Università di Pisa
Dipartimento di Informatica

Technical Report: TR-05-09

Generalized inverses of band matrices

R. Bevilacqua E. Bozzo

March 8, 2005

Address: via F. Buonarroti 2, 56127 Pisa, Italy. Tel: +39 050 2212700 Fax: +39 050 2212726

Generalized inverses of band matrices

R. Bevilacqua E. Bozzo

March 8, 2005

Abstract

Edgar Asplund in [1] presented a geometric lemma that largely anticipated many results on the structure of inverses of band matrices. In this note we discuss an extension of Asplund's lemma that addresses the concept of generalized inverse, in particular the Moore-Penrose inverse.

1 Introduction and motivations

The paper [1] of Edgar Asplund contains the first result on the structure of the inverse of a band matrix without assumptions on its elements. The strength of the paper, in our opinion, is in its abstract, geometric formulation. Actually, the importance of the work of Asplund is recognized in [5] where relations between rank properties of certain subblocks of a matrix and its inverse are reviewed. These properties play a fundamental role in the work of various authors, aiming at a precise definition of the concept of rank structure, see for example [3, 4].

Generalized inverses have important applications in matrix theory, see [2]. The (perhaps) best known generalized inverse, named after Moore and Penrose, is strictly related to the problem of the solution, in the least squares sense, of an overdetermined linear system, see again [2].

In this note we extend Asplund's results to the Moore-Penrose inverse of a band matrix, since an extensive search did not allow us to find any result in this direction. We try to follow [1] in the spirit and in the form. For this reason, we do not pursue here any application or specific inversion formula. Both of them will be shown elsewhere.

2 A geometric lemma

Let V be a finite dimensional vector space over the real or complex field, and let $f : V \to V$ be a linear mapping. We denote with f^+ its Moore-Penrose inverse, for the definition see [2]. As well known, if f is nonsingular $f^+ = f^{-1}$. More generally, let $\mathrm{null}(f)$ be the nullspace of f, and let $H = \mathrm{rank}\,(f^+)$. The definition of the operator f^+ implies that H is the orthogonal complement of $\mathrm{null}(f)$ so that $V = \mathrm{null}(f) \oplus H$. Here, as customary, $\oplus$ denotes the direct sum of subspaces. Analogously, $K = \mathrm{rank}\,(f)$ is the orthogonal complement of null(f^+). If h denotes the projection on H along null(f) it is well known that $f^+ \circ f = h$. At this point, let R and S, T and U be two couples of complementary subspaces of V, so that

$$V = R \oplus S = T \oplus U.$$

Let r indicate the projection on R along S, s indicate the projection on S along R and let t and u have analogous meanings. In this setting Asplund proves an elegant result, that here we restate by using the notation $|\cdot|$ to indicate the dimension of a linear space.

Lemma 1 *If f is nonsingular then $|tf(S)| = 0$ if and only if*

$$|rf^{-1}(U)| \leq |R| - |T|.$$

In order to drop the hypothesis of nonsingularity we need to separate the two implications. We begin with the "only if" part of the lemma.

Lemma 2 *If $|tf(S)| = 0$ then*

$$|rf^+(U)| \leq |U| - |S \cap H|.$$

Proof. The assumption $|tf(S)| = 0$ implies $f(S) \subseteq U$. Now let us consider a basis in $f(S)$ and let us extend it to a basis of U. Since $f^+(f(S)) = h(S)$ we obtain

$$|rf^+(U)| \leq |U| - |f(S)| + |rh(S)|. \tag{1}$$

Obviously $|f(S)| = |S| - |S \cap \text{null}(f)|$. Analogously $|h(S)| = |S| - |S \cap \text{null}(h)| = |S| - |S \cap \text{null}(f)|$, since $\text{null}(h) = \text{null}(f)$. Thus $|rh(S)| = |h(S)| - |h(S) \cap S| = |S| - |S \cap \text{null}(f)| - |S \cap H|$, since $h(S) \cap S = H \cap S$. This leads to the thesis. ∎

The main difference with the nonsingular case is that here we have to consider $h(S)$ instead of S. Let us observe that if f is nonsingular then $|U| - |S \cap H| = |U| - |S| = |R| - |T|$. More generally, since H and null$f$ are complementary $|S| - |S \cap H| \leq |\text{null}(f)|$ and this implies a useful inequality that helps to compare the two lemmas 1 and 2

$$|U| - |S \cap H| \leq |U| - (|S| - |\text{null}(f)|) = |R| - |T| + |\text{null}(f)|.$$

We now extend the second part of Asplund's result.

Lemma 3 *If*

$$|rf^+(U \cap K)| \leq |U \cap K| - |S \cap H|$$

then $|tf(S \cap H)| = 0$.

Proof. Clearly $f^+(U) \subseteq H$ and since $S \cap H$ is the nullspace of r over H then it turns out that $S \cap H \subseteq f^+(U \cap K)$ so that $f(S \cap H) \subseteq U \cap K$ leading to the thesis. ∎

3 Band matrices and Green's Matrices

We now turn to band matrices and their (generalized) inverses. Our definitions are closely related to Asplund's ones.

Definition 1 *A band matrix of order p is a square matrix $A = (a_{ij})$ over the real or complex field whose elements satisfy $a_{ij} = 0$ for $j > i + p$. If in addition $a_{ij} \neq 0$ for $j = i + p$ the band matrix is said to be strict.*

Definition 2 *A (p, q) Green's matrix is a square matrix $A = (a_{ij})$ over the real or complex field whose submatrices have rank $\leq q$ if their elements belong to the part of (a_{ij}) for which $j + p > i$.*

Now we follow both the statement and the proof of one of the implications Theorem 1 in [1].

Theorem 1 *If A is band matrix of order $p \geq 0$ then A^+ is a $(p, p + |\mathrm{null}(A)|)$ Green's matrix.*

Proof. Let $|V| = n$ and let A be the representation of a linear mapping f with respect to the basis vectors b_i where $i = 1 : n$. Take some submatrix of A^+ whose elements are all from the part of the matrix where $j + p > i$. For some q in the range $0 : (n-p)$ there must be some submatrix $A^+(1 : (p+q), (q+1) : n)$ including the given submatrix. We want to prove that its rank is less or equal to $p + |\mathrm{null}(A)|$. Let R, S, T, U be the subspaces of V spanned by the vectors b_i for $i = 1 : (p+q)$, $i = (p+q+1) : n$, $i = 1 : q$, $i = (q+1) : n$ respectively (we agree that the empty set spans the trivial subspace $\{0\}$). Clearly $|tf(S)| = 0$. Thus we can apply lemma 2 to get the thesis. ■

Asplund noticed that when strictness is assumed a simple description of the inverse can be obtained involving a matrix of low rank. An analogous result holds in general. We need a lemma from matrix theory.

Lemma 4 *If A and X are square matrices then $\mathrm{rank}(X) \leq 2\,|\mathrm{null}(A)| + \mathrm{rank}(AXA)$.*

Proof. Given two square matrices A and B it is known and simple to prove that $\mathrm{rank}(AB) \geq \mathrm{rank}(B) - |\mathrm{null}(A)|$ and analogously $\mathrm{rank}(AB) \geq \mathrm{rank}(A) - |\mathrm{null}(B)|$. Then

$$\mathrm{rank}(AXA) \geq \mathrm{rank}(AX) - |\mathrm{null}(A)| \geq \mathrm{rank}(X) - 2\,|\mathrm{null}(A)|.$$

■

The promised result follows.

Theorem 2 *If A is a strict band matrix of order $p \geq 0$ then its Moore-Penrose inverse is the sum of a strict band matrix of order $-p$ and a matrix of rank $p + |\mathrm{null}(A)|$ (and hence also a $(p, p + |\mathrm{null}(A)|)$ Green's matrix).*

Proof. If A is a strict band matrix of order $p \geq 0$ then it can be represented in the form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{12} is a $(n-p)$-square invertible lower triangular matrix. We want to show that the matrix

$$A^+ - \begin{pmatrix} 0 & 0 \\ A_{12}^{-1} & 0 \end{pmatrix}$$

has rank at most $p + |\text{null}(A)|$ since this clearly implies the thesis.

Since $AA^+A = A$ one easily verifies that

$$A\left(A^+ - \begin{pmatrix} 0 & 0 \\ A_{12}^{-1} & 0 \end{pmatrix}\right)A = \begin{pmatrix} 0 & 0 \\ S & 0 \end{pmatrix}$$

where $S = A_{21} - A_{22}A_{12}^{-1}A_{11}$ is the Schur complement of A_{12} in A. As well known $|\text{null}(S)| = |\text{null}(A)|$ so that $\text{rank}(S) = p - |\text{null}(S)| = p - |\text{null}(A)|$. By using lemma 4 we obtain what claimed. ∎

4 Extensions

With the due modifications to the definitions and the notations, all the results proved so far hold if we consider instead of the Moore-Penrose inverse of f any reflexive type g-inverse of f, i.e., a linear application g such that $f \circ g \circ f = f$ and $g \circ f \circ g = g$. Actually the orthogonality between the range of g and the nullspace of f which is the distictive feature of the Moore-Penrose inverse, does not play any active role in the proofs of Section 1.

Since the concept of orthogonality is not directly involved, the proofs in Section 2 hold for matrices in arbitrary fields. Their validity in more abstract settings is under consideration.

References

[1] E. Asplund, Inverses of matrices $\{a_{ij}\}$ which satisfy $a_{ij} = 0$ for $j > i + p$, Math. Scand. **7** (1959), 57-60.

[2] A. Ben-Israel, T. N. E. Greville, Generalized inverses: theory and applications, 2nd ed., Springer, New York (2003).

[3] Y. Eidelman, I. Gohberg, On a new class of structured matrices, Integral Eq. and Oper. Theory, **34** (1999), 293-324.

[4] M. Fiedler, Structure ranks of matrices, Linear Alg. Appl., **179** (1993), 119-127.

[5] G. Strang, T. Nguyen, The interplay of ranks of submatrices, SIAM Rev., **46** (2005), 637-646.