The Moore–Penrose Generalized Inverse for Sums of Matrices

James Allen Fill* and
Donniell E. Fishkind†

September 18, 1998

Short title: Moore–Penrose Generalized Inverse for Sums

Abstract

In this paper we exhibit, under suitable conditions, a neat relationship between the Moore–Penrose generalized inverse of a sum of two matrices and the Moore–Penrose generalized inverses of the individual terms. We include an application to the parallel sum of matrices.

AMS 1991 *subject classifications.* Primary 15A09; secondary 15A18.

Key words and phrases. Moore–Penrose generalized inverse, Sherman–Morrison–Woodbury formula, singular value decomposition, rank additivity, parallel sum.

*Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, Maryland 21218-2682, jimfill@jhu.edu. Research supported by NSF grants DMS–96–26756 and DMS–98–03780.

†Department of Mathematics and Statistics, University of Southern Maine, Portland, Maine 04104, fishkind@payson.usm.maine.edu. Research carried out while this author was affiliated with The Johns Hopkins University, and supported by US Department of Education GAANN grant P200A40303.

1 Background and Main Result

In the late 1940s and the 1950s Sherman and Morrison [11] [12], Woodbury [13], Bartlett [2], and Bodewig [4] discovered the following result. As in [7], $M_{m,n}$ denotes the space of complex-valued $m \times n$ matrices and, when $m = n$, this is shortened to M_n.

Theorem 1 (Sherman–Morrison–Woodbury) *For $s \leq n$, let $A \in M_n$ and $G \in M_s$ both be invertible, and let $Y, Z \in M_{n,s}$. Then $A + YGZ^*$ is invertible if and only if $G^{-1} + Z^*A^{-1}Y$ is invertible, in which case*

$$(A + YGZ^*)^{-1} = A^{-1} - A^{-1}Y(G^{-1} + Z^*A^{-1}Y)^{-1}Z^*A^{-1}.$$

The Sherman–Morrison–Woodbury (SMW) formula and related formulas are reviewed in Henderson and Searle [6]. The SMW formula has been used in a wide variety of applications; an excellent review by Hager [5] describes some of the applications to statistics, networks, structural analysis, asymptotic analysis, optimization, and partial differential equations.

In 1992, Riedel [10] proved an analogous formula (Theorem 2) for some cases where A is singular. All matrices, including singular and even nonsquare matrices, have a Moore–Penrose generalized inverse. Given a matrix $A \in M_{m,n}$, the *Moore–Penrose generalized inverse* of A, denoted $A^\dagger$, is the unique matrix in $M_{n,m}$ satisfying the conditions

$$
\begin{aligned}
AA^\dagger A &= A, & (1)\\
A^\dagger AA^\dagger &= A^\dagger, & (2)\\
&AA^\dagger \text{ is Hermitian, and} & (3)\\
&A^\dagger A \text{ is Hermitian.} & (4)
\end{aligned}
$$

In particular, if $A = U\Sigma V^*$ is a singular value decomposition of A (that is, if $U \in M_m$ and $V \in M_n$ are unitary and $\Sigma \in M_{m,n}$ has $\Sigma_{i,i} \geq 0$ for $1 \leq i \leq \min(m,n)$ and $\Sigma_{i,j} = 0$ otherwise) then it may be verified (by checking (1)–(4)) that $A^\dagger = V\Sigma^\dagger U^*$, where $\Sigma^\dagger$ is defined by

$$
\Sigma^\dagger_{i,j} := \begin{cases} \frac{1}{\Sigma_{i,i}} & \text{if } i = j \text{ and } \Sigma_{i,i} \neq 0 \\ 0 & \text{otherwise.} \end{cases}
$$

Classical references on generalized inverses are [3] and [9].

Theorem 2 (Riedel) *Let s and n be positive integers with $s \leq n$; $A \in M_n$; $G \in M_s$; $Y, Y_p \in M_{n,s}$; $Z, Z_p \in M_{n,s}$. Assume $R(Y) \subseteq R(A)$, $R(Y_p) \perp R(A)$, $R(Z) \subseteq R(A^*)$, $R(Z_p) \perp R(A^*)$, G is invertible, Y_p is of full rank, and Z_p is of full rank. Assume also that $R(Y_p) = R(Z_p)$. Then*

$$
(A + (Y + Y_p)G(Z + Z_p)^*)^\dagger = A^\dagger - DZ^*A^\dagger - A^\dagger YC^* + D(G^{-1} + Z^*A^\dagger Y)C^*,
$$

where $C := Y_p(Y_p^*Y_p)^{-1}$ and $D := Z_p(Z_p^*Z_p)^{-1}$.

The matrices $(Y + Y_p)G(Z + Z_p)^*$ in Theorem 2 and YGZ^* in Theorem 1 are referred to as the *update matrices* to the *initial matrix* A. A version of Riedel's theorem

(Theorem 2) for the special case where we seek the Moore–Penrose generalized inverse of a rank-one update to the initial matrix can be found in [9].

Riedel verifies Theorem 2 by checking conditions (1)–(4). It must, however, be noted that the hypothesis $R(Y_p) = R(Z_p)$ of Theorem 2 is nowhere used in the verification of Theorem 2, and thus *Theorem 2 is true without this part of the hypothesis.* It is this key observation that allows us to make use of Theorem 2 in this paper. **When we refer to Theorem 2 henceforth, we will be referring to this theorem without the aforementioned unnecessary hypothesis.**

The purpose of this paper, given matrices A and B and suitable conditions, is to relate $(A+B)^\dagger$ cleanly to $A^\dagger$ and $B^\dagger$. This is done in Theorem 3, our main result, using Riedel's theorem. (For a subspace Ω we denote by P_Ω the orthogonal projection onto Ω.)

Theorem 3 *Let* $A, B \in M_n$ *with* $\operatorname{rank}(A+B) = \operatorname{rank}A + \operatorname{rank}B$. *Then*

$$(A+B)^\dagger = (I - S)\, A^\dagger\, (I - T) + SB^\dagger T \qquad (5)$$

$$\text{where} \qquad S := (P_{R(B^*)} P_{R(A^*)^\perp})^\dagger \qquad \text{and}$$
$$T := (P_{R(A)^\perp} P_{R(B)})^\dagger.$$

Example 4 Without the rank-additivity hypothesis [$\operatorname{rank}(A+B) = \operatorname{rank}A + \operatorname{rank}B$], the conclusion of Theorem 3 is (in general) false. For example, let A and B be 1×1 matrices with 1 as their only entry. In the notation of Theorem 3, we compute

$S = ([1][0])^\dagger = [0]$ and $T = ([0][1])^\dagger = [0]$. Hence,

$$(I - S)A^\dagger(I - T) + SB^\dagger T = [1][1]^\dagger[1] + [0][1]^\dagger[0] = [1]$$

while $(A+B)^\dagger = [\frac{1}{2}]$, contrary to the assertion of Theorem 3. But note $\text{rank}[2] = 1 \neq 2 = \text{rank}[1] + \text{rank}[1]$. □

Is it possible, however, that the rank-additivity hypothesis in the statement of Theorem 3 can be eliminated in favor of a weaker condition? We show in Proposition 5 that the rank-additivity hypothesis cannot be avoided in any proof of Theorem 3 which employs Riedel's theorem (Theorem 2), since rank additivity is shown to be implied by the hypotheses of Riedel's theorem. (As mentioned, our proof of Theorem 3 relies on Theorem 2.)

For conditions when $\text{rank}(A + B) = \text{rank}A + \text{rank}B$, see [8].

Remark. The matrices S and T appearing in (5) are far from determined by (5). For example, let x and y be orthonormal vectors in $\mathbf{C}^n$ with $n \geq 3$, and let

$$A := xx^*, \quad B := yy^*.$$

Applying Theorem 3 we obtain

$$(A + B)^\dagger = (I - yy^*)A(I - yy^*) + (yy^*)B(yy^*),$$

which simplifies to

$$(xx^* + yy^*)^\dagger = xx^* + yy^*. \tag{6}$$

But applying Theorem 3 with the roles of A and B reversed we obtain the different formula

$$(A + B)^\dagger = (xx^*)A(xx^*) + (I - xx^*)B(I - xx^*)$$

which, however, also simplifies to (6).

1.1 Derivation of Main Result (Theorem 3)

Our proof of Theorem 3 is based on the following proposition.

Proposition 5 *Let s and n be positive integers with $s \leq n$; $A \in M_n$; $G \in M_s$; $Y, Y_p \in M_{n,s}$; $Z, Z_p \in M_{n,s}$. Assume $R(Y) \subseteq R(A)$, $R(Y_p) \perp R(A)$, $R(Z) \subseteq R(A^*)$, and $R(Z_p) \perp R(A^*)$.*
Of the following statements, 1 *implies* 2*. Conversely,* 2 *and* 3 *imply* 1*.*

1. *Y_p and Z_p are of full rank.*

2. $\mathrm{rank}[A + (Y + Y_p)G(Z + Z_p)^*] = \mathrm{rank}A + \mathrm{rank}[(Y + Y_p)G(Z + Z_p)^*]$.

3. $\mathrm{rank}[(Y + Y_p)G(Z + Z_p)^*] = s$.

Proposition 5 is used in proving Theorem 3, but it also demonstrates that rank additivity (of the initial matrix and the update matrix) is implied by the hypotheses of Theorem 2; since our proof of Theorem 3 relies on Theorem 2, the rank additivity hypothesis of Theorem 3 is, for us, unavoidable.

Proof of Proposition 5: Using the assumption $R(Y) \subseteq R(A)$ we find

$$R[A + (Y + Y_p)G(Z + Z_p)^*] \subseteq R(A) + R(Y) + R(Y_p) = R(A) + R(Y_p).$$

Thus, if Statements 2 and 3 hold, then

$$\mathrm{rank}A + s \;=\; \mathrm{rank}[A + (Y + Y_p)G(Z + Z_p)^*] \leq \mathrm{rank}A + \mathrm{rank}Y_p,$$

from which we conclude that $\mathrm{rank} Y_p \geq s$, that is, that Y_p (and similarly Z_p) is of full rank (Statement 1).

Conversely, suppose Y_p and Z_p are of full rank (Statement 1). We have

$$\mathrm{rank} Y_p = s \geq \mathrm{rank} G \geq \mathrm{rank}\left[(Y + Y_p)G(Z + Z_p)^*\right]. \tag{7}$$

In [10], Riedel points out that (when Y_p and Z_p are of full rank)

$$\left[A + (Y + Y_p)G(Z + Z_p)^*\right]\left[A + (Y + Y_p)G(Z + Z_p)^*\right]^\dagger = AA^\dagger + Y_p Y_p^\dagger.$$

By the orthogonality of $R(A)$ and $R(Y_p)$, we have $\mathrm{rank}(AA^\dagger + Y_p Y_p^\dagger) = \mathrm{rank}(AA^\dagger) + \mathrm{rank}(Y_p Y_p^\dagger)$. (Without loss of generality, $AA^\dagger$ and $Y_p Y_p^\dagger$ share the same unitary matrices in their singular value decompositions because of this orthogonality.) Thus,

$$\begin{aligned}
\mathrm{rank}[A + (Y + Y_p)G(Z + Z_p)^*] &= \mathrm{rank}(AA^\dagger + Y_p Y_p^\dagger) \\
&= \mathrm{rank}(AA^\dagger) + \mathrm{rank}(Y_p Y_p^\dagger) \\
&= \mathrm{rank} A + \mathrm{rank} Y_p \\
&\geq \mathrm{rank} A + \mathrm{rank}[(Y + Y_p)G(Z + Z_p)^*],
\end{aligned}$$

the last inequality holding by (7). Because (trivially)

$$\mathrm{rank}[A + (Y + Y_p)G(Z + Z_p)^*] \leq \mathrm{rank} A + \mathrm{rank}[(Y + Y_p)G(Z + Z_p)^*],$$

we conclude

$$\mathrm{rank}[A + (Y + Y_p)G(Z + Z_p)^*] = \mathrm{rank} A + \mathrm{rank}[(Y + Y_p)G(Z + Z_p)^*],$$

that is, Statement 2 holds. □

In proving Theorem 3 we will need also the following three facts about the Moore–Penrose generalized inverse that can be verified directly from (1)–(4). For positive integers t and n such that $t \leq n$, let $L_{n,t}$ denote a matrix of size $n \times t$ with ones on the diagonal and zeros elsewhere. Let r, s, p, and q be positive integers with $s \leq p$ and $r \leq q$, and let $A \in M_{r,s}$, $U \in M_r$, and $V \in M_s$ with U and V unitary. Then

$$(L_{q,r} A L_{p,s}^*)^\dagger = L_{p,s} A^\dagger L_{q,r}^* \tag{8}$$

and

$$(UAV^*)^\dagger = V A^\dagger U^*. \tag{9}$$

If A is of full rank with $r \geq s$, then

$$A^\dagger = (A^* A)^{-1} A^*. \tag{10}$$

Proof of Theorem 3: To simplify notation, and since n is fixed, we shorten $L_{n,t}$ to L_t for $t \leq n$. Let A and B have respective singular value decompositions $U_A \Sigma_A V_A^*$ and $U_B \Sigma_B V_B^*$, where, without loss of generality, exactly the first s diagonal entries of Σ_B are nonzero and exactly the first r diagonal entries of Σ_A are zero.

Note that

$$A + B = A + U_B L_s L_s^* \Sigma_B L_s L_s^* V_B^* = A + (Y + Y_p) G (Z + Z_p)^*, \tag{11}$$

where we define

$$\begin{aligned} G &:= L_s^* \Sigma_B L_s, \\ Y &:= P_{R(A)} U_B L_s = [U_A (I - L_r L_r^*) U_A^*] U_B L_s, \end{aligned}$$

$$
\begin{aligned}
Y_p &:= P_{R(A)^\perp} U_B L_s = [U_A L_r L_r^* U_A^*] U_B L_s, \\
Z &:= P_{R(A^*)} V_B L_s = [V_A (I - L_r L_r^*) V_A^*] V_B L_s, \\
Z_p &:= P_{R(A^*)^\perp} V_B L_s = [V_A L_r L_r^* V_A^*] V_B L_s.
\end{aligned}
$$

Note that G, Y, Y_p, Z, and Z_p satisfy all of the hypotheses of Theorem 2 since Y_p and Z_p are of full rank by Proposition 5 (because $\mathrm{rank} B = s$ and $\mathrm{rank}(A+B) = \mathrm{rank} A + \mathrm{rank} B$).

We next observe that (with D and C defined as in Theorem 2)

$$
\begin{aligned}
D G^{-1} C^* &= D L_s^* \Sigma_B^\dagger L_s C^* \\
&= D L_s^* V_B^* V_B \Sigma_B^\dagger U_B^* U_B L_s C^* \\
&= D L_s^* V_B^* B^\dagger U_B L_s C^* \\
&= D (Z^* + Z_p^*) B^\dagger (Y + Y_p) C^*, \qquad (12)
\end{aligned}
$$

and thus by Theorem 2 and (12) we have that

$$
(A+B)^\dagger = (I - DZ^*) A^\dagger (I - YC^*) + (DZ^* + DZ_p^*) B^\dagger (YC^* + Y_p C^*). \qquad (13)
$$

This is the basic form of $(A+B)^\dagger$ that we seek, and we proceed to compute DZ^*, YC^*, DZ_p^*, and $Y_p C^*$.

Because

$$
n \geq \mathrm{rank}(A+B) = \mathrm{rank} A + \mathrm{rank} B = n - r + s,
$$

we have $r \geq s$. By this, the fact that projection matrices are Hermitian and idempotent, and (8)–(10), we get

$$
YC^* = Y (Y_p^* Y_p)^{-1} Y_p^*
$$

$$
\begin{aligned}
&= P_{R(A)}U_BL_s(L_s^*U_B^*P_{R(A)^\perp}^*P_{R(A)^\perp}U_BL_s)^{-1}L_s^*U_B^*P_{R(A)^\perp}^* \\
&= P_{R(A)}U_BL_s(L_s^*U_B^*P_{R(A)^\perp}U_BL_s)^{-1}L_s^*U_B^*P_{R(A)^\perp} \\
&= P_{R(A)}U_BL_s(L_s^*U_B^*U_AL_rL_r^*U_A^*U_BL_s)^{-1}L_s^*U_B^*U_AL_rL_r^*U_A^* \\
&= P_{R(A)}U_BL_s[(L_r^*U_A^*U_BL_s)^*(L_r^*U_A^*U_BL_s)]^{-1}(L_r^*U_A^*U_BL_s)^*L_r^*U_A^* \\
&= P_{R(A)}U_BL_s(L_r^*U_A^*U_BL_s)^\dagger L_r^*U_A^* \\
&= P_{R(A)}U_B(L_rL_r^*U_A^*U_BL_sL_s^*)^\dagger U_A^* \\
&= P_{R(A)}(U_AL_rL_r^*U_A^*U_BL_sL_s^*U_B^*)^\dagger \\
&= P_{R(A)}(P_{R(A)^\perp}P_{R(B)})^\dagger \\
&= P_{R(A)}T \qquad (14)
\end{aligned}
$$

and also

$$
\begin{aligned}
DZ^* &= Z_p(Z_p^*Z_p)^{-1}Z^* \\
&= P_{R(A^*)^\perp}V_BL_s(L_s^*V_B^*P_{R(A^*)^\perp}^*P_{R(A^*)^\perp}V_BL_s)^{-1}L_s^*V_B^*P_{R(A^*)}^* \\
&= P_{R(A^*)^\perp}V_BL_s(L_s^*V_B^*P_{R(A^*)^\perp}V_BL_s)^{-1}L_s^*V_B^*P_{R(A^*)} \\
&= V_AL_rL_r^*V_A^*V_BL_s[L_s^*V_B^*V_AL_rL_r^*V_A^*V_BL_s]^{-1}L_s^*V_B^*P_{R(A^*)} \\
&= V_AL_r(L_r^*V_A^*V_BL_s)[(L_r^*V_A^*V_BL_s)^*(L_r^*V_A^*V_BL_s)]^{-1}L_s^*V_B^*P_{R(A^*)} \\
&= V_AL_r(L_r^*V_A^*V_BL_s)^{*\dagger}L_s^*V_B^*P_{R(A^*)} \\
&= V_AL_r(L_s^*V_B^*V_AL_r)^\dagger L_s^*V_B^*P_{R(A^*)} \\
&= V_A(L_sL_s^*V_B^*V_AL_rL_r^*)^\dagger V_B^*P_{R(A^*)} \\
&= (V_BL_sL_s^*V_B^*V_AL_rL_r^*V_A^*)^\dagger P_{R(A^*)} \\
&= (P_{R(B^*)}P_{R(A^*)^\perp})^\dagger P_{R(A^*)}
\end{aligned}
$$

$$= \quad SP_{R(A^*)}. \tag{15}$$

Similarly, we get

$$\begin{aligned} Y_p C^* &= P_{R(A)^\perp} T \qquad \text{and} \\ D Z_p^* &= S P_{R(A^*)^\perp}. \end{aligned} \tag{16}$$

By plugging (14)–(16) into (13), and noting that $P_{R(A^*)}A^\dagger = A^\dagger$ and $A^\dagger P_{R(A)} = A^\dagger$, the assertion of Theorem 3 follows. □

2 Application to the Parallel Sum

It is well known in elementary electronics that if two resistors with resistances r_1 and r_2 are placed in parallel, then the cumulative resistance r is computed by the formula

$$r = r_1(r_1 + r_2)^{-1} r_2 = \left(\frac{1}{r_1} + \frac{1}{r_2}\right)^{-1}. \tag{17}$$

With the idea of generalizing this notion to matrices, Anderson and Duffin [1] define, for $A, B \in M_n$, the *parallel sum* of A and B as

$$A : B \quad := \quad A(A + B)^\dagger B, \tag{18}$$

which, in the case that A and B are (scalar) resistances, is exactly the formula in (17). An alternative definition for the parallel sum of A and B can be found in Rao and Mitra [9], where it is defined as

$$A \| B \quad := \quad \left(A^\dagger + B^\dagger\right)^\dagger, \tag{19}$$

which, in the case that A and B are (scalar) resistances, is again exactly the formula in (17). Given some assumptions on A and B, [9] presents necessary and sufficient conditions for the two definitions of parallel sum to agree.

The following result uses Theorem 3 to provide, under certain conditions, a neat equation relating $A\|B$ to A and B.

Corollary 6 *Let* $A, B \in M_n$ *with* $\mathrm{rank}(A\|B) = \mathrm{rank}A + \mathrm{rank}B$. *Then*

$$\begin{aligned} A\|B &= (I-R)A(I-W) + RBW \\ \textit{where} \qquad R &:= \left(P_{R(B)}P_{R(A)^\perp}\right)^\dagger \qquad \textit{and} \\ W &:= \left(P_{R(A^*)^\perp}P_{R(B^*)}\right)^\dagger . \end{aligned}$$

Corollary 6 is an immediate corollary of Theorem 3, where $A^\dagger$ and $B^\dagger$ of Theorem 6 play the roles of A and B in Theorem 3.

Acknowledgements. The authors thank Adi Ben-Israel, Shih-Ping Han, Roger Horn, and Kurt Riedel for valuable discussions and two anonymous referees for their helpful suggestions.

References

[1] W. N. Anderson, Jr. and R. J. Duffin, Series and parallel addition of matrices, *Journal of Mathematical Analysis and Applications* **26** (1969), 576–594.

[2] M. S. Bartlett, An inverse matrix adjustment arising in discriminant analysis, *Ann. Math. Statist.* **22** (1951), 107–111.

[3] A. Ben-Israel and T. Greville, *Generalized Inverses, Theory and Applications*, Wiley, New York, 1974.

[4] E. Bodewig, *Matrix Calculus*, North Holland, Amsterdam, 1959.

[5] W. W. Hager, Updating the inverse of a matrix, *SIAM Review* **31** (1989), 221–239.

[6] H. V. Henderson and S. R. Searle, On deriving the inverse of a sum of matrices, *SIAM Review* **23** (1981), 53–60.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[8] G. Marsaglia and G. P. H. Styan, When does $\mathrm{rank}(A+B) = \mathrm{rank}A + \mathrm{rank}B$?, *Canadian Mathematical Bulletin* **15** (1972), 451–452.

[9] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*, Wiley, New York, 1971.

[10] K. S. Riedel, A Sherman–Morrison–Woodbury identity for rank augmenting matrices with application to centering, *SIAM J. Matrix Anal. Appl.* **13** (1992), 659–662.

[11] J. Sherman and W. J. Morrison, Adjustment of an inverse matrix corresponding to changes in the elements of a given column or given row of the original matrix, *Ann. Math. Statist.* **20** (1949), 621.

[12] J. Sherman and W. J. Morrison, Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, *Ann. Math. Statist.* **21** (1950), 124–127.

[13] M. A. Woodbury, Inverting modified matrices, Technical Report 42, Statistical Research Group, Princeton University, Princeton, NJ (1950).