wobei $\beta_{ik} = \alpha_i + \beta_k$ und $B = (a_{ik} + \alpha_i + \beta_k)$ v-symmetrisch ist. Dies hat zur Folge

$$\alpha_i = t_i(\alpha). \tag{39}$$

Der Punkt α im R^p liegt entweder im Innern J einer Röhre T_0 oder er gehört nur endlich vielen Röhren $T_1, \ldots, T_m$ an und liegt im Innern J der Vereinigung dieser Röhren (wegen der Abgeschlossenheit der Röhren und ihrer endlichen Anzahl). Für α gilt somit (38) mit $T = T_0$ bzw. mit $T = T_1, \ldots, T_m$. Setzen wir daher $x_i = \alpha_i + \xi_i$, $x_i' = \alpha_i + \xi_i'$, so folgt mit (39)

$$\xi_i' = \tfrac{1}{4}(\xi_{j_1} + \xi_{j_2} + \xi_{j_3} + \xi_{j_4}) \text{ für } x \in J. \tag{40}$$

Wir wählen nun ein $r_0 > 0$, so daß die Umgebung

$$\{x : |x_i - \alpha_i| < r_0\}$$

in J enthalten ist; wenn dann $|\xi_i| < \varepsilon < r_0$, so ist nach (40) auch $|\xi_i'| < \varepsilon$. Wegen $x_i^{(n')} + (\alpha_{i_0} - x_{i_0}^{(n')}) \to \alpha_i$ (i_0 beliebig aber fest), so gibt es zu jedem ε mit $0 < \varepsilon < r_0$ ein n_1', so daß mit $u := \alpha_{i_0} - x_{i_0}^{(n_1')}$ gilt:

$$|x_i^{(n_1')} + u - \alpha_i| < \varepsilon \text{ für } i = 1, \ldots, p.$$

Setzen wir allgemein $x_i^{(n)} + u = \alpha_i + \xi_i^{(n)}$, so ergibt sich $x_i^{(n_1'+1)} + u = t_i(x^{(n_1')} + \vec{u}) = \alpha_i + \xi_i^{(n_1'+1)}$, also nach dem Vorausgehenden $|\xi_i^{(n_1'+1)}| < \varepsilon$, und in Fortsetzung dieses Schlusses allgemein $|\xi_i^{(n)}| < \varepsilon$ für alle i und alle $n \geq n_1'$. Dies besagt, daß $\xi^{(n)} \to 0$, also $x_i^{(n)} \to \alpha_i - u$ und wegen der Stetigkeit von g_k somit schließlich $A^{(2n)} \to B$, also allgemein $A^{(n)} \to B$, w. z. z. w.

Die 2. und 3. betreffenden Behauptungen ergeben sich aus dem Hilfssatz in 12.

15. Nach dem Bewiesenen können wir hinsichtlich der Struktur der Gesamtheit $\mathfrak{S}$ der zu einer Matrix $A = (a_{ik})$ gehörigen v-symmetrischen Verwandten und der Beziehungen zwischen $\mathfrak{S}$, der Gesamtheit $\mathfrak{M}$ der Lösungen des Matrixproblems im Sinne der Minimalisierung von $\max_{i,k} |a_{ik} + x_i + y_k|$ und der Gesamtheit $\mathfrak{V}$ aller Verwandten von A folgendes sagen:

1) Es ist $\mathfrak{S} \subset \mathfrak{M} \subset \mathfrak{V}$;

2) Die Anwendung der alternierenden Symmetrisierung auf $\mathfrak{V}$ ergibt

$$\mathfrak{V}^* := \{V^* : V \in \mathfrak{V}\} = \mathfrak{S}.$$

Dem sei noch hinzugefügt:

3) Es gibt Beispiele, wo $\mathfrak{S} = \mathfrak{M}$;

4) Es gibt Beispiele, wo die konvexe Hülle von $\mathfrak{S}$ ein echter Teil von $\mathfrak{M}$ ist.

Hier sind also noch einige Fragen zu klären; darüber soll später berichtet werden.