Appeared in *Lin. Algeb. and its Appl.*, **167**(1992), 87-111

A volume associated with $m \times n$ matrices *

Adi Ben-Israel †

June 25, 1990
Revised March 9, 1991

Abstract

Let $A \in \mathbb{R}_r^{m\times n}$ with nonzero singular values $\sigma_1, \sigma_2, \cdots, \sigma_r$. The **volume** of A, $\operatorname{vol} A$, is defined as zero if $r = 0$, and otherwise,

$$\operatorname{vol} A = \prod_{i=1}^{r} \sigma_i \ , \quad \text{or equivalently,} \quad \operatorname{vol} A = \sqrt{\sum \det^2 A_{IJ}} \ ,$$

summing over all $r\times r$ nonsingular submatrices A_{IJ}. The matrix volume $\operatorname{vol} A$ generalizes the "absolute value of determinant" from nonsingular to arbitrary matrices. Any r-dimensional unit cube in $R(A^T)$ is mapped, under A, into a parallelepiped of volume $\operatorname{vol} A$.

This paper gives properties and applications of $\operatorname{vol} A$. In particular, the Moore-Penrose inverse of A is a convex combination of (ordinary) inverses of its maximal nonsingular submatrices, [4],

$$A^\dagger = \sum_{I,J} \lambda_{IJ} \ \widehat{A_{IJ}^{-1}} \ ,$$

or as a convex combination of Moore-Penrose inverses of its maximal full-rank submatrices,

$$A^\dagger = \sum_{I} \lambda_{I*} \ \widehat{A_{I*}^{\dagger}} = \sum_{J} \lambda_{*J} \ \widehat{A_{*J}^{\dagger}} \ ,$$

where the convex weights λ are proportional to the squares of volumes of the corresponding submatrices.

Since $A = A^{\dagger\,\dagger}$, the matrix A is a convex combination of inverses of submatrices of $A^\dagger$. Interestingly, the weights of these convex combinations are the above weights (in transposed position), see Theorem 6.2.

Key words: Singular values. Determinants. Volume. Least squares solutions. Moore-Penrose inverse.

*Supported by National Science Foundation Grant DDM-8996112.

†RUTCOR-Rutgers Center for Operations Research and Department of Mathematics, Rutgers University, P.O. Box 5062, New Brunswick, NJ 08903-5062, U.S.A.

1 Introduction

For $A \in \mathbb{R}_r^{m\times n}$ we denote by $\mathcal{N}(A)$, or by $\mathcal{N}$, the **index-set of nonsingular** $r \times r$ **submatrices** A_{IJ}. Berg [4] showed that the Moore-Penrose inverse $A^\dagger$ is a convex combination of ordinary inverses $\{A_{IJ}^{-1} : (I,J) \in \mathcal{N}\}$,

$$A^\dagger = \sum_{(I,J)\in\mathcal{N}(A)} \lambda_{IJ}\, \widehat{A_{IJ}^{-1}} \ , \tag{1}$$

where $\widehat{X}$ denotes that X is padded with the right number of zeros in the right places.

Equivalently, for any $\mathbf{b} \in \mathbb{R}^m$, the **minimum-norm least-squares solution**[1] of the linear equations

$$A\mathbf{x} = \mathbf{b} \ , \tag{2}$$

is

$$A^\dagger\,\mathbf{b} = \sum_{(I,J)\in\mathcal{N}(A)} \lambda_{IJ}\, \widehat{A_{IJ}^{-1}\,\mathbf{b}_I} \ , \tag{3}$$

a convex combination of **basic solutions** $A_{IJ}^{-1}\,\mathbf{b}_I$, where $\mathbf{b}_I$ is the I^{th} subvector of $\mathbf{b}$. The representation (3) was given by Ben-Tal and Teboulle [3] for A of full column-rank, from which the general case follows easily.

What is curious about these convex combinations is that the weights are proportional to the squares of the determinants of the A_{IJ}'s,

$$\lambda_{IJ} = \frac{\det^2 A_{IJ}}{\displaystyle\sum_{(K,L)\in\mathcal{N}} \det^2 A_{KL}} \ , \quad (\,(I,J) \in \mathcal{N}\,) \ . \tag{4}$$

For the sake of motivation, consider a single equation

$$\sum_{j=1}^{n} a_j x_j = b \ , \tag{5}$$

with $\sum_{j=1}^n a_j^2 \neq 0$. The minimum norm solution $\mathbf{x} = (x_1, \cdots, x_n)$ is orthogonal to the hyperplane (5). Therefore $x_j = \theta\, a_j \quad (j = 1, \ldots, n)$, where θ is determined by substitution in (5),

$$\theta = \frac{b}{\displaystyle\sum_{k=1}^{n} a_k^2} \ .$$

The minimum norm solution is therefore a convex combination

$$\mathbf{x} = \sum_{a_j\neq 0} \lambda_j\, \frac{b}{a_j}\mathbf{e}_j \tag{6}$$

of "basic solutions" $\frac{b}{a_j}\,\mathbf{e}_j$, where $\mathbf{e}_j$ is the jth unit vector, $\frac{b}{a_j}$ is the intercept of the hyperplane (5) with the j^{th} coordinate axis, and the weights are

$$\lambda_j = \frac{a_j^2}{\displaystyle\sum_{k=1}^{n} a_k^2} \qquad (j = 1, \ldots, n) \ , \tag{7}$$

[1] Throughout this paper, **norm** means Euclidean norm.

which explains (4) for the case $m = 1$. This explanation works also for the general case, since by taking exterior products the system of equations (2) reduces to a single equation whose nonzero coefficients are the $r \times r$ determinants $\det A_{IJ}, \quad (I,J) \in \mathcal{N}$. The general result (3) then follows by applying (6) to the r^{th} compound matrix $C_r(A)$.

This paper studies the role played by the squares of determinants in results like (3). The sum of squares in the denominator of (4),

$$\sum_{(I,J)\in\mathcal{N}(A)} \det{}^2 A_{IJ} \ , \tag{8}$$

is shown to have special significance. If A is a full row-rank [column-rank], then (8) is the **Gramian** of A, which is the square of the volume of the parallelepiped generated by the rows [columns] of A , e. g. [7, Vol. I, Chapter IX, § 5].

For general matrices $A \in \mathbb{R}_r^{m\times n}$ the sum of squares (8) still has a volume interpretation: It is the square of the volume of a parallelepiped, with "basis " in the r-dimensional subspace $R(A^T)$, and "height" in the complementary orthogonal subspace $N(A)$. The "basis" has the square root of (8) as its r-dimensional volume, the "height" has volume 1. Hence the following:

Definition 1.1 (Volume of matrix) The **volume** of a matrix $A \in \mathbb{R}_r^{m\times n}$, denoted $\mathrm{vol}\, A$ or $\mathrm{vol}_r\, A$, is defined as 0 if $r = 0$, and otherwise

$$\mathrm{vol}\, A := \sqrt{\sum_{(I,J)\in\mathcal{N}(A)} \det{}^2 A_{IJ}} \ , \tag{9}$$

or more constructively,

$$\mathrm{vol}\, A := \prod_{i=1}^{r} \sigma_i \ , \tag{10}$$

where

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \tag{11}$$

are the nonzero **singular values** of A. The equivalence of (9) and (10) is proved in (27) below.

For convenience, we denote the sum of squares (8) by

$$\Delta_r^2(A) := \sum_{(I,J)\in\mathcal{N}(A)} \det{}^2 A_{IJ} \ . \tag{12}$$

If the matrix C is of full column-rank r then, by the Cauchy-Binet Theorem,

$$\Delta_r^2(C) = \det C^T C, \tag{13}$$

the Gramian of the columns of C. Similarly, if R is of full row-rank r,

$$\Delta_r^2(R) = \det RR^T \ . \tag{14}$$

The "volume" $\mathrm{vol}\, A$ generalizes to arbitrary matrices the "absolute value of determinant". Its applicability is illustrated in § 5 below. Other applications include:

- computing principal angles between subspaces [14],
- checking existence of integer solutions [17],
- counting spanning trees in a digraph [9, § 4.9],
- exact computation of $A^\dagger$ using residue-arithmetic [16].

The representation (10) implies some properties of $\mathrm{vol}\, A$ which are inherited from singular values. For example,

$$\mathrm{vol}_r\, (AB) \leq \mathrm{vol}_r\, A \ \mathrm{vol}_r\, B$$

is immediate from well-known inequalities for singular values. However this paper concentrates on "determinantal properties" of the matrix volume.

The plan of this paper is as follows. Preliminaries on full-rank factorizations are given in § 2. Connections with exterior products and compound matrices are outlined in § 3.

Section 4 develops the main determinantal properties of vol A. The main results here are that the volume can be written as a single determinant of a matrix obtained from A by bordering (Theorem 4.1) or complementing (Theorem 4.2).

Section 5 presents applications of geometric nature. Finally Section 6 gives various representations of the Moore-Penrose inverse. The main result here is Theorem 6.2, stating that both A and $A^\dagger$ have the same weights (4) in their convex decompositions.

2 Full rank factorizations

Let $A \in \mathbb{R}_r^{m\times n},\ r > 0$. The **full-rank factorization** of A

$$A = CR,\ \text{where } C \in \mathbb{R}_r^{m\times r},\quad R \in \mathbb{R}_r^{r\times n}\ , \tag{15}$$

is not unique, but the results below hold for <u>all</u> such C, R. Denote the set of increasing sequences of r elements from $\{1,\dots,m\}$ by

$$Q_{r,m} = \{I = \{i_1,\cdots,i_r\} : 1 \le i_1 < i_2 < \cdots < i_r \le m\} \tag{16}$$

The set $Q_{r,m}$ is ordered lexicographically. The following denote the **index-sets**

$$\mathcal{I}(A) = \{I \in Q_{r,m} :\ \text{rank } A_{I*} = r\}\ , \tag{17}$$
$$\mathcal{J}(A) = \{J \in Q_{r,n} :\ \text{rank } A_{*J} = r\}\ , \tag{18}$$
$$\mathcal{N}(A) = \{(I,J) \in Q_{r,m}\times Q_{r,n} :\ \text{rank } A_{IJ} = r\}\ , \tag{19}$$

of **maximal sets** of **linearly independent rows** and **columns**, and of **maximal nonsingular submatrices**, respectively. Clearly

$$\mathcal{I}(A) = \mathcal{I}(C),\quad \mathcal{J}(A) = \mathcal{J}(R)\ . \tag{20}$$

The index sets $\mathcal{I}(A)$, $\mathcal{J}(A)$ and $\mathcal{N}(A)$ will be abbreviated here by $\mathcal{I}$, $\mathcal{J}$ and $\mathcal{N}$ respectively.

Lemma 2.1

$$\mathcal{N} = \mathcal{I}\times\mathcal{J}\ . \tag{21}$$

<u>Proof</u>: $\mathcal{N} \subset \mathcal{I}\times\mathcal{J}$ is obvious. The converse

$$\mathcal{N} \supset \mathcal{I}\times\mathcal{J}\ ,$$

follows since every A_{IJ} is the product

$$A_{IJ} = C_{I*}R_{*J}\ .\ \ \square \tag{22}$$

An interpretation of (21), for $A \in \mathbb{R}_r^{m\times n}$ and $r > 0$, is: *The intersection of any r linearly independent rows and any r linearly independent columns is a nonsingular submatrix.*

An immediate generalization: *Let $A \in \mathbb{R}_r^{m\times n}$, $0 \le k \le r/2$. Then the intersection of any $r-k$ linearly independent rows, and any $r-k$ linearly independent columns, is a matrix of rank $\ge r-2k$.*
<u>Proof</u>: Follows from (21) since dropping a row and a column decreases the rank by at most 2. $\square$

Lemma 2.2 Let $A \in \mathbb{R}_r^{m\times n}$, $r > 0$, and let $A = CR$ be any full rank factorization. Then

$$\begin{aligned}
\Delta_r^2(A) &= \sum_{I\in\mathcal{I}} \Delta_r^2(A_{I*}) \ , && (23)\\
&= \sum_{J\in\mathcal{J}} \Delta_r^2(A_{*J}) \ , && (24)\\
&= \Delta_r^2(C)\,\Delta_r^2(R) \ . && (25)
\end{aligned}$$

Proof: Follows from Definition (12), and from (21) and (22). □

Example 2.1 (Singular value decomposition) Let $A \in \mathbb{R}_r^{m\times n}$, $r > 0$. Then the **singular value decomposition** (SVD) gives a full rank factorization of A,

$$\begin{aligned}
&A = U\Sigma V^T \ , \qquad \text{with } \ \Sigma = \ \mathrm{diag}(\sigma_i) \in \mathbb{R}^{r\times r} \ ,\\
&U = (\mathbf{u}_1, \ldots, \mathbf{u}_r) \ , \quad V = (\mathbf{v}_1, \ldots, \mathbf{v}_r)
\end{aligned} \qquad (26)$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are the nonzero **singular values** of A, and the sets $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ are o.n. bases of $R(A)$ and $R(A^T)$, respectively. Applying (25) we calculate

$$\begin{aligned}
\Delta_r^2(A) &= \Delta_r^2(U)\Delta_r^2(\Sigma V^T) = \Delta_r^2(U)\Delta_r^2(\Sigma)\Delta_r^2(V^T)\\
&= \prod_{i=1}^{r} \sigma_i^2 && (27)
\end{aligned}$$

since $U^TU = V^TV = I$. Moreover, from (26),

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad (i = 1, \ldots, r) \qquad (28)$$

showing that the r-dimensional **unit cube** $\square(\{\mathbf{v}_1, \cdots, \mathbf{v}_r\})$ is mapped under A into the cube of sides $\sigma_i\,\mathbf{u}_i$ ($i = 1, \ldots, r$), whose (r-dimensional) volume is

$$\prod_{i=1}^{r} \sigma_i \ ,$$

the volume of A. Since the singular values are unitarily invariant, it follows that all r-dimensional unit cubes in $R(A^T)$ are mapped under A into parallelepipeds of volume $\mathrm{vol}\,A$.

3 The volume in multilinear setting

The setting of multilinear algebra is natural for the matrix volume, allowing simplification of statements and proofs. The main results here are Lemmas 3.1 and 3.2.

Let $V = \mathbb{R}^n$, $U = \mathbb{R}^m$ and denote the set of linear transformations: $V \to U$ by $L(V, U)$. We use the same letter to denote both a linear transformation in $L(V, U)$ and its matrix representation with respect to fixed bases in V and U.

Let $\overset{k}{\wedge} V$ be the kth **exterior space** over V, spanned by **exterior products** $\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_k$ of elements $\mathbf{x}_i \in V$, e.g [10], [11] and [15]. For $A \in \mathbb{R}_r^{m\times n}$, $r > 0$ and $k = 1, \cdots, r$, the k-**compound matrix** $C_k(A)$ is an element of $L(\overset{k}{\wedge} V, \overset{k}{\wedge} U)$, given by

$$A\mathbf{x}_1 \wedge \cdots \wedge A\mathbf{x}_k = C_k(A)\,(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_k), \quad \forall\ \mathbf{x}_i \in V, \qquad (29)$$

see e.g [11, § 4.2, p. 94]. Then $C_k(A)$ is an $\binom{m}{k} \times \binom{n}{k}$ matrix of rank $\binom{r}{k}$, and its singular values are all products $\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_k}$ of singular values of A. It follows that $C_r(A)$ is of rank 1, and its nonzero singular value equals $\mathrm{vol}\,A$.

To any r-dimensional subspace $W \subset V$ there corresponds a unique 1-dimensional subspace $\overset{r}{\wedge} W \subset \overset{r}{\wedge} V$, spanned by the exterior product

$$\mathbf{w}^{\wedge} = \mathbf{w}_1 \wedge \cdots \wedge \mathbf{w}_r \tag{30}$$

where $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ is any basis of W, e.g. [15]. The $\binom{n}{r}$ components of $\mathbf{w}^{\wedge}$ (determined up to a multiplicative constant) are the **Plücker coordinates** of W. In particular, let o.n. bases of $R(A)$ and $R(A^T)$ be given by the $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ of the SVD (28). Then the Plücker coordinates of $R(A)$ are given by

$$\mathbf{u}^{\wedge} = \mathbf{u}_1 \wedge \cdots \wedge \mathbf{u}_r$$

and those of $R(A^T)$ by

$$\mathbf{v}^{\wedge} = \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_r$$

Moreover

$$C_r(A)\ \mathbf{v}^{\wedge} = (\mathrm{vol}\, A)\ \mathbf{u}^{\wedge}, \quad C_r(A^{\dagger})\ \mathbf{u}^{\wedge} = \frac{1}{\mathrm{vol}\, A}\ \mathbf{v}^{\wedge} \tag{31}$$

correspond to the facts that A is invertible as a mapping: $R(A^T) \to R(A)$, and $A^{\dagger}$ invertible as a mapping: $R(A) \to R(A^T)$, see also [12]. In particular,

$$C_r(A^{\dagger}) = (C_r(A))^{\dagger}, \quad \text{and } \mathrm{vol}\,(A^{\dagger}) = \frac{1}{\mathrm{vol}\, A}\ .$$

Results relating volumes, compound matrices and full rank factorizations are collected in the following lemma. The proofs are omitted.

Lemma 3.1 (Volume and compounds) Let $r > 0$, $A \in \mathbb{R}_r^{m\times n}$, $C \in \mathbb{R}_r^{m\times r}$ have columns $\mathbf{c}^{(j)}$ and $R \in \mathbb{R}_r^{r\times n}$ have rows $\mathbf{r}^{(i)}$. Then:

(a)
$$C_r(R)\ (\mathbf{r}_{(1)} \wedge \cdots \wedge \mathbf{r}_{(r)}) = \mathrm{vol}^{\,2} R \tag{32}$$

(b)
$$C_r(C^T)\ (\mathbf{c}^{(1)} \wedge \cdots \wedge \mathbf{c}^{(r)}) = \mathrm{vol}^{\,2} C \tag{33}$$

(c) If $A = CR$ is a full rank factorization (15) of A, then

$$C_r(A) = (\mathbf{c}^{(1)} \wedge \cdots \wedge \mathbf{c}^{(r)})\ (\mathbf{r}^{(1)} \wedge \cdots \wedge \mathbf{r}^{(r)}) \tag{34}$$

is a full rank factorization of $C_r(A)$,[2]. Moreover, the volume of A is given by the inner product,

$$(\mathbf{c}^{(1)} \wedge \cdots \wedge \mathbf{c}^{(r)}\ ,\ A\mathbf{r}_{(1)} \wedge \cdots \wedge A\mathbf{r}_{(r)}) = \mathrm{vol}^{\,2} A\ , \tag{35}$$

and

$$\mathrm{vol}_r^2 A = \mathrm{vol}_1^2\, C_r(A) = \mathrm{vol}_1^2 C_r(C)\ \mathrm{vol}_1^2\, C_r(R)\ . \quad \square \tag{36}$$

Example 3.1 Consider the 3×3 matrix of rank 2, with a full rank factorization

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = CR = \begin{pmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix} \tag{37}$$

[2] This is a restatement of (22).

Then the 2-compound matrix is

$$C_2(A) = \begin{pmatrix} -3 & -6 & -3 \\ -6 & -12 & -6 \\ -3 & -6 & -3 \end{pmatrix} = C_2(C)\, C_2(R) =$$

$$= \begin{pmatrix} -3 \\ -6 \\ -3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \tag{38}$$

illustrating (34). The volume of A is calculated by (36)

$$\mathrm{vol}_2^2\, A = \mathrm{vol}_1^2(C)\ \mathrm{vol}_1^2(R) = (9+36+9)(1+4+1) = 324.$$

Lemma 3.2 (Corresponding $r \times r$ submatrices of A and $A^\dagger$) Let $A \in \mathbb{R}_r^{m\times n}$, $r > 0$. Then the determinants of corresponding (in transposed position) $r \times r$ submatrices of A and $A^\dagger$ are proportional,

$$\det\,(A^\dagger)_{JI} = \frac{\det A_{IJ}}{\mathrm{vol}^{\,2} A},\quad \forall\ (I,J) \in \mathcal{N}\ . \tag{39}$$

Proof: From (34), (32) and (33) we calculate

$$C_r(A)^\dagger = \frac{1}{\mathrm{vol}^{\,2} A}\ (\mathbf{r}_{(1)} \wedge \cdots \wedge \mathbf{r}_{(r)})\ (\mathbf{c}^{(1)} \wedge \cdots \wedge \mathbf{c}^{(r)})\ . \tag{40}$$

We conclude that $\mathcal{N}(A^\dagger) = \mathcal{J} \times \mathcal{I}$, and (39) follows from (22).

4 Determinants

The volume (9) can be computed as a single determinant by appropriately modifying A.

Given $A \in \mathbb{R}_r^{m\times n}$, $r > 0$, there is a natural way to define nonsingular matrices which carry useful information about A and about linear equations in which A appears. Let $U_0 \in \mathbb{R}^{m\times(m-r)}$ and $V_0 \in \mathbb{R}^{n\times(n-r)}$ be any two matrices whose columns are orthonormal bases of $N(A^T)$ and $N(A)$ respectively. Define the **bordered matrix** $\mathbf{B}(A) \in \mathbb{R}^{(m+n-r)\times(m+n-r)}$ by

$$\mathbf{B}(A) = \begin{pmatrix} A & U_0 \\ V_0^T & O \end{pmatrix}\ . \tag{41}$$

Then $\mathbf{B}(A)$ is nonsingular ([5], [2, § 5.6]). For any $\mathbf{b} \in \mathbb{R}^m$ the **minimum-norm least-squares solution**

$$\mathbf{x} = A^\dagger\, \mathbf{b} \tag{42}$$

of

(2) $\qquad A\mathbf{x} = \mathbf{b}$,

is exactly the solution $\mathbf{x}$ of

$$\begin{pmatrix} A & U_0 \\ V_0^T & O \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}, \tag{43}$$

and the **residual**, $P_{N(A^T)}\,\mathbf{b}$, is given by $U_0\mathbf{u}$.

These results can be stated more compactly if $m = n$, [3]. Then the **complemented matrix**[4]

$$\mathbf{C}(A) = A + U_0 V_0^T \tag{44}$$

is nonsingular, and the solution of

$$(A + U_0 V_0^T)\mathbf{x} = \mathbf{b} \tag{45}$$

is again the minimum-norm least-squares solution of (2). The orthonormality of the columns of U_0 or V_0 was not needed until now.

Lemma 4.1

(a) Let $C \in \mathbb{R}_r^{m\times r}$ and let the columns of $U_0 \in \mathbb{R}^{m\times(m-r)}$ be an orthonormal basis of $N(C^T)$. Then the bordered matrix $\mathbf{B}(C) = (C \;\; U_0)$ satisfies

$$\det{}^2 \mathbf{B}(C) = \Delta_r^2(C) \ . \tag{46}$$

(b) Let $R \in \mathbb{R}_r^{r\times n}$ and let the columns of $V_0 \in \mathbb{R}^{n\times(n-r)}$ be an orthonormal basis of $N(R)$. Then the bordered matrix

$$\mathbf{B}(R) = \begin{pmatrix} R \\ V_0^T \end{pmatrix}$$

satisfies

$$\det{}^2 \mathbf{B}(R) = \Delta_r^2(R) \ . \tag{47}$$

Proof of (a): Follows from $U_0^T U_0 = I$, $\det^2 \mathbf{B}(C) = \det \mathbf{B}(C)^T \det \mathbf{B}(C)$ and

$$\mathbf{B}(C)^T \mathbf{B}(C) = \begin{pmatrix} C^T C & O \\ O & U_0^T U_0 \end{pmatrix} \ . \ \square$$

Lemma 4.2 Let $A \in \mathbb{R}_r^{m\times n}$, and let $U_0 \in \mathbb{R}^{m\times(m-r)}$ and $V_0 \in \mathbb{R}^{n\times(n-r)}$ be matrices whose columns are orthonormal bases of $N(A^T)$ and $N(A)$, respectively. Then:
(a) The m-dimensional volume of $(A \;\; U_0)$ equals the r-dimensional volume of A

$$\Delta_m^2\,(A \;\; U_0) = \Delta_r^2(A) \ . \tag{48}$$

(b) The n-dimensional volume of $\begin{pmatrix} A \\ V_0^T \end{pmatrix}$ equals the r-dimensional volume of A

$$\Delta_n^2 \begin{pmatrix} A \\ V_0^T \end{pmatrix} = \Delta_r^2(A) \ . \tag{49}$$

Proof of (a): Every $m \times m$ nonsingular submatrix of $(A \;\; U_0)$ is of the form

$$(A_{*J} \;\; U_0) \ , \quad J \in \mathcal{J} \ ,$$

and therefore, $\Delta_m^2\,(A_{*J} \;\; U_0) = \Delta_r^2(A_{*J})$, by Lemma 4.1(a). The proof is completed by (24). $\square$

[3] Which can be assumed, without loss of generality, since A can be padded with zeros.
[4] The factor $U_0 V_0^T$ is called **bicomplementary**, [18].

Theorem 4.1 (Bordered matrices) Let A, U_0 and V_0 be as in Lemma 4.2, and consider the bordered matrix (41). Then:

$$\begin{aligned}
\Delta_r^2(A) &= \det{}^2 \mathbf{B}(A) \ , && (50)\\
&= \det\left(AA^T + U_0U_0^T\right) \ , && (51)\\
&= \det\left(A^TA + V_0V_0^T\right) \ . && (52)
\end{aligned}$$

Proof: Since the columns of

$$\begin{pmatrix} V_0 \\ O \end{pmatrix}$$

form an o.n. basis of $N((A \;\; U_0)^T)$, we can use Lemma 4.1(b) and Lemma 4.2(a) to prove (50)

$$\det{}^2 \mathbf{B}(A) = \Delta_m^2\,(A \;\; U_0) = \Delta_r^2(A),$$

and (51)

$$\Delta_m^2\,(A \;\; U_0) = \det\left((A \;\; U_0)\,(A \;\; U_0)^T\right) = \det\left(AA^T + U_0U_0^T\right) \ . \quad \square$$

Theorem 4.2 (Complemented matrices) Let $A \in \mathbb{R}_r^{m\times n}$, $m = n$, and let U_0 and V_0 be as in Lemma 4.2. Then

$$\Delta_r^2(A) = \det{}^2\left(A + U_0V_0^T\right) \ . \qquad (53)$$

Proof: Follows from previous results

$$\Delta_r^2(A) = \Delta_r^2(C)\,\Delta_r^2(R) = \det{}^2\,(C\;U_0)\det{}^2\begin{pmatrix} R \\ V_0^T \end{pmatrix} =$$

$$= \det{}^2\left(CR + U_0V_0^T\right) = \det{}^2\left(A + U_0V_0^T\right) \ . \quad \square$$

Example 4.1 Let

$$A = \begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix} \quad \text{with } r = 1,\ \Delta_1^2(A) = 5$$

and select U_0, V_0 as in Lemma 4.2

$$U_0 = \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}, \quad V_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Then

$$\mathbf{C}(A) = A + U_0V_0^T = \begin{pmatrix} 1 & -\frac{2}{\sqrt{5}} \\ 2 & \frac{1}{\sqrt{5}} \end{pmatrix},$$

with determinant $\sqrt{5}$. Also

$$\mathbf{B}(A) = \begin{pmatrix} A & U_0 \\ V_0^T & O \end{pmatrix} = \begin{pmatrix} 1 & 0 & -\frac{2}{\sqrt{5}} \\ 2 & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \end{pmatrix},$$

with determinant $= -\sqrt{5}$. The 1-dimensional volume of A is thus equal to the 2-dimensional volume of the parallelepiped defined by the columns (or rows) of $\mathbf{C}(A)$, or to the 3-dimensional volume of the parallelepiped defined similarly by $\mathbf{B}(A)$.

Corollary 4.1 Let $A,\ B$ be $n \times n$ matrices. Then

$$\mathrm{vol}\,(AB) = \mathrm{vol}\,(A)\ \mathrm{vol}\,(B) \tag{54}$$

if the null spaces

$$N(A) = N(B^T),\ N(A^T) = N(B)\ . \tag{55}$$

Proof: From (55) it follows that rank A = rank $B = r$, say, and $N(B) = N(AB) = N((AB)^T)$. Let U_0 be a matrix whose columns are an o.n. basis of $N(B)$. Then

$$\begin{aligned}
\Delta_r^2(AB) &= \det{}^2\left(AB + U_0U_0^T\right)\ ,\ \text{ by Theorem 4.2,}\\
&= \det{}^2\begin{pmatrix} A & U_0 \\ V_0^T & O\end{pmatrix}\begin{pmatrix} B & V_0 \\ U_0^T & O\end{pmatrix}\ ,\\
&= \Delta_r^2(A)\Delta_r^2(B)\ .\ \square
\end{aligned}$$

5 Some applications

Example 5.1 (A generalized Hadamard inequality) Let $A \in \mathbb{R}_r^{m\times n}$ be partitioned into two matrices $A = (A_1, A_2),\ \ A_i \in \mathbb{R}_{r_i}^{m\times n_i},\ i = 1, 2$, with $r_1 + r_2 = r$. Then

$$\mathrm{vol}_r A \leq \mathrm{vol}_{r_1} A_1\ \mathrm{vol}_{r_2} A_2\ , \tag{56}$$

with equality iff the columns of A_1 are orthogonal to those of A_2.

Proof: The full-rank case $n_i = r_i,\ i = 1, 2$, was proved in [7, Vol. I, p. 254]. The general case follows since every $m \times r$ submatrix of rank r has r_1 columns from A_1 and r_2 columns from A_2. $\square$
A statement of (56) in terms of the principal angles ([1]) between $R(A_1)$ and $R(A_2)$ is given in [14].

Example 5.2 (Orthogonal projections) Let S be a subspace of $\mathbb{R}^n$ of dimension r, and let S be spanned by vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$. Let $\mathbf{v} \in \mathbb{R}^n$ be written as $\mathbf{v} = \mathbf{v}_S + \mathbf{v}_{S^\perp}$ where $\mathbf{v}_S \in S$ and $\mathbf{v}_{S^\perp}$ is orthogonal to S. Then

$$\|\mathbf{v}_{S^\perp}\| = \frac{\mathrm{vol}_{r+1}(\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{v})}{\mathrm{vol}_r(\mathbf{v}_1, \cdots, \mathbf{v}_k)}\ , \tag{57}$$

where $(\mathbf{v}_1, \cdots, \mathbf{v}_k)$ is the matrix with $\mathbf{v}_j$ as columns.

Proof: If $\mathbf{v} \in S$ then both sides of (57) are zero. If $\mathbf{v} \notin S$ then

$$\begin{aligned}
\Delta_{r+1}^2(\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{v}) &= \Delta_{r+1}^2(\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{v}_{S^\perp})\ ,\\
&\qquad \text{by properties of determinants,}\\
&= \Delta_r^2(\mathbf{v}_1, \cdots, \mathbf{v}_k)\ \Delta_1^2(\mathbf{v}_{S^\perp})\ ,\\
&\qquad \text{by Example 5.1,}
\end{aligned}$$

which completes the proof since $\Delta_1^2(\mathbf{v}_{S^\perp}) = \|\mathbf{v}_{S^\perp}\|^2$. $\square$
This generalizes a classical result given for the case $k = r$ in [7, Vol. I, Chapter IX, § 4].

Example 5.3 (**The volumes of the** $(n-1)$**-dimensional faces of a right** n**-simplex**)

If $A \in \mathbb{R}_n^{n \times n}$ has orthogonal columns $\{A_{*j} : \; j = 1, \cdots, n\}$ then the right n-simplex

$$S = \text{conv } \{\mathbf{0}, \{A_{*j} : j = 1, \cdots, n\}\}$$

has $n+1$ faces of dimension $n-1$: The face

$$F_0 = \text{conv } \{A_{*j} : j = 1, \cdots, n\} \ , \tag{58}$$

and the n simplices of dimension $n-1$

$$F_J = \text{conv } \{\mathbf{0}, \{A_{*j} : j \in J\}\}, \quad J \in Q_{n-1,n} \ . \tag{59}$$

Then:

$$\text{vol}_{n-1} \ F_J = \frac{\sqrt{\Delta_{n-1}^2 A_{*J}}}{(n-1)!}, \quad J \in Q_{n-1,n} \ , \tag{60}$$

$$\text{vol}_{n-1} \ F_0 = \frac{\sqrt{\Delta_{n-1}^2 A}}{(n-1)!} \ . \tag{61}$$

Proof: For any k, the volume of a right k-simplex conv $\{\mathbf{0}, \mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is related to the volume of the parallelepiped $\square(\{\mathbf{v}_1, \cdots, \mathbf{v}_k\})$ by

$$\frac{\text{vol } \square(\{\mathbf{v}_1, \cdots, \mathbf{v}_k\})}{k!} \ ,$$

which proves (60). Then (61) follows from the Pythagorean theorem of Lin and Lin [8]

$$\begin{aligned} \text{vol}_{n-1}^2 \ F_0 &= \sum_{J \in Q_{n-1,n}} \text{vol}_{n-1}^2 \ F_J \ , \\ &= \frac{1}{(n-1)!^2} \sum_{J \in Q_{n-1,n}} \Delta_{n-1}^2 A_{*J}, \quad \text{by (60)}, \\ &= \frac{\Delta_{n-1}^2 A}{(n-1)!^2} \ . \quad \square \end{aligned}$$

Example 5.4 (Height of a right n-simplex) Let A, S, F_0 be as in Example 5.3. Then the distance from the origin to the hyperplane containing the face F_0 is

$$\frac{1}{\sqrt{\displaystyle\sum_{j=1}^{n} \frac{1}{\|A_{*j}\|^2}}} \ . \tag{62}$$

Proof: Let h be the distance in question. Then the volume of the simplex S is

$$\frac{|\text{det} A|}{n!} = \frac{h \, \text{vol}_{n-1} \ F_0}{n} \ ,$$

and by (61),

$$h = \frac{|\text{det} A|}{\sqrt{\Delta_{n-1}^2 A}} \ ,$$

which equals (62) by the orthogonality of columns of A. $\square$

6 Least squares solutions and the Moore-Penrose inverse

Remark 6.1 (Volumes and Cramer's rule) Volume enters linear equations through Cramer's rule which can be stated in terms of ratios of volumes: Let $A \in \mathbb{R}_n^{n\times n}$, $\mathbf{b} \in \mathbb{R}^n$ and consider the linear equation $A\mathbf{x} = \mathbf{b}$. For any $j = 1, \cdots, n$ partition $A = \left(A^{(j)}, \bar{A}\right)$ where $A^{(j)}$ is the jth column, and $\bar{A}$ is the matrix of remaining columns. Similarly partition the solution as $\mathbf{x} = (x_j, \bar{\mathbf{x}})$, so that

$$A\mathbf{x} = A^{(j)}x_j + \bar{A}\bar{\mathbf{x}} = \mathbf{b} \ . \tag{63}$$

Let $A[j \leftarrow \mathbf{b}]$ denote the matrix A with the jth column replaced by $\mathbf{b}$. Then

$$\begin{aligned} \det A[j \leftarrow \mathbf{b}] &= \det A[j \leftarrow A^{(j)}x_j + \bar{A}\bar{\mathbf{x}}] \ , \\ &= \det A[j \leftarrow A^{(j)}x_j] \ , \\ &= x_j \det A[j \leftarrow A^{(j)}] \ , \\ &= x_j \det A \ , \end{aligned}$$

which is Cramer's rule. It states that x_j is the ratio of volumes of two parallelepipeds with same "basis" $\bar{A}$, but with "heights" $\mathbf{b}$ and $A^{(j)}$ respectively. If these parallelepipeds are on different sides of the hyperplane spanned by the columns of $\bar{A}$ then $x_j < 0$.

Remark 6.2 (**Separating the least squares and minimum norm computations**)

Let $A \in \mathbb{R}_r^{m\times n}$ and consider the equation $A\mathbf{x} = \mathbf{b}$. Its minimum-norm least-squares solution $\mathbf{x} = A^\dagger\mathbf{b}$ is literally the solution of a **two-stage minimization problem**:

Stage 1

$$\text{minimize } \|A\mathbf{x} - \mathbf{b}\| \tag{64}$$

Stage 2

$$\text{minimize } \{\|\mathbf{x}\| \text{ among all solutions of } \textbf{Stage 1}\} \tag{65}$$

where norms are Euclidean. Stage 1 (least squares) has a unique solution only if $r = n$. Stage 2 has the (unique) solution $\mathbf{x} = A^\dagger\mathbf{b}$, [5].

For any full-rank factorization $A = CR$ the above two stages can be separated

Stage 1

$$\text{minimize } \|C\mathbf{y} - \mathbf{b}\| \tag{66}$$

Stage 2

$$\text{minimize } \{\|\mathbf{x}\| \text{ among all solutions of } R\mathbf{x} = \mathbf{y}\} \tag{67}$$

with the advantage that Stage 1 now has the unique solution $\mathbf{y} = C^\dagger\mathbf{b}$. This is an implementation of the fact that

$$A^\dagger = R^\dagger C^\dagger \tag{68}$$

is a full-rank factorization of $A^\dagger$.

[5] These two stages can be combined (in the limit):

$$\text{minimize } \|A\mathbf{x} - \mathbf{b}\|^2 + \alpha^2\|\mathbf{x}\|^2 \ ,$$

where $\alpha \to 0$.

Lemma 6.1 (Solution of full-rank systems)

(a) Let $C \in \mathbb{R}_r^{m\times r}$, $\mathbf{b} \in \mathbb{R}^m$. Then the (unique) least-squares solution y of

$$C\mathbf{y} = \mathbf{b} \ , \tag{69}$$

is

$$\mathbf{y} = \sum_{I\in\mathcal{I}(C)} \mu_{I*} \ C_{I*}^{-1}\mathbf{b}_I \ , \tag{70}$$

where μ_{I*} is given by

$$\mu_{I*} = \frac{\mathrm{vol}^2 C_{I*}}{\mathrm{vol}^2 C} \ . \tag{71}$$

(b) Let $R \in \mathbb{R}_r^{r\times n}$, $\mathbf{y} \in \mathbb{R}^r$. Then the minimum norm solution of

$$R\mathbf{x} = \mathbf{y} \ , \tag{72}$$

is

$$\mathbf{x} = \sum_{J\in\mathcal{J}(R)} \nu_{*J} \ R_{*J}^{-1}\mathbf{y} \ , \tag{73}$$

where ν_{*J} is given by

$$\nu_{*J} = \frac{\mathrm{vol}^2 R_{*J}}{\mathrm{vol}^2 R} \ . \tag{74}$$

Proof: (a) The coefficients y_i satisfy the **normal equation** $C^TC\mathbf{y} = C^T\mathbf{b}$, rewritten as,

$$\begin{aligned} C^T\mathbf{c}^{(1)} \wedge \cdots \wedge C^T\mathbf{c}^{(i-1)} \wedge C^T\mathbf{b} \wedge C^T\mathbf{c}^{(i+1)} \wedge \cdots \wedge C^T\mathbf{c}^{(r)} = \\ = y_i \ (C^T\mathbf{c}^{(1)} \wedge \cdots \wedge C^T\mathbf{c}^{(r)}) \ . \end{aligned} \tag{75}$$

The LHS is $C_r(C^T) \ (\mathbf{c}^{(1)} \wedge \cdots \wedge \mathbf{c}^{(i-1)} \wedge \mathbf{b} \wedge \mathbf{c}^{(i+1)} \wedge \cdots \wedge \mathbf{c}^{(r)})$ which simplifies to

$$\begin{aligned} \text{LHS (75)} &= \sum_{I\in\mathcal{I}(C)} \det C_{I*} \ \det C_{I*}[i \leftarrow \mathbf{b}_I] = \\ &= \sum_{I\in\mathcal{I}(C)} \det{}^2 C_{I*} \ (C_{I*}^{-1}\mathbf{b}_I)_i \end{aligned}$$

and RHS (75) is y_i times (33). The Cramer rule for the least squares solution is therefore

$$y_i = \sum_{I\in\mathcal{I}(C)} \mu_{I*} \ (C_{I*}^{-1}\mathbf{b}_I)_i \ , \tag{76}$$

with μ_{I*} given by (71), and $(C_{I*}^{-1}\mathbf{b}_I)_i$ the i^{th} component of the solution $C_{I*}^{-1}\mathbf{b}_I$ of the $r \times r$ system

$$C_{I*} \ \mathbf{y} = \mathbf{b}_I \ . \tag{77}$$

Combining (76) for $i = 1, \ldots, r$, we obtain the least squares solution $\mathbf{y}$ as the convex combination (70) of "basic solutions"[6] □

Combining Remark 6.2 and Lemma 6.1, we prove Berg's Theorem in geometric form.

[6]This derivation follows that of Marcus [10, § 3.1, Example 1.5(c)] and Ben-Tal and Teboulle [3].

Theorem 6.1 (Berg, [4]) Let $A \in \mathbb{R}_r^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. Then the minimum-norm least-squares solution of

(2) $\qquad A\mathbf{x} = \mathbf{b}$,

is the convex combination

$$\mathbf{x} = \sum_{(I,J)\in\mathcal{N}(A)} \lambda_{IJ}\, \widehat{A_{IJ}^{-1}\mathbf{b}}_I \ , \tag{78}$$

with weights given by (4).

Proof: Follows by substituting (70) in the RHS of (72). Then (78) follows from (73) with weights

$$\lambda_{IJ} = \mu_{I*}\,\nu_{*J}$$

which, by (71) and (74), are (4). □

Since (78) holds for all $\mathbf{b}$, we proved Berg's representation (1) of the Moore-Penrose inverse as a convex combination of ordinary inverses of $r \times r$ submatrices,

(1) $\qquad A^\dagger = \sum_{(I,J)\in\mathcal{N}} \lambda_{IJ}\ \widehat{A_{IJ}^{-1}}$,

where $\widehat{A_{IJ}^{-1}}$ is an $n \times m$ matrix with the inverse of A_{IJ} in position (J, I), and zeros elsewhere.

Remark 6.3 Other representations of $A^\dagger$, e.g. [6], can be obtained by summing (1) in special ways.
Summing (1) over $I \in \mathcal{I}$ we obtain, using (21) and (22),

$$A^\dagger = \sum_{J\in\mathcal{J}} \lambda_{*J}\ \widehat{A_{*J}^\dagger} \ , \tag{79}$$

a convex combination of the Moore-Penrose inverses of maximal full column-rank submatrices A_{*J}, with weights λ_{*J}

$$\lambda_{*J} = \frac{\operatorname{vol}^2 A_{*J}}{\operatorname{vol}^2 A} \ , \tag{80}$$

and $\widehat{A_{*J}^\dagger}$ is the $n \times m$ matrix with $A_{*J}^\dagger$ in rows J and zeros elsewhere.
Similarly, summing (1) over $J \in \mathcal{J}$ gives

$$A^\dagger = \sum_{I\in\mathcal{I}} \lambda_{I*}\ \widehat{A_{I*}^\dagger} \ , \tag{81}$$

where

$$\lambda_{I*} = \frac{\operatorname{vol}^2 A_{I*}}{\operatorname{vol}^2 A} \ , \tag{82}$$

and $\widehat{A_{I*}^\dagger}$ is an $n \times m$ matrix with $A_{I*}^\dagger$ in columns I and zeros elsewhere.

Example 6.1 Consider A of Example 3.1. We illustrate (81) where $A^\dagger$ is represented as a convex combination of the $\widehat{A_{I*}^\dagger}$, $I \in \mathcal{I}$. Each $A_{I*}^\dagger$ is computed using (1). This requires a listing of the 2×2 submatrices and their volumes and inverses as follows:

I	J	A_{IJ}	$\det^2 A_{IJ}$	A_{IJ}^{-1}
$1,2$	$1,2$	$\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$	3^2	$\frac{1}{3}\begin{pmatrix} -5 & 2 \\ 4 & -1 \end{pmatrix}$
$1,2$	$1,3$	$\begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix}$	6^2	$\frac{1}{6}\begin{pmatrix} -6 & 3 \\ 4 & -1 \end{pmatrix}$
$1,2$	$2,3$	$\begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix}$	3^2	$\frac{1}{3}\begin{pmatrix} -6 & 3 \\ 5 & -2 \end{pmatrix}$
$1,3$	$1,2$	$\begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}$	6^2	$\frac{1}{6}\begin{pmatrix} -8 & 2 \\ 7 & -1 \end{pmatrix}$
$1,3$	$1,3$	$\begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}$	12^2	$\frac{1}{12}\begin{pmatrix} -9 & 3 \\ 7 & -1 \end{pmatrix}$
$1,3$	$2,3$	$\begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix}$	6^2	$\frac{1}{6}\begin{pmatrix} -9 & 3 \\ 8 & -2 \end{pmatrix}$
$2,3$	$1,2$	$\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}$	3^2	$\frac{1}{3}\begin{pmatrix} -8 & 5 \\ 7 & -4 \end{pmatrix}$
$2,3$	$1,3$	$\begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}$	6^2	$\frac{1}{6}\begin{pmatrix} -9 & 6 \\ 7 & -4 \end{pmatrix}$
$2,3$	$2,3$	$\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}$	3^2	$\frac{1}{3}\begin{pmatrix} -9 & 6 \\ 8 & -5 \end{pmatrix}$

Table 1: Illustration of Example 6.1.

For $A_{\{1,2\},*} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ we calculate

$$\mathrm{vol}_2^2 A_{\{1,2\},*} = 3^2 + 6^2 + 3^2 = 54 \ ,$$

and using (1), and the first three entries in Table 1,

$$A_{\{1,2\},*}^{\dagger} = \frac{3^2}{54}\,\frac{1}{3}\begin{pmatrix} -5 & 2 \\ 4 & -1 \\ 0 & 0 \end{pmatrix} + \frac{6^2}{54}\,\frac{1}{6}\begin{pmatrix} -6 & 3 \\ 0 & 0 \\ 4 & -1 \end{pmatrix} +$$

$$+\frac{3^2}{54}\,\frac{1}{3}\begin{pmatrix} 0 & 0 \\ -6 & 3 \\ 5 & -2 \end{pmatrix} \ , = \frac{3}{54}\begin{pmatrix} -17 & 8 \\ -2 & 2 \\ 13 & -4 \end{pmatrix} \ .$$

Similarly we calculate

$$A^{\dagger}_{\{1,3\},*} = \frac{6}{216} \begin{pmatrix} -26 & 8 \\ -2 & 2 \\ 22 & -4 \end{pmatrix} ,$$

$$A^{\dagger}_{\{2,3\},*} = \frac{3}{54} \begin{pmatrix} -26 & 17 \\ -2 & 2 \\ 22 & -13 \end{pmatrix} .$$

Finally, from (81),

$$A^{\dagger} = \frac{54}{324}\,\frac{3}{54} \begin{pmatrix} -17 & 8 & 0 \\ -2 & 2 & 0 \\ 13 & -4 & 0 \end{pmatrix} + \frac{216}{324}\,\frac{6}{216} \begin{pmatrix} -26 & 0 & 8 \\ -2 & 0 & 2 \\ 22 & 0 & -4 \end{pmatrix} +$$
$$+ \frac{54}{324}\,\frac{3}{54} \begin{pmatrix} 0 & -26 & 17 \\ 0 & -2 & 2 \\ 0 & 22 & -13 \end{pmatrix} , = \frac{1}{36} \begin{pmatrix} -23 & -6 & 11 \\ -2 & 0 & 2 \\ 19 & 6 & -7 \end{pmatrix} .$$

Remark 6.4 (Weighted least squares)

Consider a **weighted** (or **scaled**) system

$$W A \mathbf{x} = W \mathbf{b} , \tag{83}$$

where W is a diagonal matrix with positive diagonal elements w_i. If $A = CR$ is a full rank factorization of A, then $WA = (WC)R$ is a full rank factorization of WA. The first stage (66) for the problem (83) is

$$\text{minimize } \|WC\mathbf{y} - W\mathbf{b}\| , \tag{84}$$

whose solution, using Lemma 6.1(a), is

$$\mathbf{y} = \sum_{I \in \mathcal{I}} \mu_{I*}(W)\; C_{I*}^{-1} \mathbf{b}_I , \tag{85}$$

with

$$\mu_{I*}(W) = \frac{\left(\prod_{i \in I} w_i^2\right) \det^2 C_{I*}}{\sum\limits_{K \in \mathcal{I}} \left(\prod_{i \in K} w_i^2\right) \det^2 C_{K*}} . \tag{86}$$

The second stage is still (67)

$$\text{minimize } \{ \|\mathbf{x}\| : R\mathbf{x} = \mathbf{y} \} ,$$

with $\mathbf{y}$ from (85). Therefore the minimum norm (weighted) least squares solution of (83) is

$$\mathbf{x} = \sum_{(I,J) \in \mathcal{N}} \lambda_{IJ}(W)\; \widehat{A_{IJ}^{-1}} \mathbf{b}_I , \tag{87}$$

with

$$\lambda_{IJ}(W) = \frac{\left(\prod_{i \in I} w_i^2\right) \det^2 A_{IJ}}{\sum\limits_{(K,L) \in \mathcal{N}} \left(\prod_{i \in K} w_i^2\right) \det^2 A_{KL}} . \tag{88}$$

Note that the weights w_i appear only in the convex weights λ_{IJ} and not in the "basic solutions" $\{\widehat{A_{IJ}^{-1}\mathbf{b}}_I : (I,J) \in \mathcal{N}\}$. We conclude that for fixed A and $\mathbf{b}$, all weighted systems (83) have their minimum-norm least-squares solutions in the convex hull of the set of basic solutions. In the full-rank case this was proved by Ben-Tal and Teboulle in [3], together with extensions from least squares to minimizing isotone functions of $|A\mathbf{x} - \mathbf{b}|$, the vector of absolute values of $A\mathbf{x} - \mathbf{b}$.

J	I	$A^\dagger_{JI}$	$(A^\dagger_{JI})^{-1}$
$1,2$	$1,2$	$\frac{1}{36}\begin{pmatrix} -23 & -6 \\ -2 & 0 \end{pmatrix}$	$\frac{36}{12}\begin{pmatrix} 0 & -6 \\ -2 & 23 \end{pmatrix}$
$1,2$	$1,3$	$\frac{1}{36}\begin{pmatrix} -23 & 11 \\ -2 & 2 \end{pmatrix}$	$\frac{36}{24}\begin{pmatrix} -2 & 11 \\ -2 & 23 \end{pmatrix}$
$1,2$	$2,3$	$\frac{1}{36}\begin{pmatrix} -6 & 11 \\ 0 & 2 \end{pmatrix}$	$\frac{36}{12}\begin{pmatrix} -2 & 11 \\ 0 & 6 \end{pmatrix}$
$1,3$	$1,2$	$\frac{1}{36}\begin{pmatrix} -23 & -6 \\ 19 & 6 \end{pmatrix}$	$\frac{36}{24}\begin{pmatrix} -6 & -6 \\ 19 & 23 \end{pmatrix}$
$1,3$	$1,3$	$\frac{1}{36}\begin{pmatrix} -23 & 11 \\ 19 & -7 \end{pmatrix}$	$\frac{36}{48}\begin{pmatrix} 7 & 11 \\ 19 & 23 \end{pmatrix}$
$1,3$	$2,3$	$\frac{1}{36}\begin{pmatrix} -6 & 11 \\ 6 & -7 \end{pmatrix}$	$\frac{36}{24}\begin{pmatrix} 7 & 11 \\ 6 & 6 \end{pmatrix}$
$2,3$	$1,2$	$\frac{1}{36}\begin{pmatrix} -2 & 0 \\ 19 & 6 \end{pmatrix}$	$\frac{36}{12}\begin{pmatrix} -6 & 0 \\ 19 & 2 \end{pmatrix}$
$2,3$	$1,3$	$\frac{1}{36}\begin{pmatrix} -2 & 2 \\ 19 & -7 \end{pmatrix}$	$\frac{36}{24}\begin{pmatrix} 7 & 2 \\ 19 & 2 \end{pmatrix}$
$2,3$	$2,3$	$\frac{1}{36}\begin{pmatrix} 0 & 2 \\ 6 & -7 \end{pmatrix}$	$\frac{36}{12}\begin{pmatrix} 7 & 2 \\ 6 & 0 \end{pmatrix}$

Table 2: Illustration of Example 6.2.

We began with Berg's representation (1) of $A^\dagger$ as a convex combination of inverses of maximal nonsingular submatrices A_{IJ} of A. The weights λ_{IJ} of this convex combination are proportional to $\det^2 A_{IJ}$. These results can be redirected at A, since $A = A^{\dagger\,\dagger}$, representing it as convex combination of ordinary inverses of maximal nonsingular submatrices $(A^\dagger)_{JI}$ of $A^\dagger$. From (39) it follows that the convex weights are again λ_{IJ}. This proves the following.

Theorem 6.2 (Convex decomposition of matrices) Let $A \in \mathbb{R}^{m\times n}_r,\ r > 0$. Then

$$A = \sum_{(I,J)\in\mathcal{N}} \lambda_{IJ}\, \widehat{(A^\dagger)^{-1}_{JI}} \tag{89}$$

where $\widehat{(A^\dagger)^{-1}_{JI}}$ is an $m \times n$ matrix with the inverse of the (J, I)th submatrix of $A^\dagger$ in position (I, J) and zeros elsewhere, and

$$\lambda_{IJ} = \frac{\det^2 A_{IJ}}{\sum\limits_{(K,L)\in\mathcal{N}} \det^2 A_{KL}}\ . \ \square$$

Interestingly, the minors of all sizes of A and $A^\dagger$ have convex decompositions analogous to (89) and (1), with weights (4), [13].

Example 6.2 We illustrate (89) for the matrix A of Example 3.1, using the Moore-Penrose inverse from Example 6.1,

$$A^\dagger = \frac{1}{36}\begin{pmatrix} -23 & -6 & 11 \\ -2 & 0 & 2 \\ 19 & 6 & -7 \end{pmatrix} .$$

A listing of the 2×2 submatrices of $A^\dagger$ is given in Table 2. Their determinants are given by (39) and Example 6.1. Then (89) reconstructs A as a convex combination of the above inverses

$$A = \frac{3^2}{324}\,\frac{36}{12}\begin{pmatrix} 0 & -6 & 0 \\ -2 & 23 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{6^2}{324}\,\frac{36}{24}\begin{pmatrix} -2 & 11 & 0 \\ 0 & 0 & 0 \\ -2 & 23 & 0 \end{pmatrix} +$$

$$+\frac{3^2}{324}\,\frac{36}{12}\begin{pmatrix} 0 & 0 & 0 \\ -2 & 11 & 0 \\ 0 & 6 & 0 \end{pmatrix} + \frac{6^2}{324}\,\frac{36}{24}\begin{pmatrix} -6 & 0 & -6 \\ 19 & 0 & 23 \\ 0 & 0 & 0 \end{pmatrix} +$$

$$+\frac{12^2}{324}\,\frac{36}{48}\begin{pmatrix} 7 & 0 & 11 \\ 0 & 0 & 0 \\ 19 & 0 & 23 \end{pmatrix} + \frac{6^2}{324}\,\frac{36}{24}\begin{pmatrix} 0 & 0 & 0 \\ 7 & 0 & 11 \\ 6 & 0 & 6 \end{pmatrix} +$$

$$+\frac{3^2}{324}\,\frac{36}{12}\begin{pmatrix} 0 & -6 & 0 \\ 0 & 19 & 2 \\ 0 & 0 & 0 \end{pmatrix} + \frac{6^2}{324}\,\frac{36}{24}\begin{pmatrix} 0 & 7 & 2 \\ 0 & 0 & 0 \\ 0 & 19 & 2 \end{pmatrix} +$$

$$+\frac{3^2}{324}\,\frac{36}{12}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 7 & 2 \\ 0 & 6 & 0 \end{pmatrix} .$$

References

[1] S. N. Afriat, "Orthogonal and oblique projectors and the characteristics of pairs of vector spaces", *Proc. Cambridge Phil. Soc.* **53** (1957), 800-816

[2] A. Ben-Israel and T.N.E. Greville, *Generalized Inverses: Theory and Applications*, Wiley-Interscience, 1974

[3] A. Ben-Tal and M. Teboulle, "A geometric property of the least squares solution of linear equations", *Lin. Algeb. Appl.* **139**(1990), 165-170.

[4] L. Berg, "Three results in connection with inverse matrices", *Lin. Algeb. Appl.* **84**(1986), 63-77

[5] J.W. Blattner, "Bordered matrices", *J. Soc. Indust. Appl. Math.* **10**(1962), 528-536

[6] J.T. Bruening, "A new formula for the Moore-Penrose inverse", pp. 65-74 in *Current Trends in Matrix Theory* (F. Uhlig and R. Grone, editors), North-Holland, 1987

[7] F.R. Gantmacher, *The Theory of Matrices* Vols. I, II, Chelsea, New York, 1959

[8] S.-Y. T. Lin and Y.-F. Lin, "The n-dimensional Pythagorean theorem", *Lin. and Multilin. Algeb.* **26**(1990), 9-13

[9] L. Lovász, *Combinatorial Problems and Exercises*, North-Holland, Amsterdam, 1979

[10] M. Marcus, *Finite Dimensional Multilinear Algebra*, Part 1, Marcel Dekker, New York, 1973

[11] M. Marcus, *Finite Dimensional Multilinear Algebra*, Part 2, Marcel Dekker, New York, 1975

[12] M. Marcus, "A unified exposition of some classical matrix theorems", *Lin. and Multilin. Algeb.* **25**(1989), 137-147

[13] J. Miao and A. Ben-Israel, "Minors of the Moore-Penrose inverse", *Lin. Algeb. and its Appl.* **195**(1993), 191-208

[14] J. Miao and A. Ben-Israel, "On principal angles between subspaces in $\mathbb{R}^n$ ", *Lin. Algeb. Appl.* **171**(1992), 81-98

[15] G.D. Mostow, J.H. Sampson and J.-P. Meyer, *Fundamental Structures of Algebra*, McGraw-Hill, New York, 1963

[16] J. Springer, "Die exacte Berechnung der Moore-Penrose Inversen einer Matrix durch Residuenarithmetik", *Zeit. Angew. Math. Mech.* **63**(1983), 203-210

[17] X. Tang and A. Ben-Israel, "Two consequences of Minkowski's 2^n Theorem", *Discrete Mathematics* **169**(1997), 279–282

[18] H.J. Werner, "On extensions of Cramer's rule for solutions of restricted linear systems", *Lin. & Multilin. Algeb.* **15**(1984), 319-330