Characterization of the oblique projector $U(VU)^{\dagger}V$ with application to constrained least squares[☆]

Aleš Černý

Cass Business School, City University London

Abstract

We provide a full characterization of the oblique projector $U(VU)^{\dagger}V$ in the general case where the range of U and the null space of V are not complementary subspaces. We discuss the new result in the context of constrained least squares minimization which finds many applications in engineering and statistics.

AMS classification: 15A09, 15A04, 90C20

Key words: oblique projection, constrained least squares, Zlobec formula

1. Introduction

Let $E \in \mathbb{C}^{m\times m}$ be idempotent, $E^2 = E$. The null space and range of any idempotent matrix are complementary, cf. [1, Theorem 2.8],

$$R(E) + N(E) = \mathbb{C}^m, R(E) \cap N(E) = \{0\},$$

and we say that E is an oblique projector onto $R(E)$ along $N(E)$. For any two complementary subspaces of $\mathbb{C}^m$ we denote the oblique projector onto L along M by $P_{L,M}$. The orthogonal projector onto L is denoted by $P_L := P_{L,L^\perp}$, where $L^\perp$ is the orthogonal complement of L. Oblique projectors arise in numerous engineering and statistical applications, see [1, Chapter 8], [2] and references therein. Many of their properties follow from the general solution to the matrix equation $XAX = X$ studied in 1960-ies in the context of the various pseudoinverses, cf. [3]. This literature is mature, with excellent monographs such as [1]. In particular it is very well understood how to construct an oblique projector with a prescribed range and null space.

Proposition 1.1. *Let L, M be complementary subspaces of $\mathbb{C}^m$. For any two matrices U, V with $R(U) = L$ and $N(V) = M$ one has*

$$P_{L,M} = U(VU)^{\dagger}V,$$

where the superscript "†" denotes the Moore-Penrose inverse. If U and V are in addition orthogonal projectors (i.e. they are Hermitian and idempotent) one obtains an even simpler form due to Greville [4, (3.1) and Theorem 2],

$$P_{L,M} = P_L(P_{M^\perp}P_L)^{\dagger}P_{M^\perp} = (P_{M^\perp}P_L)^{\dagger}. \quad (1)$$

☆I would like to thank Adi Ben-Israel, Laura Rebollo-Neira, Götz Trenkler and two anonymous referees for helpful comments and pointers to references. Any errors in the manuscript are my responsibility.

The converse problem of characterizing the range and null space of a given idempotent matrix has not received the same amount of attention. The motivation for studying idempotents of the form $U(VU)^{\dagger}V$ in the general case where $R(U) + N(V) \subsetneq \mathbb{C}^m$ and/or $R(U) \cap N(V) \neq \{0\}$ comes, among others, from constrained least squares optimization with a range of applications mentioned above. Briefly, the problem

$$\min_{x \in \mathbb{C}^n} \|A_1 x - b_1\|^2, \text{ subject to } A_2 x = b_2,$$

gives rise to the projector $D_2(A_1 D_2)^{\dagger} A_1$ where D_2 is an arbitrary but fixed matrix with the property $R(D_2) = N(A_2)$. In this situation we typically have neither $R(D_2) + N(A_1) = \mathbb{C}^m$ nor $R(D_2) \cap N(A_1) = \{0\}$. Oblique projectors of the form $U(VU)^{\dagger}V$ with $R(U) + N(V) = \mathbb{C}^m$ and $R(U) \cap N(V) \neq \{0\}$ feature also in signal reconstruction, cf. [5].

Given that $U(VU)^{\dagger}V$ has a wide range of applications it is desirable to understand its geometric nature. One might conjecture that in general

$$U(VU)^{\dagger}V = P_{L,M}, \text{ where} \tag{2}$$

$$L = P_{R(U)} N(V)^{\perp} = R(U) \cap (R(U) \cap N(V))^{\perp}, \tag{3}$$

$$M = N(V) + (N(V) + R(U))^{\perp}, \tag{4}$$

but the behaviour of the projector is somewhat more intricate and cannot be described based on the knowledge of $R(U)$ and $N(V)$ alone. The conjecture (2)-(4) turns out to be true only when both U and V are orthogonal projectors. Surprisingly, the main tool in proving the general result is the Zlobec formula [6] in conjunction with Proposition 1.1.

The result presented here is different from the problem discussed by Rao and Yanai [7] in which projectors onto and along two given subspaces are considered under the assumption that the subspaces are not necessarily spanning the whole space. In such a situation, the projector no longer needs to be idempotent.

The paper is organized as follows. In section 2 we introduce required terminology and notation, we establish the main tools and prove Proposition 1.1. In section 3 we state and prove the main result. In section 4 we discuss application of the main result to constrained least squares minimization and the link to the minimal norm solution of Eldén [8].

2. Preliminaries

We use notation of [1]. A^* denotes the conjugate transpose of matrix A. We write $r(A)$, $R(A)$, $N(A)$ for the rank, range and null space of A, respectively. Consider the following relations

$$AXA = A, \tag{I.1}$$

$$XAX = X, \tag{I.2}$$

$$AX = (AX)^*, \tag{I.3}$$

$$XA = (XA)^*. \tag{I.4}$$

We write $X \in A\{i, j, \ldots, k\}$, if X satisfies conditions (I.i), (I.j), ..., (I.k). $A^{\dagger}$ denotes the Moore-Penrose inverse which is the unique element of $A\{1, 2, 3, 4\}$. The following theorem is our main tool.

Theorem 2.1 ([1, Theorem 2.13]). *Let $A \in \mathbb{C}^{m\times n}, \tilde{U} \in \mathbb{C}^{n\times s}, \tilde{V} \in \mathbb{C}^{t\times m}$ and*

$$Z = \tilde{U}(\tilde{V}A\tilde{U})^{(1)}\tilde{V},$$

where $(\tilde{V}A\tilde{U})^{(1)}$ is a fixed but arbitrary element of $(\tilde{V}A\tilde{U})\{1\}$. Then

a) $Z \in A\{1\}$ if and only if $r(\tilde{V}A\tilde{U}) = r(A)$;

b) $Z \in A\{2\}$ and $R(Z) = R(\tilde{U})$ if and only if $r(\tilde{V}A\tilde{U}) = r(\tilde{U})$;

c) $Z \in A\{2\}$ and $N(Z) = N(\tilde{V})$ if and only if $r(\tilde{V}A\tilde{U}) = r(\tilde{V})$;

d) $Z = A^{(1,2)}_{R(\tilde{U}),N(\tilde{V})}$ if and only if $r(\tilde{U}) = r(\tilde{V}) = r(\tilde{V}A\tilde{U}) = r(A)$, where $A^{(1,2)}_{R(\tilde{U}),N(\tilde{V})}$ is the unique element of $A\{1,2\}$ with range $R(\tilde{U})$ and null space $N(\tilde{V})$, also known as the oblique pseudoinverse (cf. [9]).

Corollary 2.2. *The Zlobec formula [6],*

$$A^{\dagger} = A^*(A^*AA^*)^{(1)}A^*, \tag{5}$$

is now obtained by setting $\tilde{U} = \tilde{V} = A^*$ in part d) and arguing $A^{(1,2)}_{R(A^*),N(A^*)} = A^{\dagger}$.

The following is a pre-cursor to the main result in this note. The "if" part appears, for example, in [10, (3.51)].

Corollary 2.3. *$\tilde{U}(\tilde{V}\tilde{U})^{(1)}\tilde{V} = P_{R(\tilde{U}),N(\tilde{V})}$ if and only if $r(\tilde{V}\tilde{U}) = r(\tilde{V}) = r(\tilde{U})$.*

Next we show that the form $U(VU)^{\dagger}V$ covers all idempotent matrices.

Lemma 2.4. *Let $U \in \mathbb{C}^{m\times p}, V \in \mathbb{C}^{q\times m}$. $R(U)$ and $N(V)$ are complementary subspaces of $\mathbb{C}^m$ if and only if $r(U) = r(V) = r(VU)$.*

Proof. If: By Corollary 2.3 $U(VU)^{\dagger}V = P_{R(U),N(V)}$ which implies that $R(U), N(V)$ are complementary.

Only if: i) complementarity implies $\dim(R(U)) + \dim(N(V)) = m$. On rearranging we obtain $r(U) = m - \dim(N(V))$ and by the rank-nullity theorem $r(U) = r(V)$.

ii) Complementarity also implies $R(U) \cap N(V) = \{0\}$ which yields $N(VU) = N(U)$. By rank-nullity theorem we obtain $r(VU) = r(U)$. □

Proposition 2.5. *Matrix $E \in \mathbb{C}^{m\times m}$ is idempotent if and only if there are matrices $U \in \mathbb{C}^{m\times p}, V \in \mathbb{C}^{q\times m}$ such that*

$$E := U(VU)^{\dagger}V. \tag{6}$$

Proof. The 'if' statement follows easily from (6) and (I.2),

$$E^2 = U(VU)^{\dagger}VU(VU)^{\dagger}V = E.$$

The 'only if' part: construct U so that its columns form a basis of $R(E)$ and construct V^* so that its columns form the basis of $N(E)^{\perp}$. This implies $R(U) = R(E), N(V) = N(E)$. Since E is idempotent $R(U), N(V)$ are by construction complementary and from Lemma 2.4 we obtain $r(U) = r(V) = r(VU)$. By Corollary 2.3 $U(VU)^{\dagger}V = P_{R(E),N(E)} = E$. □

Remark 2.6. *A comprehensive characterization of projectors appears in [3]. Proposition 2.5 resembles a result of Mitra [11, Theorem 3a] who shows that all idempotent matrices are of the form $\tilde{U}(\tilde{V}\tilde{U})^{(1,2)}\tilde{V}$ where $(\tilde{V}\tilde{U})^{(1,2)}$ is an arbitrary element of $\tilde{V}\tilde{U}\{1,2\}$. This result is generalized further in [1, Theorem 2.13] to the form $\tilde{U}(\tilde{V}\tilde{U})^{(1)}\tilde{V}$, see Corollary 2.3. Proposition 2.5 goes in the opposite direction in order to avoid the ambiguity associated with $\{1,2\}$-inverses.*

To conclude we provide a proof of Proposition 1.1.

Proof (Proposition 1.1). The first statement follows from the 'only if' part in the proof of Proposition 2.5. The second part follows from identities $(P_{M^\perp}P_L)^\dagger = P_L(P_{M^\perp}P_L)^\dagger = (P_{M^\perp}P_L)^\dagger P_{M^\perp}$, see [1, Exercise 2.57]. □

3. Result

Theorem 3.1. *Given two arbitrary matrices $U \in \mathbb{C}^{m\times p}, V \in \mathbb{C}^{q\times m}$ the matrix $E = U(VU)^\dagger V$ is idempotent with range and null space given by*

$$R(E) = R(UU^*V^*) = R(UU^*V^*V) = R(U) \cap ((UU^*)^\dagger(R(U) \cap N(V)))^\perp, \tag{7}$$

$$N(E) = N(U^*V^*V) = N(UU^*V^*V) = N(V) \oplus (V^*V)^\dagger(R(U) + N(V))^\perp. \tag{8}$$

Proof. By Zlobec's formula (5) with $A = VU$ we obtain

$$E = UU^*V^*(U^*V^*VUU^*V^*)^{(1)}U^*V^*V.$$

Setting $\tilde{U} = UU^*V^*, \tilde{V} = U^*V^*V$ we claim $r(\tilde{U}) = r(\tilde{V}) = r(\tilde{V}\tilde{U}) = r(VU)$. Indeed,

$$r(VU) = r(VUU^*V^*) = r(VUU^*V^*VUU^*V^*) \le r(U^*V^*VUU^*V^*) = r(\tilde{V}\tilde{U}), \tag{9}$$

$$r(\tilde{V}\tilde{U}) \le r(\tilde{U}) = r(UU^*V^*) \le r(U^*V^*) = r(VU), \tag{10}$$

$$r(\tilde{V}\tilde{U}) \le r(\tilde{V}) = r(U^*V^*V) \le r(U^*V^*) = r(VU). \tag{11}$$

Corollary 2.3 yields $R(E) = R(\tilde{U}), N(E) = N(\tilde{V})$. From

$$r(VU) = r(VUU^*V^*) = r(VUU^*V^*VUU^*V^*) \le r(UU^*V^*V) \le r(U^*V^*) = r(VU),$$

and from (9)-(11) we obtain $r(VU) = r(UU^*V^*) = r(UU^*V^*V)$ which implies $R(UU^*V^*) = R(UU^*V^*V)$. The proof of $N(U^*V^*V) = N(UU^*V^*V)$ proceeds similarly by showing $r(U^*V^*V) = r(UU^*V^*V)$.

To show the last equality in (8) we observe $\mathbb{C}^m = N(V) \oplus R(V^*)$. Since $N(V) \subseteq N(U^*V^*V)$ we have

$$N(U^*V^*V) = N(V) \oplus (R(V^*) \cap N(U^*V^*V)). \tag{12}$$

Continuing with the second term on the right hand side we obtain

$$\begin{aligned} y \in R(V^*) \cap N(U^*V^*V) &\iff (V^*Vy \in N(U^*) \cap R(V^*)) \wedge (y \in R(V^*)) \\ &\iff y \in (V^*V)^\dagger(N(U^*) \cap R(V^*)), \end{aligned}$$

which yields

$$R(V^*) \cap N(U^*V^*V) = (V^*V)^\dagger(R(U)^\perp \cap N(V)^\perp) = (V^*V)^\dagger(R(U) + N(V))^\perp. \tag{13}$$

On substituting (13) into (12) we obtain the desired result.

The last equality in (7) is obtained by writing $R(UU^*V^*) = N(VUU^*)^\perp$ and then evaluating $N(VUU^*)$ by exchanging the role of U and V^* in (12) and (13). □

Remark 3.2. *Special cases of Theorem 3.1 include situations covered by Corollary 2.3 in which $r(U) = r(V) = r(VU)$ and we have $R(E) = R(U), N(E) = N(V)$; the Langenhop form [12, Lemma 2.2] with $VU = I$ is a case in point. The Greville formula (1) also falls into this category. Hirabayashi and Unser [5, Lemma 3] encounter the case $R(U) + N(V) = \mathbb{C}^m$ and $R(U) \cap N(V) \neq \{0\}$, yielding $R(E) = R(UU^*V^*), N(E) = N(V)$.*

4. Application

Proposition 4.1. *Let* $A_1 \in \mathbb{C}^{m\times n}, b_1 \in \mathbb{C}^m, A_2 \in \mathbb{C}^{k\times n}, r(A_2) = k \geq 1, b_2 \in \mathbb{C}^k$. *Solutions of the problem*

$$\min_{x\in\mathbb{C}^n} \|A_1 x - b_1\|^2\ , \text{ subject to } A_2 x = b_2, \tag{14}$$

lie in the set

$$\Xi = \{D_2(A_1D_2)^\dagger A_1 A_1^\dagger b_1 + (I - D_2(A_1D_2)^\dagger A_1)(A_2^\dagger b_2 + z) : z \in N(A_2)\}, \tag{15}$$

where D_2 *is an arbitrary but fixed matrix with the property* $R(D_2) = N(A_2)$.

Proof. See [1, Exercise 3.10]. □

In general, the projector $D_2(A_1D_2)^\dagger A_1$ will depend on how D_2 is chosen. However, Theorem 3.1 shows that there is a special case when $D_2(A_1D_2)^\dagger A_1$ is actually invariant to the choice of D_2.

Corollary 4.2. *Using the notation of Proposition 4.1 assume further* $r(A_1) = n$. *Then*

$$D_2(A_1D_2)^\dagger A_1 = P_{N(A_2),(A_1^*A_1)^{-1}R(A_2^*)},$$

and Ξ *is a singleton,*

$$\Xi = \{A_1^\dagger b_1 + (A_1^*A_1)^{-1}A_2^*(A_2(A_1^*A_1)^{-1}A_2^*)^{-1}(b_2 - A_2A_1^\dagger b_1)\}.$$

Proof. We have $N(A_1) = 0$ and by Theorem 3.1

$$R(D_2(A_1D_2)^\dagger A_1) = R(D_2) \cap (\{0\})^\perp = N(A_2),$$
$$N(D_2(A_1D_2)^\dagger A_1) = (A_1^*A_1)^{-1}R(D_2)^\perp = (A_1^*A_1)^{-1}R(A_2^*).$$

This implies $(I - D_2(A_1D_2)^\dagger A_1)z = 0$ for all $z \in N(A_2)$ and by Proposition 1.1

$$(I - D_2(A_1D_2)A_1) = (A_1^*A_1)^{-1}A_2^*(A_2(A_1^*A_1)^{-1}A_2^*)^{-1}A_2.$$

The rest follows from Proposition 4.1. □

Note that Corollary 4.2 is not covered by Corollary 2.3 since $n - k = r(D_2) = r(A_1D_2) < r(A_1) = n$. In situations where the choice of D_2 impacts on the projector $D_2(A_1D_2)^\dagger A_1$ Theorem 3.1 guides us to the convenient choice of D_2 which simplifies the geometry of the result and also helps to identify the element of Ξ with minimal distance from a given reference point.

Corollary 4.3. *Using the notation of Proposition 4.1 the following statements hold:*

1. *The constrained least squares minimizer in (14) lies in the set*

$$\Xi = \{A_1^\dagger b_1 + P_{\mathcal{Y},\mathcal{X}}(A_2^\dagger b_2 - A_1^\dagger b_1) + z : z \in N(A_1) \cap N(A_2)\}, \tag{16}$$

with

$$P_{\mathcal{X},\mathcal{Y}} = I - P_{\mathcal{Y},\mathcal{X}} = (A_1(I - A_2^\dagger A_2))^\dagger A_1, \tag{17}$$
$$\mathcal{X} = P_{N(A_2)}R(A_1^*) = N(A_2) \cap (N(A_2) \cap N(A_1))^\perp, \tag{18}$$
$$\mathcal{Y} = N(A_1) \oplus (A_1^*A_1)^\dagger(N(A_1) + N(A_2))^\perp. \tag{19}$$

2. *The element of* Ξ *with the smallest Euclidean norm is given by*

$$\xi := A_1^\dagger b_1 + P_{\mathcal{Y},\mathcal{X}}(A_2^\dagger b_2 - A_1^\dagger b_1).$$

3. *For any* $y \in \mathbb{C}^n$ *the solution of* $\min_{x\in\Xi} \|x - y\|$ *is given by*

$$\psi(y) := \xi + P_{N(A_1)\cap N(A_2)}y. \tag{20}$$

Proof. 1. On setting $D_2 = I - A_2^\dagger A_2 = P_{N(A_2)}$ Proposition 4.1 and Theorem 3.1 yield

$$\Xi = A_1^\dagger b_1 + P_{\mathcal{Y},\mathcal{X}}(A_2^\dagger b_2 - A_1^\dagger b_1 + N(A_2)), \tag{21}$$

with $P_{\mathcal{X},\mathcal{Y}}$, $\mathcal{X}$ and $\mathcal{Y}$ given in (17)-(19). From (18) we obtain $N(A_2) = \mathcal{X} \oplus (N(A_1) \cap N(A_2))$ which implies

$$P_{\mathcal{Y},\mathcal{X}} N(A_2) = P_{\mathcal{Y},\mathcal{X}}(N(A_1) \cap N(A_2)) = N(A_1) \cap N(A_2), \tag{22}$$

the last equality following from $N(A_1) \cap N(A_2) \subseteq \mathcal{Y}$. Substitution of (22) into (21) yields (16).

2. By (18) we have $\mathcal{X} \subseteq (N(A_2) \cap N(A_1))^\perp = R(A_1^*) + R(A_2^*)$. Consequently

$$P_{\mathcal{Y},\mathcal{X}}(R(A_1^*) + R(A_2^*)) = (I - P_{\mathcal{X},\mathcal{Y}})(R(A_1^*) + R(A_2^*)) \subseteq R(A_1^*) + R(A_2^*). \tag{23}$$

This implies

$$\xi \in R(A_1^*) + R(A_2^*) = (N(A_2) \cap N(A_1))^\perp. \tag{24}$$

By (16) $x - \xi \in N(A_1) \cap N(A_2)$ for any $x \in \Xi$ which together with (24) yields

$$\|x\|^2 = \|x - \xi + \xi\|^2 = \|x - \xi\|^2 + \|\xi\|^2 \text{ for all } x \in \Xi.$$

3. By (16), (20) and (24) we obtain $x - \psi(y) \in N(A_2) \cap N(A_1)$ and $\psi(y) - y \in (N(A_2) \cap N(A_1))^\perp$ which implies $\|x - y\|^2 = \|x - \psi(y) + \psi(y) - y\|^2 = \|x - \xi\|^2 + \|\xi - y\|^2$, for all $x \in \Xi$. □

Remark 4.4. *It is well known that vector* $A_1^\dagger b_1$ *has the smallest Euclidean norm among all solutions of the unconstrained least squares problem* $\min_{x\in\mathbb{C}^n} \|A_1 x - b_1\|$. *We have shown in part 2. of Corollary 4.3 that* $\xi = A_1^\dagger b_1 + P_{\mathcal{Y},\mathcal{X}}(A_2^\dagger b_2 - A_1^\dagger b_1)$ *is the shortest solution of the constrained least squares problem (14).*

Eldén [8, Theorem 2.1] studied minimal norm solutions of constrained least squares. On setting

$$h = b_2 - A_2 A_1^\dagger b_1, \quad f = x - A_1^\dagger b_1, \quad K = A_1, \quad L = A_2, \quad M = I,$$

Eldén's solution yields that

$$\zeta := A_1^\dagger b_1 + (I - P_{N(A_2)}(A_1 P_{N(A_2)})^\dagger A_1) A_2^\dagger (b_2 - A_2 A_1^\dagger b_1)$$

minimizes the Euclidean distance $\|x - A_1^\dagger b_1\|$ *among all constrained minimizers* $x \in \Xi$.

With a little bit of work one finds $\zeta = \xi - P_{\mathcal{Y},\mathcal{X}} P_{N(A_2)} A_1^\dagger b_1 = \xi$, *since* $P_{N(A_2)} A_1^\dagger \in \mathcal{X}$ *by virtue of (18). Thus part 3. of Corollary 4.3 simplifies and extends Eldén's result.*

References

[1] A. Ben-Israel, T. N. E. Greville, Generalized inverses: Theory and applications, 2nd Edition, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 15, Springer-Verlag, New York, 2003.

[2] G. Corach, A. Maestripieri, Weighted generalized inverses, oblique projections, and least-squares problems, Numer. Funct. Anal. Optim. 26 (6) (2005) 659–673.

[3] G. Trenkler, Characterizations of oblique and orthogonal projectors, in: Proceedings of the International Conference on Linear Statistical Inference LINSTAT '93 (Poznań, 1993), Vol. 306 of Math. Appl., Kluwer Acad. Publ., Dordrecht, 1994, pp. 255–270.

[4] T. N. E. Greville, Solutions of the matrix equation $XAX = X$, and relations between oblique and orthogonal projectors, SIAM J. Appl. Math. 26 (1974) 828–832.

[5] A. Hirabayashi, M. Unser, Consistent sampling and signal recovery, IEEE Trans. Signal Process. 55 (8) (2007) 4104–4115.

[6] S. Zlobec, An explicit form of the Moore-Penrose inverse of an arbitrary complex matrix, SIAM Rev. 12 (1970) 132–134.

[7] C. R. Rao, H. Yanai, General definition and decomposition of projectors and some applications to statistical problems, J. Statist. Plann. Inference 3 (1) (1979) 1–17.

[8] L. Eldén, Perturbation theory for the least squares problem with linear equality constraints, SIAM J. Num. Anal. 17 (3) (1980) 338–350.

[9] R. D. Milne, An oblique matrix pseudoinverse, SIAM J. Appl. Math. 16 (1968) 931–944.

[10] D. S. G. Pollock, The algebra of econometrics, John Wiley & Sons Ltd., Chichester, 1979, Wiley Series in Probability and Mathematical Statistics.

[11] S. K. Mitra, On a generalised inverse of a matrix and applications, Sankhyā Ser. A 30 (1968) 107–114.

[12] C. E. Langenhop, On generalized inverses of matrices, SIAM J. Appl. Math. 15 (1967) 1239–1246.