6

Appendix A: *Angenäherte Auflösung von Systemen linearer Gleichungen*

6.1 Approximate Solution for Systems of Linear Equations (1) (English translation)

Given the importance of Kaczmarz's algorithm to this thesis, and unable to find an English translation of Kaczmarz's original article, I undertook the following translation from German.

Although the approximate solution of the equation $f(x) = 0$ with an unknown variable has been extensively treated in the literature, we nevertheless know little about the solutions of sets of equations, even when they are linear. There are only two methods [1,2] which permit solutions, but they are limited to equations whose diagonal coefficients, a_{ii}, are large in comparison with the other coefficients. For this reason, one finds in the work of Mises and Pollaczek-Geiringer the condition $a_{ii} >= (n-1)a_{ik}$, and so forth.

In general, however, one can completely solve the equations by means of iterative methods. We assume that a system is given,

$$a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n + b_i = 0, \qquad i = 1, 2, ...n \tag{6.1.1}$$

which possesses only one solution, $x_1, x_2, \dots x_n$. We presuppose that equation 6.1.1 by multiplication with appropriate constants is transformed in such a way that

$$\sum_{k=1}^{n} a_{ik}^2 = 1 \tag{6.1.2}$$

is valid for all $i = 1, 2, \dots n$.

Let there exist a first estimate, $x_1^0, x_2^0, \dots x_n^0$, which can be completely arbitrary. We divide successive estimates into groups, each consisting of n (approximate) solutions,

$$\begin{array}{l} x_1^{(p,1)}, x_2^{(p,1)}, \dots x_n^{(p,1)}, \\ . \quad . \quad . \quad . \\ x_1^{(p,n)}, x_2^{(p,n)}, \dots x_n^{(p,n)}, \end{array}$$

the n solutions of the p^{th} group. We define the $(p+1)^{th}$ group as follows

$$\begin{aligned} x_i^{(p+1,1)} &= x_i^{(p,n)} - a_{1i} L_1^{(p,n)} \\ x_i^{(p+1,2)} &= x_i^{(p+1,1)} - a_{2i} L_2^{(p+1,1)} \\ . \quad . \quad & . \quad . \quad . \quad . \\ x_i^{(p+1,n)} &= x_i^{(p+1,n-1)} - a_{ni} L_n^{(p+1,n-1)} \\ (i = 1, 2, \dots n) & \end{aligned}$$

In these equations, $L_i^{(r,s)} = \sum_{k=1}^{n} a_{ik} x_k^{(r,s)} + b_i$ defines the left side of the i^{th} equation in 6.1.1, wherein the s^{th} solution to the r^{th} group is used for the variables $x_1, x_2, \dots x_n$.

Consequently, we submit that the solution

$$x_1^{(r,s)}, x_2^{(r,s)}, \dots x_n^{(r,s)} \qquad s = 1, 2, \dots n, r = 1, 2, \dots \tag{6.1.3}$$

converges toward $x_1, x_2, \ldots x_n$. From 6.1.3 it follows that

$$x_i^{(r,s+1)} - x_i = x_i^{(r,s)} - x_i - a_{s+1,i} L_{s+1}^{(r,s)}, \qquad s = 0, 1, \ldots n-1, \tag{6.1.4}$$

(for $s = 0$, we have $x_i^{(r,1)} - x_i = x_i^{(r-1,n)} - x_i - a_{1i} L_i^{(r-1,n)}$). By squaring and summing from $i = 1$ to n one obtains [1]

$$\sum_{i=1}^{n} (x_i^{(r,s+1)} - x_i)^2 = \sum_{i=1}^{n} (x_i^{(r,s)})^2 + L_{s+1}^2 - 2L_{s+1} \sum_{i=1}^{n} a_{s+1,i} (x_i^{(r,s)} - x_i) \tag{6.1.5}$$

But because

$$\sum a_{s+1,i} (x_i^{(r,s)} - x_i) = L_{s+1} - b_{i+1} + b_{i+1} = L_{s+1}, \tag{6.1.6}$$

if one substitutes

$$y_{r,s} = \sum_{i=1}^{n} (x_i^{(r,s)} - x_i)^2 \tag{6.1.7}$$

for the sake of conciseness, one then obtains from 6.1.5,

$$y_{r,s+1} = y_{r,s} - L_{s+1}^2. \tag{6.1.8}$$

The sequence $y_{r,s}$ thus falls monotonically toward the boundary y. Since $y_{r,s}$ are limited, one can find a subsequence of $x_i^{(r,s)}$ which approaches the bound $y_1, y_2, \ldots y_n$. From 6.1.8, it follows that

$$\lim_{r \to \infty} L_{s+1}^{(r,s)} = L_{s+1}(y_1, y_2, \ldots y_n) = 0, \tag{6.1.9}$$

and also $\lim x_i^{(r,s)} = y_i = x_i$, the system in 6.1.1 possesses only one solution. Therefore $y_{r,s}$ must converge toward zero, which means that $x_i^{(r,s)}$ tends toward x_i.

When the system 6.1.1 has more solutions, then naturally $y_1, y_2, \ldots y_n$ is one of the possible solution.

[1] In 6.1.5 the double subscript for L is omitted.

For the practical execution of calculations it is important to note that the condition $\sum_1^n a_{ik}^2 = 1$ does not have to be strictly fulfilled. If one sets $\sum_1^n a_{ik}^2 = c_i$, $c_i \neq 1$, then relation 6.1.5 turns into

$$y_{r,s+1} = y_{r,s} + c_{s+1} L_{s+1}^2 - 2L_{s+1}^2. \tag{6.1.10}$$

We therefore see that $y_{r,s+1} < y_{r,s}$ (and thus converges) if $c_{s+1} < 2$. One can choose coefficients c_i such that $c_i < 2$ and the multiplication becomes very easy.

Geometrically the algorithm means the following:

The point $x_1^0, \dots x_n^0$ is projected orthogonally onto the first hyperplane $L_1 = 0$; this projection is just the point $x_1^{(1,1)}, \dots x_n^{(1,1)}$, which is now cast onto $L_2 = 0$, etc. The point $x_1^{(1,n)}, \dots x_n^{(1,n)}$ is again cast onto $L_1 = 0$, giving the point $x_1^{(2,1)}, \dots x_n^{(2,1)}$, etc. In this way, the convergence of the algorithm is easily plausible.

Reference:

[1] R. Mises, H. Pollaczek-Geiringer, Praktische Verfahren der Gleichungsauflösung, Zeit. ang. Math. u. Mech. 9 (1929), S. 58-77, 152-164.

[2] J. Morris, A successive approximative process for solving simultaneous linear equations, Aeronaut. Res. Com. Rep. Number 1711 (1936) p. 1-12.

Przybliżone rozwiązywanie układów równań liniowych. — Angenäherte Auflösung von Systemen linearer Gleichungen.

Note

de M. ***S. KACZMARZ,***

présentée le 14 Juin 1937 par M. Th. Banachiewicz m. t.

Obgleich die angenäherte Auflösung der Gleichung $f(x)=0$ mit einer Unbekannten zahlreiche Bearbeitungen in der Literatur aufweist, wissen wir dennoch sehr wenig über die Auflösung von Gleichungssystemen, sogar wenn sie linear sind. Es gibt zwar einige Methoden[1]), welche die Auflösung gestatten, doch sind sie auf Gleichungen beschränkt, für welche die Diagonalkoeffizienten a_{ii} überwiegend groß sind im Vergleich mit den übrigen Koeffizienten; so findet man z. B. bei Mises und Pollaczek-Geiringer die Bedingung $a_{ii} \geqslant (n-1)\, a_{ik}$ und desgleichen.

Man kann jedoch die Aufgabe mittels eines Iterationsverfahrens ganz allgemein lösen. Nehmen wir an, daß ein System

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n + b_i = 0, \qquad (i=1,2,\ldots n) \quad (1)$$

gegeben ist, welches eine einzige Lösung $x_1, x_2, \ldots x_n$ besitzt. Wir setzen voraus, daß die Gleichungen (1) durch Multiplikation mit entsprechenden Konstanten so umgeformt wurden, daß

$$\sum_{k=1}^{n} a_{ik}^2 = 1$$

für alle $i = 1, 2 \ldots n$ gilt.

[1]) Vgl. R. Mises und H. Pollaczek-Geiringer, Praktische Verfahren der Gleichungsauflösung, Zeit. ang. Math. u. Mech. 9 (1929), S. 58—77, 152—164.

J. Morris, A successive approximative process for solving simultaneous linear equations, Aeronaut. Res. Com. Rep. Nr 1711 (1936) p. 1—12.

Es sei $x_1^0, x_2^0, \ldots x_n^0$ eine erste Annäherung, die ganz beliebig sein kann. Die folgenden Annäherungen teilen wir in Gruppen, deren jede aus n (angenäherten) Lösungen besteht. Die Lösung $x_1^0, x_2^0, \ldots x_n^0$ bildet die nullte Gruppe. Es seien

$$\begin{array}{l} x_1^{(p,1)},\ x_2^{(p,1)},\ \ldots\ x_n^{(p,1)}, \\ \ldots\quad\ldots\ldots \\ x_1^{(p,n)},\ x^{(p,n)}\ \ldots\ x_n^{(p,n)}, \end{array}$$

die n Lösungen der p-ten Gruppe; wir definieren die $(p+1)$-te Gruppe folgendermaßen

$$\begin{aligned} x_i^{(p+1,1)} &= x_i^{(p,n)} && - a_{1i} L_1^{(p,n)}, \\ x_i^{(p+1,2)} &= x_i^{(p+1,1)} && - a_{2i} L_2^{(p+1,1)} \qquad (i=1,2\ldots n) \quad (2) \\ &\ldots\ldots\ldots \\ x_i^{(p+1,n)} &= x_i^{(p+1,n-1)} && - a_{ni} L_n^{(p+1,n-1)}. \end{aligned}$$

In diesen Gleichungen bezeichnet $L_i^{(r,s)} = \sum_{k=1}^{n} a_{ik} x_k^{(r,s)} + b_i$ die linke Seite der i-ten Gleichung (1), wo für die Unbekannten $x_1, x_2, \ldots x_n$ die s-te Lösung der r-ten Gruppe eingesetzt wurde.

Wir behaupten, daß die Folge der Lösungen

$$x_1^{(r,s)}, x_2^{(r,s}, \ \ldots\ x_n^{(r,s)} \qquad s=1,2,\ldots n,\ r=1,2,\ldots$$

gegen $x_1, x_2 \ldots x_n$ konvergiert. Aus (2) folgt nämlich

$$x_i^{(r,s+1)} - x_i = x_i^{(r,s)} - x_i - a_{s+1,i} L_{s+1}^{(r,s)}, \qquad s=0,1,\ldots n-1,$$

(für $s=0$, ist $x_i^{(r,1)} - x_i = x_i^{(r-1,n)} - x_i - a_{1i} L_i^{(r-1,n)}$). Durch Quadrieren und Summieren von $i=1$ bis n erhält man [1])

$$\sum_{i=1}^{n} (x_i^{(r,s+1)} - x_i)^2 = \sum_{i=1}^{n} (x_i^{(r,s)} - x_i)^2 + L_{s+1}^2 - 2L_{s+1} \sum_{i=1}^{n} a_{s+1,i} (x_i^{(r,s)} - x_i). \quad (3)$$

Da aber

$$\sum a_{s+1,i} (x_i^{(r,s)} - x_i) = L_{s+1} - b_{i+1} + b_{i+1} = L_{s+1},$$

so bekommt man, wenn

$$y_{r,s} = \sum_{i=1}^{n} (x_i^{(r,s)} - x_i)^2$$

[1]) In (3) ist der obere Doppelindex bei L weggelassen.

der Kürze halber gesetzt wird, aus (3) die Gleichung

$$y_{r,s+1} = y_{r,s} - L_{s+1}^2. \tag{4}$$

Die Folge $\{y_{r,s}\}$ ist also monoton fallend und besitzt somit eine Grenze y; da die $y_{r,s}$ beschränkt sind, kann man eine Teilfolge in der Folge $\{x_i^{(r,s)}\}$ finden, welche gegen die Grenze $y_1, y_2, \dots y_n$ strebt. Aus der Gleichung (4) folgt, daß

$$\lim_{r \to \infty} L_{s+1}^{(r,s)} = L_{s+1}(y_1, y_2, \dots y_n) = 0,$$

also auch $\lim x_i^{(r,s)} = y_i = x_i$, da das System (1) nur eine einzige Lösung besitzt. Es muß also $y_{r,s}$ gegen Null konvergieren, was besagt, daß $x_i^{(r,s)}$ gegen x_i streben.

Wenn das System (1) mehrere Lösungen besitzt, ist natürlich $y_1, y_2, \dots y_n$ eine der möglichen Lösungen.

Für die praktische Ausführung der Rechnungen ist es wichtig, daß die Bedingung $\sum_1^n a_{ik}^2 = 1$ nicht streng erfüllt werden muß. Wenn man nämlich $\sum_1^n a_{ik}^2 = c_i$ setzt, $c_i \neq 1$, so geht die Relation (3) in

$$y_{r,s+1} = y_{r,s} + c_{s+1} L_{s+1}^2 - 2 L_{s+1}^2$$

über. Wir sehen also, daß $y_{r,s+1} < y_{r,s}$ (und somit konvergiert), wenn $c_{s+1} < 2$. Man kann also die c_i so wählen, daß $c_i < 2$ und die Multiplikation sehr einfach werde.

Geometrisch bedeutet das Verfahren folgendes:

Der Punkt $x_1^0, \dots x_n^0$ wird orthogonal auf die erste Hyperebene $L_1 = 0$ projiziert; die Projektion ist eben der Punkt $x_1^{(1,1)}, \dots x_n^{(1,1)}$, welcher nun auf $L_2 = 0$ geworfen wird usw. Der Punkt $x_1^{(1,n)}, \dots x_n^{(1,n)}$ wird wiederum auf $L_1 = 0$ geworfen und gibt den Punkt $x_1^{(2,1)}, \dots x_n^{(2,1)}$, usw. Die Konvergenz des Verfahrens ist geometrisch ohne weiteres einleuchtend.