COMPUTATIONAL MATHEMATICS
BANACH CENTER PUBLICATIONS, VOLUME 13
PWN–POLISH SCIENTIFIC PUBLISHERS
WARSAW 1984

A SURVEY OF GENERALIZED MATRIX INVERSES

GERHARD ZIELKE

Sektion Mathematik, Martin-Luther-Universität, Halle-Wittenberg, Weinbergweg 17, DDR-402 Halle/Saale, DDR

0. Introduction

Solve:

$$Ax = b,$$

$A \in C^{m\times n}$ a given complex $m\times n$ matrix, $b \in C^m$ a given vector, $x \in C^n$ solution vector.

Familiar cases:

A square and nonsingular $\square$:

$$x = A^{-1}b \text{ unique solution.}$$

A rectangular and column regular $\boxed{}$ (in general, an overdetermined system of equations) (linear independent columns)

$$x = (A^*A)^{-1}A^*b \text{ unique least-squares solution}$$

($A^* = \bar{A}^T$ the conjugate transpose or adjoint of A).

Method of least squares: $Ax-b = r \neq 0$ (residue)

$$\sum_{i=1}^{m} |r_i|^2 = r^*r = \|Ax-b\|^2 \to \min_{x\in C^n}!$$

($\|\cdot\|$ means the Euclidean vector norm).

Investigate the cases:

A rectangular and row regular $\square$ (underdetermined system) (linear independent rows)

$$\left.\begin{array}{l} A \text{ square and singular} \\ A \text{ rectangular and rank deficient} \end{array}\right\} \operatorname{rank}(A) = r(A) < \min(m, n).$$

QUESTION: Are analogous representations of the solutions $x = Mb$ or $x = Mb+Py$, y arbitrary, possible? M, P?

ANSWER: M generalized inverse, P projector (representable by a generalized inverse).

1. Penrose definition (derivation)

Wanted: "generalized inverse" A^+ which satisfies 3 *requirements*:

1. A^+ solves a consistent linear system of equations $Ax = b$ by $x = A^+b$,
2. A^+ yields an approximate solution of an inconsistent linear system of equations in the least squares sense by $x = A^+b$,
3. $(A^+)^+ = A$, analogous to $(A^{-1})^{-1} = A$.

1.1. $Ax = b$ *consistent* (*has a solution*), i.e., $b \in R(A) = \{y \in C^m \mid Ax = y,\ x \in C^n\}$ column space of A or range of A. $x = A^+b$ shall be solution of $Ax = b \Rightarrow AA^+b = b$ (condition of consistency). Since $b \in R(A)$, and hence $b = Az$, $z \in C^n$, it follows $AA^+Az = Az$, z arbitrary (depends only on b)

$\Rightarrow AA^+A = A$.

1.2. $Ax = b$ *inconsistent* (*has no solution*)
$x = A^+b$ shall be solution of $\|Ax-b\| = \min$, i.e.,

$$\|AA^+b-b\| \leqslant \|Ax-b\| \qquad \forall x \in C^n,\ b \in C^m.$$

By rewriting $Ax-b = AA^+b-b+A(x-A^+b)$ we obtain

$$\|AA^+b-b\| \leqslant \|AA^+b-b+Ay\| \qquad \forall b \text{ and } y = x-A^+b,$$

and by using the abbreviations $u = AA^+b-b$ and $v = Ay$ it follows

$$\|u\| \leqslant \|u+v\|$$

or because of $\|u\|^2 = u^*u$

$$u^*u \leqslant (u+v)^*(u+v) = u^*u+v^*v+u^*v+v^*u.$$

Because of $v^*u = \overline{u^*v}$ we obtain the inequality

$$0 \leqslant v^*v+2\,\mathrm{Re}(u^*v).$$

Since $v^*v = \|v\|^2 \geqslant 0$ and $u^*v = (AA^+b-b)^*Ay = b^*(A^{+*}A^*A-A)y$ depends on arbitrary vectors b and y, this inequality is equivalent to the requirement $u^*v = 0$ or

$$A^{+*}A^*A = A.$$

However, this requirement and the two equations

$$AA^+A = A,$$

$$(AA^+)^* = AA^+$$

are equivalent. For by substituting the second in the first equation we obtain $A^{+*}A^*A = A$. Conversely, from this follows by multiplying on the right by A^+ that

$$AA^+ = A^{+*}A^*AA^+ = (AA^+)^*AA^+$$

is Hermitian and hence

$$A = A^{+*}A^*A = (AA^+)^*A = AA^+A.$$

1.3. Equivalence to the Penrose equations.

THEOREM 1. *The requirements for* A^+ *and the four equations* (P1), ..., (P4) *are equivalent*:

$$\left.\begin{array}{lll} AA^+A &= A & \text{(F1)} \\ (AA^+)^* &= AA^+ & \text{(F2)} \\ (A^+)^+ &= A & \text{(F3)} \end{array}\right\} \Leftrightarrow \left\{\begin{array}{lll} AA^+A &= A & \text{(P1)} \\ A^+AA^+ &= A^+ & \text{(P2)} \\ (AA^+)^* &= AA^+ & \text{(P3)} \\ (A^+A)^* &= A^+A & \text{(P4)} \end{array}\right.$$

Proof. If we replace A by A^+ in (F1) and (F2) and consider (F3) we obtain the equations (P2) and (P4). Conversely, the equations (P1), ..., (P4) are symmetric in A and A^+, that means they define
A^+ as the generalized inverse of A and
A as the generalized inverse of A^+, but that means $A = (A^+)^+$.

DEFINITION 1 (Penrose [35]). The $n \times m$ matrix A^+ is called *Moore–Penrose inverse of the* $m \times n$ *matrix* A if A^+ satisfies the four conditions

$$\begin{array}{lll} AA^+A &= A & \text{(P1)} \\ A^+AA^+ &= A^+ & \text{(P2)} \\ (AA^+)^* &= AA^+ & \text{(P3)} \\ (A^+A)^* &= A^+A & \text{(P4)} \end{array}$$

(Penrose equations).

Remarks. 1. In the original definition of Penrose A^+ is called "generalized inverse".

2. Every matrix A has a unique matrix A^+ (Penrose [35]).

2. Other equivalent definitions of the Moore–Penrose inverse

DEFINITION 2. A^+ is called *Moore–Penrose inverse of* A if A^+b is the least-squares solution of $Ax = b$ with minimal Euclidean norm.

THEOREM 2. *The Definitions* 1 *and* 2 *are equivalent.*

Proof. According to 1.2 $x = Gb$ is a solution of the minimum problem $\|Ax - b\| = \min$ if and only if $AGA = A$ (P1) and $(AG)^* = AG$ (P3).

The *general* least-squares solution is

$$(*)\qquad x = Gb + (I - GA)z, \qquad z \text{ arbitrary},$$

(x is a least-squares solution because $A(I-GA)z = 0$; every least-squares solution $G_1 b$ can be represented in the form $(*)$ by choosing $z = G_1 b - Gb$; cf. 6.5). According to Definition 2 $G = A^+$ must satisfy the inequality

$$\|Gb\| \leqslant \|Gb + (I - GA)z\| \qquad \forall b, z.$$

By using the abbreviations $u = Gb$ and $v = (I-GA)z$ this inequality, analogous to Section 1.2, leads to the requirement

$$u^*v = b^*G^*(I-GA)z = b^*(G^* - G^*GA)z = 0, \qquad \forall b, z,$$

hence $G^* = G^*GA$. But it holds the equivalence

$$\begin{matrix} G^* = G^*GA \\ (\text{or } G = A^*G^*G) \end{matrix} \Leftrightarrow \begin{cases} GAG = G & (\text{P2}) \\ (GA)^* = GA & (\text{P4}), \end{cases}$$

for by substituting (P4) in (P2) we obtain $A^*G^*G = G$, and conversely, from this it follows by multiplying on the right by A that $GA = A^*G^*GA = (GA)^*GA$ is Hermitian and thus we have $G = A^*G^*G = (GA)^*G = GAG$. Thus $G = A^+$.

DEFINITION 3 (Moore [29]). A^+ is called *Moore–Penrose inverse of* A if

(1) $AA^+A = A$,
(2) the rows of A^+ are linear combinations of the rows of A^*,
(3) the columns of A^+ are linear combinations of the columns of A^*.

Remark. (2) means $A^+ = YA^*$ and (3) means $A^+ = A^*Z$ for suitable Y and Z.

DEFINITION 4 (Ben-Israel, Charnes [4]). A^+ is called *Moore–Penrose inverse of* A if

(1) $AA^+ = P_{R(A)}$, $P_{R(A)}$: orthogonal projector on $R(A)$,
(2) $A^+A = P_{R(A^+)}$, $P_{R(A^+)}$: orthogonal projector on $R(A^+)$.

Remark. $P_{R(A)}$ can be defined by the following properties:
$P_{R(A)}$ is idempotent, i.e., $P_{R(A)}^2 = P_{R(A)}$ (but unnecessary),
$P_{R(A)}^* = P_{R(A)}$ (Hermitian), $P_{R(A)}A = A$, $r(P_{R(A)}) = r(A)$.

3. Some properties of the Moore–Penrose inverse

(1) $(A^+)^+ = A$ (from Theorem 1);

(2) $(A^*)^+ = (A^+)^*$, short: $A^{*+} = A^{+*}$;

(3) $(AA^*)^+ = A^{+*}A^+$, $(A^*A)^+ = A^+A^{+*}$;

(4) $A^+ = (A^*A)^+A^* = A^*(AA^*)^+$
$= A^+A^{+*}A^* = A^*A^{+*}A^+$ (cf. Def. 3);
(P2) & (P3) (P2) & (P4)

(5) $A^* = A^*AA^+ = A^+AA^*$;
(P1) & (P3) (P1) & (P4)

(6) $(AA^+)^+ = AA^+$, $(A^+A)^+ = A^+A$;

(7) The matrices $AA^+ = P_{R(A)}$, $A^+A = P_{R(A^*)}$, $I - AA^+ = P_{N(A^*)}$, $I - A^+A = P_{N(A)}$ are Hermitian and idempotent (orthogonal projectors);

(8) A^*, AA^*, A^*A, A^+, AA^+ and A^+A have the same rank like A, moreover $r(A) = \text{trace}(AA^+) = \text{trace}(A^+A)$.

4. Moore–Penrose inverses of some special matrices

(1) A nonsingular: $A^+ = A^{-1}$ ordinary inverse;

(2) $A \in C^{m\times n}$ of full rank:
A column regular ▯, $r(A) = n$:

$$A^+ = A_L^{-1} = (A^*A)^{-1}A^* \text{ left inverse, } A^+A = I_n;$$

A row regular ▭, $r(A) = m$:

$$A^+ = A_R^{-1} = A^*(AA^*)^{-1} \text{ right inverse, } AA^+ = I_m;$$

(3) Zero matrix: $0^+ = 0^T$;

(4) Scalar a: $a^+ = \begin{cases} 1/a & \text{if } a \neq 0, \\ 0 & \text{if } a = 0; \end{cases}$

(5) Diagonal matrix:

$$D = \text{diag}(d_1, d_2, \ldots, d_n), \quad D^+ = \text{diag}(d_1^+, d_2^+, \ldots, d_n^+);$$

(6) $(c\cdot A)^+ = (1/c)A^+ (c \neq 0$ scalar), especially $(-A)^+ = -A^+$;

(7) A Hermitian ($A = A^*$), then A^+ Hermitian and $AA^+ = A^+A$;

(8) A Hermitian and idempotent ($A = A^*$, $A^2 = A$, i.e., A orthogonal projector), then $A^+ = A$;

(9) u column vector ($\neq 0$), then $u^+ = \dfrac{u^*}{u^*u} = \dfrac{u^*}{\|u\|^2}$,

v^* row vector ($\neq 0$), then $v^{*+} = \dfrac{v}{v^*v} = \dfrac{v}{\|v\|^2}$;

(10) $A = uv^*$ (i.e. $r(A) = 1$), then

$$A^+ = \frac{A^*}{(u^*u)(v^*v)} = \frac{A^*}{\|u\|^2\|v\|^2}$$

$$= \frac{A^*}{\operatorname{trace}(A^*A)}$$ (this is true for every matrix of rank 1);

(11) A $m \times n$ matrix with $a_{ij} = 1$ $\forall i, j$, then $A^+ = \frac{1}{m \cdot n} A^T$;

(12) $A = \begin{bmatrix} B \\ C \end{bmatrix}$ and $BC^* = 0$, then $A^+ = (B^+, C^+)$, especially: $\begin{bmatrix} B \\ 0 \end{bmatrix}^+ = (B^+, 0)$;

(13) $A = (B, C)$ with $B^*C = 0$, then $A^+ = \begin{bmatrix} B^+ \\ C^+ \end{bmatrix}$, especially: $(B, 0)^+ = \begin{bmatrix} B^+ \\ 0 \end{bmatrix}$;

(14) $A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$, then $A^+ = \begin{bmatrix} B^+ & 0 \\ 0 & C^+ \end{bmatrix}$;

(15) $A = BC$, B column regular, C row regular, then

$$A^+ = C^+B^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$$

(the reverse order law is not true in general!);

(16) $A = UBV^*$, where U, V have orthonormal columns (especially U, V unitary matrices), then

$$A^+ = VB^+U^*.$$

5. Classification of generalized inverses

In addition to the unique solution $X = A^+$ of the four Penrose equations

$$\begin{array}{llll} AXA = A & (\mathrm{P1}) & (AX)^* = AX & (\mathrm{P3}) \\ XAX = X & (\mathrm{P2}) & (XA)^* = XA & (\mathrm{P4}) \end{array}$$

also matrices X are of interest which satisfy only part of these equations, but at least (P1). In general, they are no longer determined uniquely. We obtain 8 classes of generalized inverses, e.g. the class of all $A^{1,4}$ or (1, 4)-inverses:

$$\{A^{1,4}\} = \{X \mid AXA = A,\ (XA)^* = XA\}.$$

For the (1)-inverse instead of A^1 we use the symbol A^- introduced by Rao [39].

Classification

Satisfied Penrose equations	Symbol	Terminology (Rao [40])
(P1)	A^-	generalized inverse
(P1), (P2)	$A^{1,2}$	reflexive inverse
(P1), (P3)	$A^{1,3}$	least-squares inverse
(P1), (P4)	$A^{1,4}$	minimum norm inverse
(P1), (P2), (P3)	$A^{1,2,3}$	reflexive least-squares inverse
(P1), (P2), (P4)	$A^{1,2,4}$	reflexive minimum norm inverse
(P1), (P3), (P4)	$A^{1,3,4}$	—
(P1), (P2), (P3), (P4)	A^+	Moore–Penrose inverse (pseudoinverse)

Inclusions: $\{A^+\} \subseteq \{A^{1,2,3}\} \subseteq \{A^{1,2}\} \subseteq \{A^-\}$ etc.
Intersection of classes, e.g. $\{A^+\} = \{A^{1,4}\} \cap \{A^{1,2,3}\}$.

6. Applications of generalized inverses to the solution of linear systems of equations (Zielke [53])

6.1. Condition of consistency (solvability of $Ax = b$).

3 equivalent conditions:

(1) $r(A) = r(A, b)$ $(r(A) = \dim R(A) = r)$,

(2) $b \in R(A)$,

(3) $AA^-b = b$ (see also 1.1).

Proof. $r(A) = r(A, b) \Rightarrow R(A) = R(A, b) \Rightarrow b \in R(A) \Rightarrow b = Ax \Rightarrow AA^-b = AA^-Ax = Ax = b \Rightarrow (A, b) = (A, AA^-b) \Rightarrow R(A, b) = R\big(A(I, A^-b)\big) \subseteq R(A)$, since $R(A) \subseteq R(A, b)$ it holds even the equality $R(A, b) = R(A) \Rightarrow r(A, b) = r(A)$.

6.2. Solution of the homogeneous system $Ax = 0$ (always solvable). For any matrix $A \in C^{m\times n}$ we denote by $N(A) = \{x \in C^n \mid Ax = 0\}$ the null space of A. It holds $N(A) = R(A^*)^\perp$, i.e. $N(A)$ is the orthogonal complement of $R(A^*)$ and conversely; for if $A^*y = z$ and $Ax = 0$ then $z^*x = y^*Ax = 0 \Leftrightarrow z \perp x$, i.e. all $z \in R(A^*)$ and $x \in N(A)$ are orthogonal to each other.

General solution of $Ax = 0$:

$$x_h \in N(A) = R(A^*)^\perp,$$

$$(*) \qquad x_h = B_{N(A)}y, \qquad y \in C^{n-r} \text{ arbitrary, } B_{N(A)} \text{ basis of } N(A).$$

For the representation of x_h by means of generalized inverses we consider $I_n - A^-A$. We show that $R(I_n - A^-A) = R(B_{N(A)}) = N(A)$. Firstly, $A(I_n - A^-A) = 0 \Rightarrow R(I_n - A^-A) \subseteq N(A)$. Secondly, from $Ax = 0 \Rightarrow A^-Ax = 0$, i.e. $x - A^-Ax = x$ or $(I_n - A^-A)x = x \Rightarrow x \in R(I_n - A^-A)$

$\forall x \in N(A) \Rightarrow N(A) \subseteq R(I_n - A^- A)$. Thus in $(*)$ $B_{N(A)}$ can be replaced by $I_n - A^- A$, and we obtain the *general solution of the homogeneous system*

$$x_h = (I_n - A^- A)y, \quad y \in C^n \text{ arbitrary, } A^- \text{ arbitrary (1)-inverse of } A.$$

Remark. $I_n - A^- A$ is a (in general, nonorthogonal or oblique) projector from C^n on $N(A)$.

6.3. Solution of the consistent inhomogeneous system $Ax = b$. General solution

$$x = x_1 + x_h,$$

$x_1 = x_1(b)$ special solution, e.g. $x_1 = A^- b$ (which follows from the condition of consistency $AA^- b = b$) or

$$x = A^- b + (I_n - A^- A)y, \quad y \in C^n \text{ arbitrary}.$$

Special case: *underdetermined row regular system:* $\overset{n}{\square} m$
$r(A) = m < n$, always solvable, because a right inverse exists and (since $AA_R^{-1} = I$) therefore $AA_R^{-1} b = b$ is true.

General solution:

$$x = A_R^{-1} b + (I_n - A_R^{-1} A)y, \quad y \in C^n \text{ arbitrary}.$$

Special right inverse: $A_R^{-1} = A^*(AA^*)^{-1}$.

6.4. Minimum norm solution (normal solution). Unique solution of the inhomogeneous system with minimal Euclidean norm: $\|x\| = \min\limits_{Ax=b}$. We choose the particular solution

$$\bar{x} \in N(A)^\perp = R(A^*),$$

then

$$\|x\|^2 = \|\bar{x}\|^2 + \|x_h\|^2 \geqslant \|\bar{x}\|^2,$$

that is, $\bar{x}$ is the solution of minimum norm. One obtains $\bar{x}$ by orthogonal projecting any particular solution onto $R(A^*)$:

$$\bar{x} = x_{\min} = P_{R(A^*)} x_1.$$

Since $P_{R(A^*)} = A^+ A = A^{1,4} A$ and $A^{1,4} A x_1 = A^{1,4} b$, we get the result

$$x_{\min} = A^{1,4} b, \quad A^{1,4} \text{ arbitrary } (1,4)\text{-inverse of } A.$$

6.5. Least-squares solution.

THEOREM 3. *The following conditions characterizing x as least-squares solution of the inconsistent system $Ax = b$ are equivalent.*

(1) $\|Ax - b\| = \min\limits_{x \in C^n}$,

$$\left.\begin{array}{l}(2)\ Ax = P_{R(A)}b,\\ (3)\ A^*Ax = A^*b\ \textit{(Gaussian transformation)}\end{array}\right\}\ \textit{consistent}.$$

Proof. (1) $\Leftrightarrow$ (2): Consider the expression

$$(*)\qquad \|Ax-b\|^2 = \|(Ax-P_{R(A)}b)-(I-P_{R(A)})b\|^2.$$

Obviously,

$$(Ax-P_{R(A)}b)\in R(A) \quad\text{and}\quad (I-P_{R(A)})b\in R(A)^{\perp}$$

because of $b^*(I-P_{R(A)})^*Az = b^*(I-P_{R(A)})Az = 0$. Using the generalized Pythagorean proposition $(*)$ can be written in the form

$$\|Ax-b\|^2 = \|Ax-P_{R(A)}b\|^2+\|(I-P_{R(A)})b\|^2.$$

Obviously, this expression takes its smallest value if and only if $Ax = P_{R(A)}b$.

(2) $\Leftrightarrow$ (3): From (2) it follows with (P1), (P3) and $A^{+*} = A^{*+}$ by multiplying on the left by A^*

$$A^*Ax = A^*P_{R(A)}b = A^*AA^+b = A^*A^{*+}A^*b = A^*b,$$

i.e. (3).

Conversely, multiplying (3) on the left by A^{+*} gives

$$A^{+*}A^*Ax = A^{+*}A^*b$$

or with (P3)

$$AA^+Ax = Ax = AA^+b = P_{R(A)}b,$$

i.e. (2).

The *general least-squares solution* is obtained, for instance, by solving the consistent system $Ax = P_{R(A)}b$, namely

$$x = A^-P_{R(A)}b+x_h$$

or

$$x = A^{1,3}b+(I_n-A^-A)y,$$

$y\in C^n$ arbitrary, and $A^{1,3}$ arbitrary $(1,3)$-inverse of A, because $A^-P_{R(A)} = A^-AA^+ = A^{1,3}$, for A^-AA^+ satisfies (P1) and (P3).

Special case: *overdetermined column regular system*: $\square^{m}_{n}$ $r(A) = n < m$. From (3) follows the unique solution

$$x = (A^*A)^{-1}A^*b.$$

$(A^*A)^{-1}A^* = A_L^{-1}$ is a special left inverse of A, because of $A_L^{-1}A = I$.

6.6. Best approximate solution (minimal least-squares solution, pseudo normal solution). Unique least-squares solution of minimum Euclidean norm, that is, the solution of the minimum problem

$$\|x\| = \min_{x \in K} \quad \text{with} \quad K = \{x \mid \|Ax-b\| = \min_{x \in C^n}\}.$$

Similarly to the minimum norm solution we obtain the best approximate solution by orthogonal projecting a least-squares solution on $R(A^*)$:

$$\begin{aligned} x_{\text{best}} &= P_{R(A^*)}[A^{1,3}b + (I_n - A^- A)y] \\ &= A^+AA^{1,3}b = A^+A(A^-AA^+)b = A^+AA^+b, \end{aligned}$$

hence

$$x_{\text{best}} = A^+b.$$

In principle: all cases can be covered by

$$x = A^+b - (I_n - A^+A)y, \qquad y \in C^n \text{ arbitrary}.$$

$y = 0$ gives the x of smallest norm. But in general the computational amount will be smaller if instead of A^+ the appropriate generalized inverses A^-, $A^{1,3}$ or $A^{1,4}$ are used.

7. Computation of generalized inverses and numerical solution of linear equations by transforming into normal form

7.1. Generalized inverses.

PROCEDURE. Transforming the given matrix into a simpler form and representing the generalized inverses by means of the transformation matrices.

THEOREM 4. *Let the matrix $A \in C^{m\times n}$ of rank $r < \min(m, n)$ be transformed into the normal form*

$$PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

by the nonsingular matrices $P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$ and $Q = (Q_1, Q_2)$, where P_1 indicates the first r rows of P and Q_1 the first r columns of Q. Then, the generalized inverses of A have the following general representations in which X, Y and Z are arbitrary matrices of suitable size.

(1) $$A^- = Q\begin{bmatrix} I_r & X \\ Y & Z \end{bmatrix}P \qquad \text{(Rohde [42])},$$

(2) $$\begin{aligned} A^{1,2} &= Q\begin{bmatrix} I_r & X \\ Y & YX \end{bmatrix}P \qquad \text{(Rohde [42])} \\ &= Q\begin{bmatrix} I_r \\ Y \end{bmatrix}(I_r, X)P, \end{aligned}$$

(3) $$A^{1,3} = Q\begin{bmatrix} I_r & -P_1P_2^+ \\ Y & Z \end{bmatrix} P,$$

(4) $$A^{1,4} = Q\begin{bmatrix} I_r & X \\ -Q_2^+Q_1 & Z \end{bmatrix} P,$$

(5) $$A^{1,2,3} = Q\begin{bmatrix} I_r & -P_1P_2^+ \\ Y & -YP_1P_2^+ \end{bmatrix} P \quad \text{(Morris, Odell [31])}$$
$$= Q\begin{bmatrix} I_r \\ Y \end{bmatrix} (I_r, \ -P_1P_2^+)P,$$

(6) $$A^{1,2,4} = Q\begin{bmatrix} I_r & X \\ -Q_2^+Q_1 & -Q_2^+Q_1X \end{bmatrix} P = Q\begin{bmatrix} I_r \\ -Q_2^+Q_1 \end{bmatrix} (I_r, X)P,$$

(7) $$A^{1,3,4} = Q\begin{bmatrix} I_r & -P_1P_2^+ \\ -Q_2^+Q_1 & Z \end{bmatrix} P,$$

(8) $$A^{+} = Q\begin{bmatrix} I_r & -P_1P_2^+ \\ -Q_2^+Q_1 & Q_2^+Q_1P_1P_2^+ \end{bmatrix} P \quad \text{(Morris, Odell [31])}$$
$$= Q\begin{bmatrix} I_r \\ -Q_2^+Q_1 \end{bmatrix} (I_r, \ -P_1P_2^+)P.$$

Proof (see Zielke [49], [52]).

Remarks. 1. Because of

$$A^{1,2} = A_1^-AA_2^- \quad (A_1^-, A_2^- \text{ arbitrary (1)-inverses})$$
$$= Q\begin{bmatrix} I_r & X_1 \\ Y_1 & Z_1 \end{bmatrix} PP^{-1}\begin{bmatrix} I_r \\ 0 \end{bmatrix} (I_r, 0)Q^{-1}Q\begin{bmatrix} I_r & X_2 \\ Y_2 & Z_2 \end{bmatrix} P = Q\begin{bmatrix} I_r \\ Y_1 \end{bmatrix} (I_r, X_2)P$$

the generalized inverses $A^{1,2}$, $A^{1,2,3}$, $A^{1,2,4}$ and A^+ can be written in this product form with arbitrary or specified Y_1, X_2.

2. Since P_2 and Q_2 have full rank, it holds

$$P_2^+ = P_2^*(P_2P_2^*)^{-1}, \quad Q_2^+ = (Q_2^*Q_2)^{-1}Q_2^*.$$

3. If A is a full-rank matrix, then, in the case

A column regular ▯ :

$$A^- = A^{1,2} = A^{1,4} = A^{1,2,4} = Q(I_n, X)P, \quad X \text{ arbitrary},$$
$$A^{1,3} = A^{1,2,3} = A^{1,3,4} = A^+ = Q(I_n, \ -P_1P_2^+)P.$$

A row regular ▭ :

$$A^- = A^{1,2} = A^{1,3} = A^{1,2,3} = Q\begin{bmatrix} I_m \\ Y \end{bmatrix} P, \quad Y \text{ arbitrary},$$

$$A^{1,4} = A^{1,2,4} = A^{1,3,4} = A^+ = Q\begin{bmatrix} I_m \\ -Q_2^+Q_1 \end{bmatrix} P.$$

4. Also transformations into other normal forms and by other matrices (e.g. orthogonal (unitary) matrices) are possible and usual.

COROLLARY. *If X, Y and Z in Theorem 4 are replaced by zero matrices, then we obtain the particular generalized inverses*

$$A_1^- = Q\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P = Q_1P_1,$$

$$A_1^{1,3} = Q\begin{bmatrix} I_r & -P_1P_2^+ \\ 0 & 0 \end{bmatrix} P = Q_1P_1(I_m - P_2^+P_2),$$

$$A_1^{1,4} = Q\begin{bmatrix} I_r & 0 \\ -Q_2^+Q_1 & 0 \end{bmatrix} P = (I_n - Q_2Q_2^+)Q_1P_1.$$

Moreover, A^+ can be written in the product form

$$A^+ = (I_n - Q_2Q_2^+)Q_1P_1(I_m - P_2^+P_2).$$

7.2. Linear equations.

THEOREM 5. *Let $A \in C^{m\times n}$ of rank r be transformed into the normal form $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ as in Theorem 4. Then, for the solution of linear equations $Ax = b$ we yield the following results:*

Condition of consistency (necessary and sufficient for the solvability of $Ax = b$)

(1) $$P_2b = 0.$$

General solution of the homogeneous system $Ax = 0$

(2) $$x = Q_2y, \quad y \in C^{n-r} \text{ arbitrary}.$$

General solution of the consistent inhomogeneous system $Ax = b$

(3) $$x = Q_1P_1b + Q_2y, \quad y \in C^{n-r} \text{ arbitrary}.$$

Minimum norm solution (solution of the minimum problem $\|x\| = \min\limits_{Ax=b}$)

(4) $$x = (I_n - Q_2Q_2^+)Q_1P_1b.$$

General least-squares solution (general solution of the minimum problem $\|Ax-b\| = \min\limits_{x\in C^n}$)

(5) $$x = Q_1P_1(I_m - P_2^+P_2)b + Q_2y, \quad y \in C^{n-r} \text{ arbitrary}.$$

Best approximate solution (solution of the minimum problem

$$\|x\| = \min_{x\in K} \quad \textit{with } K = \{x \mid \|Ax-b\| = \min_{x\in C^n}\}).$$

(6) $$x = (I_n - Q_2Q_2^+)Q_1P_1(I_m - P_2^+P_2)b.$$

Proof. According to Section 6 and Corollary to Theorem 4.

(1) With $A = P^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1}$ the condition of consistency $AA^-b = b$ can be written in the form

$$P^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1}Q\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Pb = P^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Pb = b$$

or

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} b = \begin{bmatrix} P_1 \\ 0 \end{bmatrix} b = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} b,$$

that is

$$P_2 b = 0.$$

(2)
$$\begin{aligned}
(I_n - A^-A)y &= \left(I_n - Q\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} PP^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1}\right)y \\
&= \left(I_n - Q\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1}\right)y \\
&= \left(Q\begin{bmatrix} I_r & 0 \\ 0 & I_{n-r} \end{bmatrix} Q^{-1} - Q\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1}\right)y \\
&= Q\begin{bmatrix} 0 & 0 \\ 0 & I_{n-r} \end{bmatrix} Q^{-1}y, \qquad y \in C^n \text{ arbitrary}, \\
&= (Q_1, Q_2)\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}(0, I_{n-r})\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\
&\qquad \text{with } \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = z = Q^{-1}y, \ z \in C^n \text{ arbitrary}, \\
&= Q_2 z_2, \qquad z_2 \in C^{n-r} \text{ arbitrary}.
\end{aligned}$$

(3) to (6) are obviously true.

For other transformations into block forms see Zielke [54].

7.3. Numerical example. Solve 3 linear systems $Ax = b$, where

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 1+\varepsilon \end{bmatrix}, \quad b^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad b^{(1)} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \quad b^{(2)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Transforming into normal form $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ by elementary row or column operations: interchanging, multiplication by a nonzero scalar, addition of a multiple of a row (column) to another. Determination of P and Q (not unique) according to the scheme

$$P\;\begin{array}{c|c} I_m & A \\ \hline & I_n \end{array}\;Q \;\Rightarrow\; \begin{array}{c|c} P & \begin{matrix} I_r & 0 \\ 0 & 0 \end{matrix} \\ \hline & Q \end{array}$$

Transformation steps:

$$\begin{array}{cccc|ccc} 1 & & & & 1 & 0 & 1 \\ & 1 & & & -1 & 1 & 0 \\ & & 1 & & 1 & -1 & 0 \\ & & & 1 & 0 & 1 & 1+\varepsilon \\ \hline & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{array} \quad \begin{matrix} 1. \\ 2. \\ \Rightarrow \end{matrix} \quad \begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ -1 & 0 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1+\varepsilon \\ \hline & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{array} \quad \begin{matrix} 3. \\ 4. \\ \Rightarrow \end{matrix} \quad \begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 & \varepsilon \\ \hline & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{array}$$

$$\begin{matrix} 5. \\ 6. \\ \Rightarrow \end{matrix} \quad \begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 & \varepsilon \\ \hline & & & & 1 & 0 & -1 \\ & & & & 0 & 1 & -1 \\ & & & & 0 & 0 & 1 \end{array} \quad \begin{matrix} 7. \\ \Rightarrow \end{matrix} \quad \begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 & \varepsilon \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline & & & & 1 & 0 & -1 \\ & & & & 0 & 1 & -1 \\ & & & & 0 & 0 & 1 \end{array}$$

$$\begin{matrix} 8. \\ \Rightarrow \end{matrix} \quad \begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ \frac{-1}{\varepsilon} & \frac{-1}{\varepsilon} & 0 & \frac{1}{\varepsilon} & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline & & & & 1 & 0 & -1 \\ & & & & 0 & 1 & -1 \\ & & & & 0 & 0 & 1 \end{array}$$

1.: Add the 1st row to the 2nd row.
2.: Add (-1) times the 1st row to the 3rd row.
3.: Add the 2nd row to the 3rd row.
4.: Add (-1) times the 2nd row to the 4th row.
5.: Add (-1) times the 1st column to the 3rd column.
6.: Add (-1) times the 2nd column to the 3rd column.
7.: Interchange the 3rd and the 4th row.
If ε lies within the scope of the rounding errors, set $\varepsilon = 0$. Then the numerical rank is 2.

8.: If ε lies significantly above the rounding error level, then e.g., multiply the 3rd row by $1/\varepsilon$ and hence the numerical rank is 3.

Results. If $\varepsilon = 0$ in the given matrix A, we obtain after the 6th step

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 0 & 1 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{bmatrix}, \quad Q = (Q_1, Q_2) = \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{array}\right],$$

thus $r(A) = 2$. In this case the following results are obtained.

Condition of consistency $P_2 b = 0$: because of

$$P_2 b^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad P_2 b^{(2)} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

it is satisfied for $b = b^{(1)}$, but not for $b = b^{(2)}$.

Solution of the homogeneous system $Ax = 0$:

$$x = Q_2 y = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} y, \quad y \in C^{n-r} \text{ (scalar)}.$$

Solution of the inhomogeneous consistent system $Ax = b^{(1)}$:

$$x = Q_1 P_1 b^{(1)} + Q_2 y = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} y = \begin{bmatrix} 1-y \\ 2-y \\ y \end{bmatrix}.$$

Minimum norm solution:

$$x = (I_n - Q_2 Q_2^+) Q_1 P_1 b^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \text{with} \quad Q_2^+ = \tfrac{1}{3}(-1, -1, 1).$$

Least-squares solution of $Ax = b^{(2)}$:

$$x = Q_1 P_1 (I_m - P_2^+ P_2) b^{(2)} + Q_2 y = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} y = \begin{bmatrix} 1-y \\ 1-y \\ y \end{bmatrix}$$

with

$$P_2^+ = \frac{1}{5} \begin{bmatrix} -1 & -2 \\ 2 & -1 \\ 3 & 1 \\ 1 & 2 \end{bmatrix}.$$

Best approximate solution:

$$x = (I_n - Q_2Q_2^+)Q_1P_1(I_m - P_2^+P_2)b^{(2)} = \frac{1}{3}\begin{bmatrix}1\\1\\2\end{bmatrix}.$$

Moreover, we obtain the particular *generalized inverse* (even $(1,2)$-inverse)

$$A_1^- = Q_1P_1 = \begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\begin{bmatrix}1&0&0&0\\1&1&0&0\end{bmatrix} = \left[\begin{array}{cc|cc}1&0&0&0\\1&1&0&0\\\hline 0&0&0&0\end{array}\right] = \begin{bmatrix}A_{11}^{-1}&0\\0&0\end{bmatrix},$$

and the *Moore–Penrose inverse*

$$A^+ = \frac{1}{15}\begin{bmatrix}4&-3&3&1\\1&3&-3&4\\5&0&0&5\end{bmatrix}.$$

In the case $\varepsilon \neq 0$ we obtain, also for ε with a small absolute value (and just for these values), completely different results. This shows already A_1^- which reads for $\varepsilon \neq 0$:

$$A_1^- = QP_1 = \begin{bmatrix}1&0&-1\\0&1&-1\\0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\1&1&0&0\\\frac{-1}{\varepsilon}&\frac{-1}{\varepsilon}&0&\frac{1}{\varepsilon}\end{bmatrix} = \begin{bmatrix}1+\frac{1}{\varepsilon}&\frac{1}{\varepsilon}&0&\frac{-1}{\varepsilon}\\1+\frac{1}{\varepsilon}&1+\frac{1}{\varepsilon}&0&\frac{-1}{\varepsilon}\\\frac{-1}{\varepsilon}&\frac{-1}{\varepsilon}&0&\frac{1}{\varepsilon}\end{bmatrix}.$$

From this comparison we learn that the elements of generalized inverses do not continuously depend on the elements of the given matrix if the rank changes.

7.4. Minimum rank. In general, an exact determination of the rank of (nearly) rank deficient matrices is not possible in a finite arithmetic. In order to avoid an incorrect increase of the rank by rounding errors one uses the so-called minimum rank.

DEFINITION 5. $r_\varepsilon(A)$ is called *minimum rank* (or *numerical rank, pseudo rank*) of the matrix A with respect to a tolerance ε and a norm $\|\cdot\|$, if

$$r_\varepsilon(A) = \min_{\|\Delta A\|/\|A\| \leqslant \varepsilon} r(A + \Delta A)$$

for a sufficiently large $\varepsilon > 0$.

Because a slight increase of ε can decrease the minimum rank we need an upper bound on ε so that for such ε $r_\varepsilon(A) < r(A)$ does not occur. By using the spectral norm $\|A\|_2 = (\text{max. eigenvalue of } A^*A)^{1/2}$ a proper choice of ε is $\varepsilon < 1/\text{cond}_2(A)$, where $\text{cond}_2(A) = \|A\|_2 \cdot \|A^+\|_2$ is the condition number of A. But since the condition number is mostly unkown, a proper choice of ε is difficult, but not if the singular value decomposition of a matrix is used for determining the rank. See also Golub, Klema, Stewart [19]. In order to determine the minimum rank by transformation into normal form we have to check whether the pivots are smaller than a tolerance δ which depends on the number of digits of the computer and the condition number. This procedure is practicable, although it can fail in sophisticated counter-examples, cf. Peters, Wilkinson [37].

In our computations, which were done on a R 300 computer (8 decimal digits), and for the test matrices used the following orders of $\delta/\|A\| = \varepsilon$ proved to be feasible:

ε between 10^{-8} and 10^{-5} for matrices with condition numbers up to about 10^3 ("well-conditioned"),

ε about 10^{-8} for matrices with larger condition numbers ("ill-conditioned").

We used the Frobenius norm $\|A\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2}$. It is advisable for matrices which are not known to be well-conditioned to check the correctness of the computations (and hence the correct determination of the rank) by inserting the computed A^+ into the Penrose equations.

8. Solution of the matrix equation $AXB = C$

THEOREM 6. *For the solution of the matrix equation $AXB = C$ the following results hold, in which Y is an arbitrary matrix of suitable size.*

Condition of consistency (necessary and sufficient for the solvability of $AXB = C$):

(1) $$C = AA^-CB^-B.$$

General solution of the homogeneous system $AXB = 0$:

(2) $$X = Y - A^-AYBB^-.$$

General solution of the consistent inhomogeneous system $AXB = C$:

(3) $$X = A^-CB^- + Y - A^-AYBB^-.$$

Minimum norm solution (solution of $\|X\|_F = \min\limits_{AXB=C}$):

(4) $$X = A^{1,4}CB^{1,3}.$$

General least-squares solution (*solution of* $\|AXB-C\|_F = \min\limits_X$):

(5) $$X = A^{1,3}CB^{1,4}+Y-A^-AYBB^-.$$

Best approximate solution (*solution of* $\|X\|_F = \min\limits_{X\in K}$, $K = \{X \mid \|AXB-C\|_F = \min\limits_X\}$):

(6) $$X = A^+CB^+.$$

Remark. (1), (2), (3) are due to Penrose [35], (4) to Hearon [24], (5) to Zielke [53], (6) to Penrose [36]. In order to solve $AXB = C$ numerically the matrices A and B can be transformed into normal form and then the generalized inverses required can be computed according to Theorem 4.

9. Computation of column spaces, null spaces and projectors

A further application of the transformation into normal form.

THEOREM 7. *Let* $A \in C^{m\times n}$ *of rank* r *be transformed into the normal form* $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ *by the nonsingular matrices* $P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$ *and* $Q = (Q_1, Q_2)$ *as in Theorem* 4. *Then we obtain: the bases*

(1) $B_{R(A)} = AQ_1$ *basis for the column space of* A (*r linear independent columns of* A),

(2) $B_{N(A)} = Q_2$ *basis for the null space of* A ($x = Q_2 y$ *is the general solution of the homog. system* $Ax = 0$),

(3) $B_{R(A^*)} = (P_1A)^*$,

(4) $B_{N(A^*)} = P_2^*$;

the orthogonal projectors

(5) $$\begin{aligned} P_{R(A)} &= AQ_1(AQ_1)^+ \text{ orthogonal projector on } R(A) \\ &= AQ_1[(AQ_1)^*AQ_1]^{-1}(AQ_1)^* && (r < m/2) \\ &= I_m - P_{N(A^*)}, && (r \geqslant m/2) \end{aligned}$$

(6) $$\begin{aligned} P_{N(A)} &= Q_2Q_2^+ = Q_2(Q_2^*Q_2)^{-1}Q_2^* && (r \geqslant n/2) \\ &= I_n - P_{R(A^*)}, && (r < n/2) \end{aligned}$$

(7) $$\begin{aligned} P_{R(A^*)} &= (P_1A)^+P_1A = (P_1A)^*[P_1A(P_1A)^*]^{-1}P_1A && (r < n/2) \\ &= I_n - P_{N(A)}, && (r \geqslant n/2) \end{aligned}$$

(8) $$\begin{aligned} P_{N(A^*)} &= P_2^+P_2 = P_2^*(P_2P_2^*)^{-1}P_2 && (r \geqslant m/2) \\ &= I_m - P_{R(A)} && (r < m/2) \end{aligned}$$

Proof. See Zielke [49].

Remark. The inequalities $r < m/2$ etc. refers only to the smallest computational amount, but not to a limitation of the validity.

10. Other methods for computing generalized inverses

Main method: Full-rank factorization (Egerváry [16], Greville [23]).

THEOREM 8. *Any matrix $A \in C^{m\times n}$ of rank r can be expressed as a product*

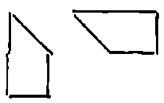
$$A = BC, \quad \textit{where} \quad B \in C^{m\times r} \textit{ and } C \in C^{r\times n}$$

are full-rank matrices, both of rank r. Then

$$A^+ = C^+B^+ \quad \textit{with} \quad C^+ = C^*(CC^*)^{-1},\ B^+ = (B^*B)^{-1}B^*.$$

PROCEDURES:

1. *LU decomposition* (Kublanowskaja [26], Peters, Wilkinson [37])

$A = LU =$  L: lower trapezoidal, with units on the principal diagonal,
U: upper trapezoidal.

Computation of L and U by means of Gauss elimination.

Note: Usually, the decomposition is performed along with some form of pivoting (interchanging rows and/or columns), that means, in reality we start from a matrix $\tilde{A}$ and obtain the factorization

$$T_1\tilde{A}T_2 = A = LU,$$

where T_1 and T_2 are permutation matrices. Then

$$A^+ = U^+L^+ = U^*(UU^*)^{-1}(L^*L)^{-1}L^*.$$

First, compute X from $L^*LX = L^*$, Y from $UU^*Y = X$, then $A^+ = U^*Y$ and $\tilde{A}^+ = T_2A^+T_1$.

2. *Orthogonal decomposition*

$$A = QR,$$

Q: orthonormal columns, i.e. $Q^*Q = I_r$, R: upper trapezoidal.

Computation of Q and R by means of Householder transformation (Korganoff, Pavel-Parvu [25] or modified Gram–Schmidt orthogonalization (Osborne [34]) or Givens method (Kublanowskaja [26]). Then,

$$A^+ = R^*(RR^*)^{-1}Q^*.$$

3. *Singular value decomposition* (Golub, Kahan [18])

$$A = USV^* = \sum_{i=1}^{r} \sigma_i u_i v_i^*,$$

$S = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$, $\sigma_i = \sqrt{\lambda_i(A^*A)} > 0$ singular values of A,

$U = (u_1, \ldots, u_r)$, $V = (v_1, \ldots, v_r)$ with orthonormal columns u_i, v_i (singular vectors, the eigenvectors of AA^* and A^*A, respectively, corresponding to the eigenvalues $\lambda_i \neq 0$ of A^*A (or AA^*)).

Then

$$A^+ = VS^{-1}U^* = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^*.$$

Note. If the computer yields $S = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$, $\sigma_1 \geqslant \ldots \geqslant \sigma_n$, then A has minimal rank $r_\varepsilon(A) = r$ relative to $\|\cdot\|_2$ if and only if

$$\frac{\sigma_{r+1}}{\sigma_1} \leqslant \varepsilon < \frac{\sigma_r}{\sigma_1} = \frac{1}{\operatorname{cond}_2(A)},$$

compare Golub, Klema, Stewart [19]. That means, all the singular values σ_i, $i = r+1, \ldots, n$, for which $\sigma_i \leqslant \varepsilon\sigma_1$ holds must be set equal to zero. A reasonable choice of ε is $\varepsilon = \eta$ where η is the relative machine precision $\eta = 0.5B^{-t+1}$ (t number of digits, B base of the number system).

4. *Block decomposition* (Noble [33]). Obviously, any matrix of known rank $r < \min(m, n)$ can be partitioned into the block form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A_{11} \in C^{r\times r} \text{ nonsingular},$$

by interchanging rows and columns if necessary. Then, because of $A_{22} = A_{21}A_{11}^{-1}A_{12}$, it holds

$$A = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} A_{11}^{-1}(A_{11}, A_{12}), \qquad A_1^- = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

$$\begin{aligned} A^+ &= (A_{11}, A_{12})^+ A_{11} \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^+ \\ &= \begin{bmatrix} A_{11}^* \\ A_{12}^* \end{bmatrix} (A_{11}A_{11}^* + A_{12}A_{12}^*)^{-1} A_{11} (A_{11}^*A_{11} + A_{21}^*A_{21})^{-1} (A_{11}^*, A_{21}^*). \end{aligned}$$

The general forms of the other generalized inverses can be found in Zielke [49], [52].

5. *Transformation into Hermite normal form*

$$PAT = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix}, \quad P \text{ nonsingular, } T \text{ permutation matrix.}$$

Computation of P, T, K by Gauss elimination. Then

$$A^+ = T \begin{bmatrix} I_r \\ K^* \end{bmatrix} (I_r + KK^*)^{-1} (I_r, -P_1P_2^+) P \quad \text{(Zielke [49], [52]).}$$

The equation

$$(I_r+KK^*)^{-1} = I_r-K(I_{n-r}+K^*K)^{-1}K^*$$

can be used to reduce the computational amount if r is relatively large or small. We propose to use the left-hand side if $r < \frac{2}{3}n$, otherwise the right-hand side. In this way and by a suitable choice of the succession of matrix multiplications it is possible to achieve for large $m = n = r$ the minimal computational amount of about n^3 multiplications.

6. *Generalized iteration of Schulz* (Ben-Israel [2], [3], Ben-Israel, Cohen [5]). Starting from

$$X_0 = aA^*, \quad 0 < a < \frac{2}{\lambda_{\max}(A^*A)}, \quad \text{e.g.} \quad a = \frac{1}{\|A^*A\|},$$

the sequence

$$(*) \qquad X_{k+1} = X_k+X_k(I-AX_k), \quad k = 0, 1, \ldots$$

converges to A^+ as $k\to\infty$. Practically, the algorithm needs at least $2\log_2[\mathrm{cond}_2(A)]$ iterations (e.g. $\mathrm{cond}_2(A) = 10^3 \Rightarrow k > 20$).

Disadvantage of the method: not self-correcting, numerically divergent, not suitable for iterative refinement of an approximation to A^+.

Remark. Any approximation X_0 to A^+ can be adapted to the space conditions $X = YA^* = A^*Z$ (see Definition 3) by orthogonal projecting $P_{R(A^*)}X_0P_{R(A)} = A^+AX_0AA^+$. But the exact A^+ is unknown.

An approximation $\tilde{X}_0$ which exactly satisfies the space conditions can be obtained without knowing A^+ by using the Penrose equations (P3) $AA^+ = A^{+*}A^*$ and (P4) $A^+A = A^*A^{+*}$ and then replacing A^+ by X_0:

$$(**) \qquad \tilde{X}_0 = A^*X_0^*X_0X_0^*A^*.$$

The iterative method $(*)$ with $\tilde{X}_0$ as initial approximation converges to A^+ if X_0 is a sufficiently close approximation to A^+ and $\mathrm{cond}(A)$ is not too large. In order to improve the accuracy compared with that obtained by a good direct method the iterative process must be performed in double precision arithmetic. This process is self-correcting if the (oblique) projection $(**)$ is repeated after some iterations (Zielke [50]). See also Shinozaki, Sibuya, Tanabe [43], [44].

11. Test matrices and numerical results

More than 30 methods for computing A^+ were tested by test matrices whose condition numbers can be changed by a real or complex parameter a. Three of the simplest matrices are the following:

$$A_1 = \begin{bmatrix} a & a & a-1 & a \\ a+1 & a & a & a \\ a & a & a-1 & a \\ a+1 & a & a & a \\ a+1 & a+1 & a & a+1 \end{bmatrix}, \quad A_1^+ = \frac{1}{4}\begin{bmatrix} 2a & 2 & 2a & 2 & -4a \\ 0 & -1 & 0 & -1 & 2 \\ -2a-2 & 0 & -2a-2 & 0 & 4a \\ 0 & -1 & 0 & -1 & 2 \end{bmatrix},$$

$\operatorname{rank}(A_1) = 3$,

a	0	1	10	10^2	10^3	10^4
$\operatorname{cond}_F(A_1)$	3.5	14	$8.0_{10}2$	$7.8_{10}4$	$7.8_{10}6$	$7.8_{10}8$

$$A_2 = \begin{bmatrix} a+1 & a & a & a+1 \\ a+2 & a+1 & a+1 & a+2 \\ a+3 & a+2 & a+2 & a+3 \\ a+1 & a+1 & a & a+2 \\ a & a & a-1 & a+1 \end{bmatrix},$$

$$A_2^+ = \frac{1}{60}\begin{bmatrix} 12a+44 & 20 & -12a-4 & -6a-27 & 6a-3 \\ -12a-56 & -20 & 12a+16 & 6a+33 & -6a-3 \\ -12a-12 & 0 & 12a+12 & 6a-9 & -6a-21 \\ 12a & 0 & -12a & -6a+15 & 6a+15 \end{bmatrix},$$

$\operatorname{rank}(A_2) = 3$,

a .	0	1	10	10^2	10^3	10^4
$\operatorname{cond}_F(A_2)$	11	21	$3.7_{10}2$	$2.9_{10}4$	$2.8_{10}6$	$2.8_{10}8$

$$A_3 = \begin{bmatrix} a & a+1 & a+2 & a+3 & a \\ a & a+2 & a+3 & a+5 & a+1 \\ a+1 & a+2 & a+3 & a+4 & a+2 \\ a+2 & a+3 & a+4 & a+5 & a+3 \\ a+3 & a+4 & a+5 & a+6 & a+5 \\ a+5 & a+5 & a+6 & a+6 & a+7 \end{bmatrix},$$

$$A_3^+ = \tfrac{1}{8}\begin{bmatrix} 4 & -1 & -8 & 7 & -5 & 3 \\ -8 & 2a+13 & -8a-28 & 6a+17 & -2a-3 & 2a+1 \\ 10 & -2a-11 & 8a+18 & -6a-9 & 2a-1 & -2a+1 \\ -2 & 3 & -2 & 1 & 1 & -1 \\ -4 & -2 & 12 & -10 & 6 & -2 \end{bmatrix},$$

$\operatorname{rank}(A_3) = 4$,

a	0	1	10	10^2	10^3
$\operatorname{cond}_F(A_3)$	$1.3_{10}2$	$2.0_{10}2$	$1.8_{10}3$	$1.1_{10}5$	$1.0_{10}7$

Numerical results. For each tested method and each of the test matrices the "minimum number of correct decimal digits", i.e. the quantity

$$-\log_{10}(\text{maximum relative error}) = -\log_{10}\left[\max_{g_{ij}\neq 0}\left|\frac{\tilde{g}_{ij}-g_{ij}}{g_{ij}}\right|\right]$$

was determined, where $\tilde{g}_{ij}$ is the computed value of the element g_{ij} of $A^+ = (g_{ij})$. If $g_{ij} = 0$ then the maximum absolute error was used. Of course, $\tilde{g}_{ij} = g_{ij}$ $\forall i, j$ means full precision.

Abbreviations:

– rank false, · rank correct

Lim $A^+ = \lim\limits_{\varepsilon \to +0} (A^*A + \varepsilon I)^{-1} A^*$

GREV	Greville's method (successive bordering of columns)
Iter	Iterative method of Ben-Israel
SVD	Singular value decomposition
GIV	Givens' method
MGS	Modified Gram–Schmidt orthogonalization
HOUS	Householder's method
LU	LU decomposition (Gauss elimination)
HT	Transformation into Hermite normal form

Table. Number of correct decimal digits (8 decimal digits mantissa)

	a	Lim	GREV	Iter	SVD	GIV	MGS	HOUS	LU	HT
A_1	0	3.52	7.48	7.40	6.69	6.80	7.70	7.09	6.66	7.00
	1	3.86	6.26	6.80	6.45	6.80	6.92	6.69	5.80	7.00
	10	0.86	1.92	2.67	4.88	4.05	5.40	4.60	5.74	6.68
	100	0.01	–	–	3.14	1.38	1.49	3.03	4.70	5.95
	1000	–	–	–	1.24	·	·	0.86	3.00	3.00
A_2	0	2.93	6.27	5.66	6.00	5.83	6.48	6.15	6.15	6.64
	1	2.00	4.29	4.40	5.92	5.92	6.22	6.09	5.92	6.22
	10	1.29	0.79	2.91	4.67	4.94	4.21	4.74	3.41	5.49
	100	0.09	–	–	1.64	2.96	1.68	1.91	3.49	3.41
	1000	–	–	–	·	0.69	1.57	·	2.04	2.04
A_3	0	1.63	3.38	3.50	4.38	5.43	4.97	5.43	5.24	5.72
	1	0.98	3.54	3.75	3.60	4.27	4.67	4.99	5.15	5.25
	10	0.56	0.58	0.62	4.11	3.96	3.25	3.38	4.14	4.14
	100	–	–	–	0.62	2.12	0.88	0.74	1.34	1.33
	1000	–	–	–	·	·	·	·	·	·
Sum		17.73	34.51	37.71	53.34	55.15	55.44	55.70	64.78	69.87
Op (n^3)		2	1.5	$2k$	7–8	5	3.5	3.5	3.5	1
Precision		very poor	poor		good				good or very good	

Remarks. 1. Op(n^3): number of multiplications for $m = n = r \gg 1$, only of order n^3, A real, k: number of iterations.

2. The transformation into normal form ($\frac{5}{3}n^3$ op.) yields results similar to LU decomposition.

3. Concerning the rank test, SVD is the most reliable method.

12. Final remarks

12.1. Short historical notes. The theory of generalized inverses of matrices has been developed during the past 25 years caused by two papers by R. Penrose [35], [36] (generalized inverse A^+) who showed that the least squares solution of the inconsistent system $Ax = b$ with the smallest Euclidean norm is given by $x = A^+b$.

Already previously, E. H. Moore [29], [30] (general reciprocal) obtained fundamental results on generalized inverses, probably much earlier, about in 1906, see Ben-Israel, Greville [6], p. 5.

Extension to generalized inverses of linear operators in Hilbert space by Y. Y. Tseng [46]–[48] a student of Moore.

The concept of a generalized inverse was already mentioned by I. Fredholm [17] (pseudoinverse of a linear integral operator).

Further main contributions to generalized inverses of matrices were made, among others, by:

A. Bjerhammar [7], [8] (different types of generalized inverses, application to linear systems),

T. N. E. Greville [22], [23] (fundamental papers),

C. R. Rao [39], [40] (classification of generalized inverses),

A. Ben-Israel, A. Charnes [4] (state of the art in generalized inverses).

Generalized inverses solve linear algebraic systems. But there are other generalizations of inverses, for instance the Drazin inverse (Drazin [15], see also Campbell, Meyer [13]) which can solve systems of linear differential and difference equations. But how to compute the Drazin inverse?

12.2. References to books and algorithms:

Korganoff, Pavel-Parvu [25] (earliest book on generalized inverses, extensive bibliography, detailed treatment of numerical methods),

Boullion, Odell [11] (proceedings of a symposium, comprehensive bibliography),

Boullion, Odell [12] (textbook, containing exercises and 362 references),

Pringle, Rayner [38] (accurately written monograph with statistics applications),

Rao, Mitra [41] (full-length monograph, concise style, several applications of generalized inverses, especially in statistics), Albert [1] (textbook with applications in statistics, restricted to the Moore–Penrose inverse, numerous problems with solutions in a separate booklet),

Bjerhammar [9] (generalized inverses are treated only to a shorter extent, the terminology used is not up to date),

Ben-Israel, Greville [6] (a comprehensive and very well readable book with more than 450 exercises, containing a chapter on generalized inverses of linear operators in Hilbert spaces),

Kuhnert [27] (short introduction to the solution of linear equations by means of the Moore–Penrose inverse, regularization methods are also considered),

Nashed [32] (proceedings of an advanced seminar, collection of excellent survey papers, containing the most comprehensive annotated bibliography on generalized inverses and applications, 1776 references),

Groetsch [21] (representation and approximation of generalized inverses of linear operators in Hilbert spaces),

Campbell, Meyer [13] (well-written unified treatment of generalized inverses including the theory and applications of Drazin inverses).

In addition, two books are mentioned which use generalized inverses in connection with the numerical solution of linear equations:

Stewart [45] (introduction to matrix computation, textbook),

Lawson, Hanson [28] (detailed numerical treatment of least-squares problems).

Algorithms. Some algorithms and computer codes for computing least-squares solutions and Moore–Penrose inverses of matrices of arbitrary rank, which are well elaborated and tested and highly recommended in the literature, can be found in:

Björck [10] (modified Gram-Schmidt orthogonalization, ALGOL procedures),

Golub, Reinsch [20] (singular value decomposition, ALGOL procedures),

Lawson, Hanson [28] (Householder transformation, singular value decomposition, FORTRAN codes),

Dongarra, Bunch, Moler, Stewart [14] (Householder transformation singular value decomposition, FORTRAN subroutines).

References

[1] A. Albert, *Regression and the Moore–Penrose pseudoinverse,* Academic Press, New York 1972, 180 pp.

[2] A. Ben-Israel, *An iterative method for computing the generalized inverse of an arbitrary matrix,* Math. Comp. 19 (1965), 452–455.

[3] —, *A note on an iterative method for generalized inversion of matrices,* ibid. 20 (1966), 439–440.

[4] A. Ben-Israel, A. Charnes, *Contributions to the theory of generalized inverses*, SIAM J. Appl. Math. 11 (1963), 667–699.

[5] A. Ben-Israel, D. Cohen, *On iterative computation of generalized inverses and associated projections*, ibid. 3 (1966), 410–419.

[6] A. Ben-Israel, T. N. E. Greville, *Generalized inverses: Theory and applications*, Wiley-Interscience, New York 1974, 395 pp.

[7] A. Bjerhammar, *Application of calculus of matrices to method of least squares; with special reference to geodetic calculations*, Kungl. Tekn. Högsk. Handl. (Trans. Royal Inst. of Technology), Stockholm, 49 (1951), 1–86.

[8] —, *A generalized matrix algebra*, ibid. 124 (1958), 1–32.

[9] —, *Theory of errors and generalized inverse matrices*, Elsevier Sci. Pub. Comp., Amsterdam 1973, 420 pp.

[10] Å. Björck, *Iterative refinement of linear least squares solutions II*, BIT 8(1968), 8–30.

[11] T. L. Boullion, P. L. Odell (Eds.), *Proceedings of the Symposium on Theory and Applications of Generalized Inverses of Matrices, held at the Department of Mathematics, Texas Technological College*, Lubbock, Texas, March 1968, 315 pp.

[12] —, —, *Generalized inverse matrices*, Wiley-Interscience, New York 1971, 103 pp.

[13] S. L. Campbell, C. D. Meyer, Jr., *Generalized inverses of linear transformations*, Pitman Pub. Co., London–San Francisco–Melbourne 1979, 272 pp.

[14] J. J. Dongarra, J. R. Bunch, C. B. Moler, G. W. Stewart, *Linpack users' guide*, Society for Industrial and Applied Mathematics, Philadelphia 1979.

[15] M. P. Drazin, *Pseudo-inverses in associative rings and semigroups*, Amer. Math. Monthly 65 (1958), 506–514.

[16] J. Egerváry, *Az inverz matrix általánositása* (*On a generalized inverse for matrices*), Magyar tud. akad. Mat. kutató int. közl. (Publ. Math. Inst. Hungar. Acad. Sci.) 1 (1956), 315–324.

[17] I. Fredholm, *Sur une classe d'équations fonctionnelles*, Acta Math. 27 (1903), 365–390.

[18] G. H. Golub, W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM J. Numer. Anal. 2 (1965), 205–224.

[19] G. H. Golub, V. Klema, G. W. Stewart, *Rank degeneracy and least squares problems*, STAN-CS-76-559, Computer Science Dept., Stanford Univ., Stanford, Calif., 1976, 38 pp.

[20] G. H. Golub, C. Reinsch, *Singular value decomposition and least squares solutions*, Numer. Math. 14 (1970), 403–420. Also in: J. H. Wilkinson, C. Reinsch (Eds.), *Handbook for automatic computation*, Vol. II: *Linear algebra*, Springer-Verlag, Berlin 1971, 134–151.

[21] C. W. Groetsch, *Representation and approximation of generalized inverses of linear operators*, Marcel Dekker, Inc., New York – Basel, 1977, 165 pp.

[22] T. N. E. Greville, *The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations*, SIAM Review 1 (1959), 38–43.

[23] —, *Some applications of the pseudoinverse of a matrix*, ibid. 2 (1960), 15–22.

[24] J. Z. Hearon, *Generalized inverses and solutions of linear systems*, J. Res. Mat. Bur. Standards, Sect. B, 72B (1968), 303–308.

[25] A. Korganoff, M. Pavel-Parvu, *Méthodes de calcul numérique — 2. Eléments de théorie des matrices carrées et rectangles en analyse numérique*, Dunod, Paris 1967, 441 pp.

[26] W. N. Kublanowskaja, *On calculation of a generalized inverse matrix and the projector* (Russ.), Ž. Vyčisl. Mat. i Mat. Fiz. 6 (1966), 326–332.

[27] F. Kuhnert, *Pseudoinverse Matrizen und die Methode der Regularisierung*, Teubner Verlagsges., Leipzig 1976, 88 pp.

[28] C. L. Lawson, R. J. Hanson, *Solving least squares problems*, Prentice-Hall, Englewood Cliffs, New Jersey 1974, 340 pp.

[29] E. H. Moore, *On the reciprocal of the general algebraic matrix (Abstract)*, Bull. Amer. Math. Soc. 26 (1920), 394–395.

[30] —, *General analysis*, Part I. Memoirs Amer. Philos. Soc., (1935), 231 pp., esp. 197–209.

[31] G. L. Morris, P. L. Odell, *A characterization for generalized inverses of matrices*, SIAM Review 10 (1968), 208–211.

[32] M. Z. Nashed (Ed.), *Generalized inverses and applications*, Academic Press, New York–San Francisco–London 1976, 1054 pp.

[33] B. Noble, *A method for computing the generalized inverse of a matrix*, SIAM J. Numer. Anal. 3 (1966), 582–584.

[34] E. E. Osborne, *Smallest least squares solutions of linear equations*, ibid. 2 (1965), 300–307.

[35] R. Penrose, *A generalized inverse for matrices*, Proc. Cambridge Philos. Soc. 51 (1955), 406–413.

[36] —, *On best approximate solutions of linear matrix equations*, ibid. 52 (1956), 17–19.

[37] G. Peters, J. H. Wilkinson, *The least squares problem and pseudo-inverses*, Computer J. 13 (1970), 309–316.

[38] R. M. Pringle, A. A. Rayner, *Generalized inverse matrices with applications to statistics*, Griffin, London 1971, 127 pp.

[39] C. R. Rao, *A note on a generalized inverse of a matrix with applications to problems in mathematical statistics*, J. Royal Statist. Soc., Ser. B, 24 (1962), 152–158.

[40] —, *Calculus of generalized inverses of matrices*. Part I: *General theory*, Sankhyā, Ser. A, 29 (1967), 317–342.

[41] C. R. Rao, S. K. Mitra, *Generalized inverse of matrices and its applications*, Wiley & Sons, New York 1971, 240 pp.

[42] C. A. Rohde, *Contributions to the theory, computation and application of generalized inverses*, Doctoral Dissertation, North Carolina State University, Raleigh, N. C., 1964.

[43] N. Shinozaki, M. Sibuya, K. Tanabe, *Numerical algorithms for the Moore–Penrose inverse of a matrix: Direct methods*, Ann. Inst. Statist. Math. 24 (1972), 193–203.

[44] —, —, —, *Numerical algorithms for the Moore–Penrose inverse of a matrix: Iterative methods*, ibid. 24 (1972), 621–629.

[45] G. W. Stewart, *Introduction to matrix computations*, Academic Press, New York–London 1973, 441 pp.

[46] Y. Y. Tseng, *Sur les solutions des équations opératrices fonctionnelles entre les espaces unitaires, solutions extrêmales, solutions virtuelles*, C. R. Acad. Sci. Paris, 228 (1949), 640–641.

[47] —, *Properties and classification of generalized inverses of closed operators*, Dokl. Akad. Nauk SSSR (N. S.) 67 (1949), 607–610.

[48] —, *Generalized inverses of unbounded operators between two unitary spaces*, ibid. 67 (1949), 431–434.

[49] G. Zielke, *Beiträge zur Theorie und Berechnung von verallgemeinerten inversen Matrizen*, Diss. B, Martin–Luther–Univers. Halle-Wittenberg, Fak. f. Naturwiss. d. Wiss. Rates, Halle 1977, 325 pp.

[50] G. Zielke, *Iterative refinement of generalized matrix inverses now practicable*, SIGNUM Newsletter 13.4 (1978), 9–10.
[51] —, *Test matrices for generalized inverses*, ibid. 13.4 (1978), 10–12.
[52] —, *Motivation und Darstellung von verallgemeinerten Matrixinversen*, Beiträge zur Numer. Math. 7 (1979), 177–218.
[53] —, *Die Auflösung linearer Gleichungssystems mittels verallgemeinerter inverser Matrizen*, Wiss. Z. Univ. Halle XXIX'80 M (1980), H. 1, 5–15.
[54] —, *Die Auflösung beliebiger linearer algebraischer Gleichungssysteme durch Blockzerlegung*, Beiträge zur Numer. Math. 8 (1980), 181–199.

Presented to the Semester
Computational Mathematics
February 20 – May 30, 1980