OBLIQUE PROJECTIONS, PSEUDOINVERSES, AND STANDARD-FORM TRANSFORMATIONS*

PER CHRISTIAN HANSEN†

Abstract. Standard-form transformation is a technique for transforming a discrete ill-posed problem in general form (with a seminorm as regularizing term) into a standard-form problem. We give a simple geometric explanation why the weighted pseudoinverse is the correct inverse to use in the standard-form transformation. Our presentation is based on oblique projections and oblique pseudoinverses – tools that are often overlooked in numerical analysis.

Key words. Oblique projection, oblique pseudoinverse, weighted pseudoinverse.

AMS subject classifications. 65F22

1. Introduction. The weighted pseudoinverse plays an important role in the numerical treatment of discrete ill-posed problems in general form, i.e., when we want to solve Tikhonov regularization problems of the general form

$$\min \left\{ \|A\,x - b\|_2^2 + \lambda^2 \|L\,x\|_2^2 \right\} \tag{1.1}$$

where L is a matrix with full row rank that determines the smoothness properties of the regularized solution. We refer to [5] for various applications and motivations for solving this problem. The matrix L is very often a discrete approximation to a derivative operator, and it is intentionally rank deficient such that its null space reflects that of the operator. The problem (1.1) can conveniently be transformed into a problem in standard form, i.e., a problem of the form

$$\min \left\{ \|\bar{A}\,\bar{x} - \bar{b}\|_2^2 + \lambda^2 \|\bar{x}\|_2^2 \right\} \qquad \text{with} \qquad \bar{x} = L\,x \tag{1.2}$$

which can then be solved by standard methods. It is suggested in several works, such as [4] and [5], that $\bar{A}$ and $\bar{b}$ take the form

$$\bar{A} = A\,L_A^\dagger, \qquad \bar{b} = b - A\,x_0, \tag{1.3}$$

where $L_A^\dagger$ is the A-weighted pseudoinverse of L, and x_0 is the component of the solution in the null space of L (we return to the details in §5). Finally, $\bar{x}$ is transformed back to the original setting by the relation $x = L_A^\dagger \bar{x}$.

This idea seems to arise in work by Hilgers [6] who considered regularization problems in a Hilbert space setting, and suggested a standard-form transformation where the derivative operator (corresponding to our matrix L) is "absorbed" into the integral operator (here represented by our matrix A). The idea was formulated in the matrix setting by Eldén [3] who made the connection to the weighted pseudoinverse of a matrix, and also showed the relation to the GSVD of the matrix pair (A, L).

The purpose of this tutorial note is to demonstrate that the weighted pseudoinverse is closely related to oblique projections and the oblique pseudoinverse. This allows us to give a simple geometric explanation why it is precisely the A-weighted pseudoinverse of L that should be used in the standard-form transformation (1.3).

*This work was carried out during a visit to Dept. of Mathematics and Computer Science, Emory University, supported by grant 21-03-0574 from the Danish Natural Science Research Council.

†Informatics and Mathematical Modelling, Building 321, The Technical University of Denmark, DK-2800 Lyngby, Denmark (email: pch@imm.dtu.dk).

While oblique projections are now used in signal processing [1] they are not part of standard textbook material (an exception is Meyer's book [8]), so we start with a brief introduction to oblique projections in §2 and oblique pseudoinverses in §3. Our treatment is general, with no restrictions on the rank of the matrices involved. Our main analysis is given in §§4–5 where we derive the relationships between the oblique and the weighted pseudoinverses.

Our notation: $\mathcal{R}(X)$ and $\mathcal{N}(X)$ denotes the range and null space of the matrix X, respectively, $\mathcal{X}^\perp$ denotes the orthogonal complement of the subspace $\mathcal{X}$, and $X^\dagger$ denotes the Moore-Penrose generalized inverse (pseudoinverse) of the matrix X.

2. The Oblique Projection. We start with a brief discussion of projections, and we recall that the only requirement of a matrix E to be a projector is that it is *idempotent*, i.e., $E^2 = E$.

Let $\mathcal{X}$ and $\mathcal{Y}$ be subspaces of $\mathbb{R}^m$ that intersect trivially, i.e., $\mathcal{X} \cap \mathcal{Y} = \{0\}$. Then the *projector* on $\mathcal{X}$ along $\mathcal{Y}$ is the linear operator $E_{\mathcal{X},\mathcal{Y}}$ which satisfies the three requirements:

a) $\forall\, x \in \mathcal{X} : \quad E_{\mathcal{X},\mathcal{Y}}\, x = x$,
b) $\forall\, y \in \mathcal{Y} : \quad E_{\mathcal{X},\mathcal{Y}}\, y = 0$,
c) $\forall\, z \in \mathbb{R}^m : \; E_{\mathcal{X},\mathcal{Y}}\, z \in \mathcal{X}$.

The geometric interpretation of $E_{\mathcal{X},\mathcal{Y}}$ is the following: if we decompose an arbitrary vector $s \in \mathbb{R}^m$ into three components

$$x \in \mathcal{X}, \qquad y \in \mathcal{Y}, \qquad z \in (\mathcal{X} \cup \mathcal{Y})^\perp,$$

then $x = E_{\mathcal{X},\mathcal{Y}}\, s$. The following theorem gives the matrix representation of $E_{\mathcal{X},\mathcal{Y}}$.

THEOREM 2.1. *Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^m$ with $\mathcal{X} \cap \mathcal{Y} = \{0\}$. Assume that we are given two matrices X and Y_0 such that*

$$\mathcal{X} = \mathcal{R}(X), \qquad \mathcal{Y}^\perp = \mathcal{R}(Y_0). \tag{2.1}$$

Then a matrix representation of the projector $E_{\mathcal{X},\mathcal{Y}}$ is given by

$$E_{\mathcal{X},\mathcal{Y}} = X\,(Y_0^T X)^\dagger Y_0^T. \tag{2.2}$$

Proof. First note that $E_{\mathcal{X},\mathcal{Y}}^2 = X\,(Y_0^T X)^\dagger Y_0^T X\,(Y_0^T X)^\dagger Y_0^T = X\,(Y_0^T X)^\dagger Y_0^T = E_{\mathcal{X},\mathcal{Y}}$ showing that $E_{\mathcal{X},\mathcal{Y}}$ is idempotent. To verify a) we must show that $E_{\mathcal{X},\mathcal{Y}} X = X$. Introducing the skinny QR factorizations $X = Q_X R_X$ and $Y_0 = Q_{Y_0} R_{Y_0}$, we can write $E_{\mathcal{X},\mathcal{Y}} X = Q_X (Q_{Y_0}^T Q_X)^\dagger Q_{Y_0}^T Q_X R_X$. We know that the number of columns in Q_X and Q_{Y_0} are $n_X = \dim(\mathcal{X})$ and $n_{Y_0} = \dim(\mathcal{Y}^\perp) = m - \dim(\mathcal{Y})$. Since $\mathcal{X}$ and $\mathcal{Y}$ intersect trivially, we have $\dim(\mathcal{X}) + \dim(\mathcal{Y}) \le m$ and therefore $n_X \le m - \dim(\mathcal{Y}) = n_{Y_0}$. Hence $Q_{Y_0}^T Q_X$ has full column rank n_X, and thus $(Q_{Y_0}^T Q_X)^\dagger Q_{Y_0}^T Q_X = I_{n_X}$. We conclude that $E_{\mathcal{X},\mathcal{Y}} X = Q_X\, I_{n_X}\, R_X = X$, showing that requirement a) is fulfilled. Due to our choice of Y_0 it is clear that $E_{\mathcal{X},\mathcal{Y}}\, Y = X\,(Y_0^T X)^\dagger Y_0^T Y = 0$ and hence b) is fulfilled. Finally, it is obvious from (2.2) that $\mathcal{R}(E_{\mathcal{X},\mathcal{Y}}) \subseteq \mathcal{X}$ and thus c) is fulfilled. ∎

The *orthogonal* projector $P_{\mathcal{X}}$ on $\mathcal{X}$ is the special case for which $\mathcal{Y} = \mathcal{X}^\perp$. The well-known formula $P_{\mathcal{X}} = X\,X^\dagger$ follows from the relation

$$P_{\mathcal{X}} = E_{\mathcal{X},\mathcal{X}^\perp} = X\,(X^T X)^\dagger X^T = X\,X^\dagger,$$

which shows that $P_{\mathcal{X}}$ is symmetric (this is one of the four Moore-Penrose conditions on $X^\dagger$).

Whenever $\mathcal{Y} \neq \mathcal{X}^\perp$ the projector $E_{\mathcal{X},\mathcal{Y}}$ is called the *oblique projector*. This matrix is not symmetric, and from the relation $E_{\mathcal{X},\mathcal{Y}}^T = Y_0(X^T Y_0)^\dagger X^T$ and the dimensions of $\mathcal{X}$ and $\mathcal{Y}$ it follows that $E_{\mathcal{X},\mathcal{Y}}^T$ is not necessarily an oblique projector; but

$$E_{\mathcal{X},\mathcal{Y}}^T = E_{\mathcal{Y}^\perp,\mathcal{X}^\perp} \qquad \text{when} \qquad \mathcal{X} + \mathcal{Y} = \mathbb{R}^m. \tag{2.3}$$

Orthogonal projectors obey the simply relation $P_\mathcal{X} + P_{\mathcal{X}^\perp} = I$. The corresponding relation for oblique projectors is expressed in the following theorem.

THEOREM 2.2. *If $P_{\mathcal{X}+\mathcal{Y}}$ denotes the orthogonal projector on $\mathcal{X}+\mathcal{Y}$, then*

$$E_{\mathcal{X},\mathcal{Y}} + E_{\mathcal{Y},\mathcal{X}} = P_{\mathcal{X}+\mathcal{Y}} \tag{2.4}$$

and

$$E_{\mathcal{X},\mathcal{Y}} + E_{\mathcal{Y},\mathcal{X}} + P_{(\mathcal{X}+\mathcal{Y})^\perp} = I_m. \tag{2.5}$$

Proof. Obviously $E_{\mathcal{X},\mathcal{Y}}\, E_{\mathcal{Y},\mathcal{X}} = 0$ and $E_{\mathcal{Y},\mathcal{X}}\, E_{\mathcal{X},\mathcal{Y}} = 0$; hence $(E_{\mathcal{X},\mathcal{Y}} + E_{\mathcal{Y},\mathcal{X}})^2 = E_{\mathcal{X},\mathcal{Y}}^2 + E_{\mathcal{Y},\mathcal{X}}^2 = E_{\mathcal{X},\mathcal{Y}} + E_{\mathcal{Y},\mathcal{X}}$ showing that $E_{\mathcal{X},\mathcal{Y}} + E_{\mathcal{Y},\mathcal{X}}$ is a projector. Now consider $s \in \mathcal{X}+\mathcal{Y}$ and write $s = x + y$ with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$; then $(E_{\mathcal{X},\mathcal{Y}} + E_{\mathcal{Y},\mathcal{X}})s = E_{\mathcal{X},\mathcal{Y}}x + E_{\mathcal{X},\mathcal{Y}}y + E_{\mathcal{Y},\mathcal{X}}x + E_{\mathcal{Y},\mathcal{X}}y = x + y = s$. Next consider $v \in (\mathcal{X}+\mathcal{Y})^\perp = \mathcal{X}^\perp \cap \mathcal{Y}^\perp$; then obviously $(E_{\mathcal{X},\mathcal{Y}} + E_{\mathcal{Y},\mathcal{X}})v = 0$. Finally it is clear from Thm. 2.1 that $(E_{\mathcal{X},\mathcal{Y}} + E_{\mathcal{Y},\mathcal{X}})z \in \mathcal{X}+\mathcal{Y}$ for all $z \in \mathbb{R}^m$. We have thus proved that $E_{\mathcal{X},\mathcal{Y}}x + E_{\mathcal{X},\mathcal{Y}}$ is the orthogonal projector on $\mathcal{X}+\mathcal{Y}$. The second equation following immediately from the relation $P_{\mathcal{X}+\mathcal{Y}} + P_{(\mathcal{X}+\mathcal{Y})^\perp} = I_m$. ∎

There is a rich geometric theory related to projectors and angles between subspaces; see [7] and [8] for details. We omit these details here because they are not important for the results we are aiming at in §§4–5.

3. The Oblique Pseudoinverse. Given a matrix $X \in \mathbb{R}^{m\times n}$ with $m \geq n$ and a subspace $\mathcal{Y}$ such that $\mathcal{X} = \mathcal{R}(X)$ and $\mathcal{Y}$ intersect trivially, we define the *oblique pseudoinverse* $X_\mathcal{Y}^\dagger$ of X by

$$X_\mathcal{Y}^\dagger \equiv X^\dagger E_{\mathcal{X},\mathcal{Y}}. \tag{3.1}$$

In the appendix we show that the oblique pseudoinverse satisfies the following four conditions from [3], similar to the Penrose conditions for the ordinary pseudoinverse:

$$\begin{array}{llll} \text{i)} & X\,X_\mathcal{Y}^\dagger\,X \;=\; X & \text{ii)} & X_\mathcal{Y}^\dagger\,X\,X_\mathcal{Y}^\dagger \;=\; X_\mathcal{Y}^\dagger \\ \text{iii)} & \big(X_\mathcal{Y}^\dagger\,X\big)^T \;=\; X_\mathcal{Y}^\dagger\,X & \text{iv)} & \big(Y_0\,Y_0^T X\,X_\mathcal{Y}^\dagger\big)^T \;=\; Y_0\,Y_0^T X\,X_\mathcal{Y}^\dagger. \end{array} \tag{3.2}$$

Clearly, if $\mathcal{Y} = \mathcal{R}(X)^\perp = \mathcal{X}^\perp$ then $E_{\mathcal{X},\mathcal{X}^\perp} = P_\mathcal{X}$ and thus $X_{\mathcal{X}^\perp}^\dagger = X^\dagger P_\mathcal{X} = X^\dagger$. The following theorems give some important relations involving the oblique pseudoinverse.

THEOREM 3.1. *Let $X_\mathcal{Y}^\dagger$ be defined by (3.1), then*

$$X\,X_\mathcal{Y}^\dagger = E_{\mathcal{X},\mathcal{Y}}, \qquad X_\mathcal{Y}^\dagger X = P_{\mathcal{R}(X^T)}, \qquad X^\dagger = X_\mathcal{Y}^\dagger P_\mathcal{X}. \tag{3.3}$$

Also, if $\mathcal{R}(Y_0) = \mathcal{Y}^\perp$ then

$$X_\mathcal{Y}^\dagger = (Y_0^T X)^\dagger Y_0^T. \tag{3.4}$$

Proof. To show the first identity in (3.3), let $\mathcal{X} = \mathcal{R}(\widehat{X})$ where $\widehat{X}$ has full column rank; then $XX_\mathcal{Y}^\dagger = XX^\dagger\widehat{X}(Y_0^T\widehat{X})^\dagger Y_0^T = \widehat{X}\widehat{X}^\dagger\widehat{X}(Y_0^T\widehat{X})^\dagger Y_0^T = \widehat{X}(Y_0^T\widehat{X})^\dagger Y_0^T = E_{\mathcal{X},\mathcal{Y}}$

because $XX^\dagger = P_{\mathcal{R}(X)} = P_{\mathcal{X}} = \widehat{X}\,\widehat{X}^\dagger$. The second identity follows from $X^\dagger_{\mathcal{Y}}X = X^\dagger E_{\mathcal{X},\mathcal{Y}}X = X^\dagger X = P_{\mathcal{R}(X^T)}$. The third identity follows by inserting $P_{\mathcal{X}} = \widehat{X}\widehat{X}^\dagger$ into $X^\dagger_{\mathcal{Y}}P_{\mathcal{X}} = X^\dagger\widehat{X}(Y_0^T\widehat{X})^\dagger Y_0^T\widehat{X}\widehat{X}^\dagger = X^\dagger\widehat{X}\widehat{X}^\dagger = X^\dagger$. Equation (3.4) follows from the relations $X^\dagger_{\mathcal{Y}} = P_{\mathcal{R}(X^T)}(Y_0^TX)^\dagger Y_0^T$ and $\mathcal{R}\big((Y_0^TX)^\dagger\big) = \mathcal{R}(X^TY_0) \subseteq \mathcal{R}(X^T)$. ☐

THEOREM 3.2. *Consider the weighted least squares problem* $\min\|W(A\,x-b)\|_2$. *If* W *is rank deficient and* $\mathcal{R}(A)\cap\mathcal{N}(W) = \{0\}$, *then the unique solution of minimum 2-norm is given by* $\hat{x} = A^\dagger_{\mathcal{N}(W)}b$.

Proof. The minimum-norm solution is $\hat{x} = (WA)^\dagger W\,b$, and from Thm. 3.1 we see that $(WA)^\dagger W = A^\dagger_{\mathcal{Y}}$ with $\mathcal{Y} = \mathcal{R}(W^T)^\dagger = \mathcal{N}(W)$. ☐

We can use these results to obtain an interesting perspective on the least squares problem associated with the general linear model $b = A\,x + e$ where e is noise with covariance matrix $C = R_C^TR_C$. If C has full rank, then it is well known that one should solve the "prewhitened" system $\min\|R_C^{-T}(A\,x-b)\|_2$, whose solution $\hat{x} = (R_C^{-T}A)^\dagger R_C^{-T}b$ is the best linear unbiased estimate [2]. Introducing $Y_0 = R_C^{-1}R_C^{-T}A = (R_C^TR_C)^{-1}A = C^{-1}A$, we can also write the solution as $\hat{x} = (Y_0^TA)^\dagger Y_0^Tb$, showing that this solution is obtained by means of an oblique pseudoinverse:

$$\hat{x} = A^\dagger_{\mathcal{Y}}\,b \qquad \text{with} \qquad \mathcal{Y} = \mathcal{R}(C^{-1}A)^\perp = \mathcal{R}(C\,W_0),$$

in which $\mathcal{R}(W_0) = \mathcal{N}(A^T)$. Due to (3.1) we can thus write $\hat{x} = A^\dagger E_{\mathcal{R}(A),\mathcal{Y}}\,b$, showing that if we split b according to

$$b = E_{\mathcal{R}(A),\mathcal{Y}}\,b + (I - E_{\mathcal{R}(A),\mathcal{Y}})\,b$$

then $\hat{x}$ is the ordinary least squares solution to the problem $\min\|A\,x - E_{\mathcal{R}(A),\mathcal{Y}}\,b\|_2$. Why do we bother to include the oblique projector $E_{\mathcal{R}(A),\mathcal{Y}}$ in the solution procedure? The answer lies in the covariance matrix for the errors in the solution.

THEOREM 3.3. *Assume that* A *has full column rank, let* $A = Q\,R_A$ *be its skinny QR factorization, and let* C *be a full-rank covariance matrix for the errors in the right-hand side* b. *Let* C_x *and* $C_{\hat{x}}$ *denote the covariance matrices for the errors in the solution* $x = A^\dagger b$ *and* $\hat{x} = A^\dagger_{\mathcal{Y}}b = A^\dagger E_{\mathcal{R}(A),\mathcal{Y}}b$ *with* $\mathcal{Y} = \mathcal{R}(C^{-1}A)^\perp$. *Then*

$$C_x = A^\dagger\,C\,(A^\dagger)^T = R_A^{-1}Q^TC\,Q\,R_A^{-T},$$
$$C_{\hat{x}} = A^\dagger_{\mathcal{Y}}\,CA^\dagger_{\mathcal{Y}} = R_A^{-1}(Q^TC^{-1}\,Q)^{-1}R_A^{-T}$$

and $\|(Q^TC^{-1}Q)^{-1}\|_2 \le \|Q^TC\,Q\|_2$.

Proof. The first relation follows simply from inserting $A^\dagger = R^{-1}Q^T$ into the expression for C_x. The second relation is proved by inserting $A^\dagger_{\mathcal{Y}} = A^\dagger E_{\mathcal{R}(A),\mathcal{Y}} = (A^TC^{-1}A)^{-1}A^TC^{-1}$ into the expression $C_{\hat{x}} = A^\dagger_{\mathcal{Y}}\,C\,(A^\dagger_{\mathcal{Y}})^T$. To prove the third expression, let Q_0 be a matrix with orthonormal columns such that (Q,Q_0) is orthogonal, and consider the matrix $(Q,Q_0)^TC(Q,Q_0)$. Then

$$\begin{pmatrix} Q^TCQ & Q^TCQ_0 \\ Q_0^TCQ & Q_0^TCQ_0 \end{pmatrix}^{-1} = (Q,Q_0)^TC^{-1}(Q,Q_0) = \begin{pmatrix} Q^TC^{-1}Q & Q^TC^{-1}Q_0 \\ Q_0^TC^{-1}Q & Q_0^TC^{-1}Q_0 \end{pmatrix}$$

and via the Schur complement formulas we obtain

$$\left(Q^TC^{-1}Q\right)^{-1} = Q^TCQ - Q^TCQ_0\left(Q_0^TCQ_0\right)^{-1}Q_0^TCQ.$$

To proceed, we use the Cholesky factorization $C = R_C^T R_C$; then

$$\begin{aligned}\left(Q^T C^{-1} Q\right)^{-1} &= (R_C Q)^T \left(I - R_C Q_0 \left((R_C Q_0)^T (R_C Q_0)\right)^{-1} (R_C Q_0)^T\right) R_C Q \\ &= (R_C Q)^T \widehat{P} (R_C Q) = (P R_C Q)^T \widehat{P} R_C Q,\end{aligned}$$

in which $\widehat{P} = I - R_C Q_0 \left((R_C Q_0)^T (R_C Q_0)\right)^{-1} (R_C Q_0)^T$ is an orthogonal projector. Hence $\|Q^T C Q\|_2 = \|R_C Q\|_2^2$ and

$$\left\|\left(Q^T C^{-1} Q\right)^{-1}\right\|_2 = \|\widehat{P} R_C Q\|^2 \le \|\widehat{P}\|_2^2 \, \|R_C Q\|_2^2 = \|R_C Q\|_2^2 = \|Q^T C Q\|_2,$$

because $\|\widehat{P}\|_2 = 1$. This completes the proof. □

Due to the common factors R_A^{-1} and R_A^{-T} in C_x and $C_{\hat{x}}$, we can conclude that the error magnification, from right-hand side errors to errors in the solution, is likely to be smaller when using the oblique projection.

Finally we need the following notation for the oblique pseudoinverse of a matrix with more columns than rows. If $X \in \mathbb{R}^{p \times n}$ with $p \le n$ and if $\mathcal{N}(X) \cap \mathcal{Y} = \{0\}$, then we define[1]

$$X_{\mathcal{Y}}^{\dagger} \equiv E_{\mathcal{Y},\mathcal{N}(X)} X^{\dagger} \tag{3.5}$$

and, using the techniques of the appendix, one can show that this $X_{\mathcal{Y}}^{\dagger}$ satisfies conditions i) and ii) in (3.2) while the other two conditions are replaced by

$$\text{iii)} \quad (X_{\mathcal{Y}}^{\dagger} X \, Y \, Y^T)^T = X_{\mathcal{Y}}^{\dagger} X \, Y \, Y^T \qquad \text{iv)} \quad (X \, X_{\mathcal{Y}}^{\dagger})^T = X \, X_{\mathcal{Y}}^{\dagger}.$$

This definition of the oblique pseudoinverse of a "flat" matrix is useful for the developments in the next section.

THEOREM 3.4. *If $X_{\mathcal{Y}}^{\dagger}$ is defined by (3.5), then*

$$X \, X_{\mathcal{Y}}^{\dagger} = P_{\mathcal{X}}, \qquad X_{\mathcal{Y}}^{\dagger} X = E_{\mathcal{Y},\mathcal{N}(X)}, \qquad X^{\dagger} = P_{\mathcal{R}(X^T)} X_{\mathcal{Y}}^{\dagger}$$

and if $\mathcal{Y} = \mathcal{R}(Y)$ then

$$X_{\mathcal{Y}}^{\dagger} = Y \, (X \, Y)^{\dagger}.$$

The proof is similar to that of Thm. 3.1 and is omitted here.

4. Tikhonov Regularization. Equipped with the oblique projection and the oblique pseudoinverse, we now return to the main subject, namely, to derive the most convenient standard-form Tikhonov problem, given the problem in general form (1.1).

The generalized SVD (GSVD) of the matrix pair (A, L) is important here; see §4.2 in [2] for details. Given $A \in \mathbb{R}^{m \times n}$ and $L \in \mathbb{R}^{p \times n}$ with $\operatorname{rank}(L) = p \le n \le m$, the GSVD has the form

$$A = (\, U_1 \, , \, U_2 \,) \begin{pmatrix} \Sigma & 0 \\ 0 & I_o \end{pmatrix} X^{-1}, \qquad L = V \, (\, M \, , \, 0 \,) \, X^{-1}, \qquad X = (\, X_1 \, , \, X_2 \,)$$

where the matrices $(\, U_1 \, , \, U_2 \,)$ and V have orthonormal columns, Σ and M are diagonal $p \times p$ matrices, I_o is the $(n-p) \times (n-p)$ identity matrix, U_1 and X_1 have p columns, and X is a nonsingular matrix. If we introduce the matrix

$$\Theta = (\, \Theta_1 \, , \, \Theta_2 \,) = X^{-T}$$

[1] If $\mathcal{N}(X) + \mathcal{Y} = \mathbb{R}^n$ then, due to (2.3), $X_{\mathcal{Y}}^{\dagger} = ((X^T)^{\dagger} E_{\mathcal{R}(X^T),\mathcal{Y}^{\perp}})^T = ((X^T)_{\mathcal{Y}^{\perp}}^{\dagger})^T$.

then clearly $A = U_1\,\Sigma\,\Theta_1^T + U_2\,\Theta_2^T$ and $L = V\,M\,\Theta_1^T$; moreover $\mathcal{R}(L^T) = \mathcal{R}(\Theta_1)$ and $\mathcal{N}(L) = \mathcal{R}(X_2)$.

We now introduce the splitting $\mathbb{R}^n = \mathcal{R}(X_1) + \mathcal{R}(X_2)$ with $\mathcal{R}(X_1) \cap \mathcal{R}(X_2) = \{0\}$ (i.e., the two subspaces are complementary), and the two oblique projectors associated with the splitting:

$$E_1 \equiv E_{\mathcal{R}(X_1),\mathcal{R}(X_2)}, \qquad E_2 \equiv E_{\mathcal{R}(X_2),\mathcal{R}(X_1)}, \qquad E_1 + E_2 = I_n.$$

The next step is to write the solution to (1.1) as $x = E_1 x + E_2 x$, and consider the quantities $A\,x = A\,E_1 x + A\,E_2 x$ and $L\,x = L\,E_1 x + L\,E_2 x = L\,E_1 x$. If we introduce the two orthogonal projectors

$$P_1 = P_{\mathcal{R}(U_1)} = U_1\,U_1^T, \qquad P_2 = P_{\mathcal{R}(U_2)} = U_2\,U_2^T,$$

then it is easy to show, by straightforward insertion of the GSVD, that

$$A\,E_1 = P_1\,A\,E_1 = P_1\,A, \qquad A\,E_2 = P_2\,A\,E_2 = P_2\,A.$$

We can therefore rewrite the objective function $\Psi_\lambda = \|A\,x - b\|_2^2 + \lambda^2 \|L\,x\|_2^2$ in the general-form Tikhonov problem (1.1) as follows

$$\begin{aligned} \Psi_\lambda &= \|A\,E_1 x + A\,E_2 x - b\|_2^2 + \lambda^2 \|L\,E_1 x\|_2^2 \\ &= \|P_1\,A\,E_1 x + P_2\,A\,E_2 x - P_1 b - P_2 b\|_2^2 + \beta_0^2 + \lambda^2 \|L\,E_1 x\|_2^2 \\ &= \|P_1 (A\,E_1 x - b)\|_2^2 + \lambda^2 \|L\,E_1 x\|_2^2 + \|P_2 (A\,E_2 x - b)\|_2^2 + \beta_0^2 \end{aligned}$$

where $\beta_0 = \|(I - P_1 - P_2) b\|_2$ is a constant. Since $\|P_1 (A\,E_1 x - b)\|_2 = \|A\,E_1 x - b\|_2$ and $\|P_2 (A\,E_2 x - b)\|_2 = \|A\,E_2 x - b\|_2$, we thus arrive at two separate minimization problems: a Tikhonov problem for $E_1 x$,

$$\|A\,E_1 x - b\|_2^2 + \lambda^2 \|L\,E_1 x\|_2^2,$$

and a least squares problem for $E_2 x$,

$$\min \|A\,E_2 x - b\|_2.$$

According to Thm. 3.4 we have

$$E_1 = E_{\mathcal{R}(X_1),\mathcal{R}(X_2)} = L^\dagger_{\mathcal{R}(X_1)} L, \qquad L\,E_1 = L\,L^\dagger_{\mathcal{R}(X_1)} L = P_{\mathcal{R}(L)} L = L.$$

If we insert these expressions into the first problem, and introduce the new quantities

$$\bar{A} = A\,L^\dagger_{\mathcal{R}(X_1)} \qquad \text{and} \qquad \bar{x} = L\,x,$$

then we obtain the following standard-form Tikhonov problem in $\bar{x}$:

$$\min\{\|\bar{A}\,\bar{x} - b\|_2^2 + \lambda^2 \|\bar{x}\|_2^2\}. \tag{4.1}$$

For the second minimization problem, we note that $x_0 = E_2 x \in \mathcal{N}(L)$ and that we can write $x_0 = X_2 z$, which leads to

$$z = \operatorname{argmin} \|A\,X_2\,z - b\|_2 = (A\,X_2)^\dagger b$$

and thus $x_0 = X_2\,(A\,X_2)^\dagger b$. Hence, if $\bar{x}_\lambda$ denotes the solution to (4.1), then we can write the solution to the original problem as

$$x_\lambda = L^\dagger_{\mathcal{R}(X_1)} \bar{x}_\lambda + x_0, \qquad x_0 = X_2\,(A\,X_2)^\dagger b, \tag{4.2}$$

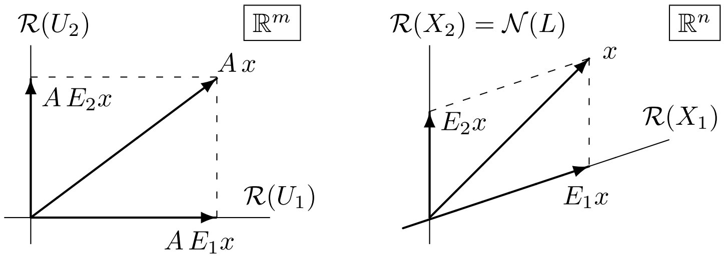

FIG. 4.1. *The geometry of the subspace splittings in the standard-form transformation.*

i.e., a sum of the oblique pseudoinverse $L^\dagger_{\mathcal{R}(X_1)}$ times the solution $\bar{x}_\lambda$ to the standard-form problem (4.1) plus the component x_0 in the null-space of L.

The vector $x_0 = E_2x$ is the component of the solution in the null space of L, and it is independent of the regularization parameter λ. This is natural because any component in $\mathcal{N}(L)$ cannot "feel" the presence of the regularization term $\|L\,x\|_2$ in the Tikhonov formulation.

The above derivation shows precisely why we must use *oblique* projections in the splitting of the solution $x = E_1x + E_2x$, namely, because we want this splitting to correspond to writing $\|A\,x - b\|_2^2$ as a sum of two terms – which is only possible when the columns of the two matrices $A\,E_1$ and $A\,E_2$ are orthogonal, such that the residual vector splits into two orthogonal components. This is illustrated in Fig. 4.1.

One might be tempted to use the orthogonal projector $P_{\mathcal{N}(L)}$ in the splitting $x = P_{\mathcal{N}(L)}x + (I_n - P_{\mathcal{N}(L)})x$, because $L\,P_{\mathcal{N}(L)}x = 0$ and $L\,(I_n - P_{\mathcal{N}(L)}) = L$. However, this does not allow the desired splitting of the residual, because the columns of the two matrices $A\,P_{\mathcal{N}(L)}$ and $A\,(I_n - P_{\mathcal{N}(L)})$ are not necessarily orthogonal.

5. The Weighted Pseudoinverse. It remains to be shown that the two approaches, using the weighted pseudoinverse $L^\dagger_A$ and the oblique pseudoinverse $L^\dagger_{\mathcal{R}(X_1)}$, are identical. It is proved in [3] that the weighted pseudoinverse can be expressed easily in terms of the GSVD of (A, L) as

$$L^\dagger_A = X \begin{pmatrix} M^{-1} \\ 0 \end{pmatrix} V^T = X_1\,M^{-1}V^T.$$

The oblique pseudoinverse, according to Thm. 3.4, is given by

$$\begin{aligned} L^\dagger_{\mathcal{R}(X_1)} &= X_1\,(L\,X_1)^\dagger = X_1\big(V\,(\,M\,,\,0\,)\,X^{-1}X_1\big)^\dagger \\ &= X_1(V\,M)^\dagger = X_1\,M^{-1}V^T = L^\dagger_A. \end{aligned}$$

Thus we have proved that the two pseudoinverses are identical.

One may wonder about the discrepancy between the transformed right-hand side $\bar{b} = b - A\,x_0$ in Eq. (1.3) and the fact that the original b appears in Eq. (4.1). The point is that it does not matter which right-hand side is used, because the component $A\,x_0$ is orthogonal to $\mathcal{R}(\bar{A})$. This follows immediately from the relations $\bar{A} = A\,L^\dagger_{\mathcal{R}(X_1)} = A\,E_1\,L^\dagger = P_1\,A\,L^\dagger$ and $A\,x_0 = A\,E_2x = P_2\,A\,x$. For numerical reasons it may be preferable to use $\bar{b}$.

Equation (2.33) in [5] gives the alternative expression $x_0 = \big(A\,(I - L^\dagger L)\big)^\dagger b$. It is not hard to show that this expression is identical to the one in (4.2), because[2]

$$\big(A\,(I - L^\dagger L)\big)^\dagger = (A\,P_{\mathcal{N}(L)})^\dagger = (A\,X_2 X_2^\dagger)^\dagger = X_2 (A\,X_2)^\dagger.$$

Also in [5] we find the expression $L_A^\dagger = \big(I_n - (A\,(I_n - L^\dagger L))^\dagger A\big) L^\dagger$. To analyze this expression, note that $A\,X_2 = U_2$, and therefore

$$\big(A\,(I - L^\dagger L)\big)^\dagger A = X_2\, U_2^T A = X_2\, W_2^T = E_{\mathcal{R}(X_2),\mathcal{R}(X_1)}$$

and hence $I_n - (A\,(I_n - L^\dagger L))^\dagger A = I_n - E_{\mathcal{R}(X_2),\mathcal{R}(X_1)} = E_{\mathcal{R}(X_1),\mathcal{R}(X_2)}$, from which it follows again that $L_A^\dagger = E_{\mathcal{R}(X_1),\mathcal{R}(X_2)}\, L^\dagger = L^\dagger_{\mathcal{R}(X_1)}$. Yet another expression $L_A^\dagger = (I_n - X_2\,(A^T A\,X_2)^T)\, L^\dagger$ is found in §4.3 in [4]; the equivalence to the relation above follows from $X_2\,(A^T A\,X_2)^T = X_2\,(A^T U_2)^T = X_2\, W_2^T$.

Appendix. Here we prove that $X_\mathcal{Y}^\dagger$ satisfies the four conditions in (3.2). Since X is not guaranteed to have full column rank, we must write $X = \widehat{X}\,Q$ where Q has dimensions $n \times \dim(\mathcal{X})$. Then $X_\mathcal{Y}^\dagger = X^\dagger \widehat{X} (Y_0^T \widehat{X})^\dagger Y_0^T$, and since $(Y_0^T \widehat{X})^\dagger Y_0^T \widehat{X} = I$ we get

$$X\,X_\mathcal{Y}^\dagger\,X = X\,X^\dagger \widehat{X}(Y_0^T \widehat{X})^\dagger Y_0^T \widehat{X}\,Q = X\,X^\dagger \widehat{X}\,Q = X\,X^\dagger X = X,$$

and

$$\begin{aligned} X_\mathcal{Y}^\dagger\,X\,X_\mathcal{Y}^\dagger &= X^\dagger \widehat{X}(Y_0^T \widehat{X})^\dagger Y_0^T \widehat{X}\,Q X^\dagger \widehat{X}(Y_0^T \widehat{X})^\dagger Y_0^T \\ &= X^\dagger X\,X^\dagger \widehat{X}(Y_0^T \widehat{X})^\dagger Y_0^T = X^\dagger \widehat{X}(Y_0^T \widehat{X})^\dagger Y_0^T = X_\mathcal{Y}^\dagger \end{aligned}$$

thus verifying conditions i) and ii). Moreover we have

$$X_\mathcal{Y}^\dagger X = X^\dagger \widehat{X}(Y_0^T \widehat{X})^\dagger Y_0^T \widehat{X}\,Q = X^\dagger X$$

and since $X^\dagger X$ is symmetric we have verified condition iii). Finally due to (3.3):

$$Y_0\,Y_0^T X\,X_\mathcal{Y}^\dagger = Y_0\,Y_0^T E_{\mathcal{X},\mathcal{Y}} = Y_0\,Y_0^T \widehat{X}(Y_0^T \widehat{X})^\dagger Y_0^T,$$

and since the middle matrix $Y_0^T \widehat{X}(Y_0^T \widehat{X})^\dagger Y_0^T$ is symmetric, the whole matrix is symmetric; thus we have also verified condition iv).

Acknowledgements. I highly appreciate discussions with Michele Benzi and Jim Nagy during the preparation of this manuscript.

REFERENCES

[1] R. T. Behrens and L. L. Scharf, *Signal processing applications of oblique projection operators*, IEEE Trans. Signal Proc., 42 (1994), pp. 1413–1424.
[2] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[3] L. Eldén, *A weighted pseudoinverse, generalized signuar values, and constrained least squares problems*, BIT, 22 (1982), pp. 487–502.
[4] M. Hanke and P. C. Hansen, *Regularization methods for large-scale problems*, Surveys Math. Indust., 3 (1993), pp. 253–315.
[5] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
[6] J. W. Hilgers, *On the equivalence of regularization and certain reproducing kernel Hilbert space appraoches for solving first kind problems*, SIAM J. Numer. Anal., 13 (1976), pp. 172–184.
[7] I. C. F. Ipsen and C. D. Meyer, *The angle between complementary subspaces*, Amer. Math. Monthly, 102 (1995), pp. 904–911.
[8] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Piladephia, 2000.

[2]We use that $(A\,B)^\dagger = B^\dagger A^\dagger$ when A and B have full column and row rank, respectively.