Projection Matrices in Least Squares Fits

Matrices of the form $(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ (or $(X^{\mathrm{T}}X)^{-}X^{\mathrm{T}}$ if X is not of full rank) occur often in linear models. This is the matrix that transforms the response vector y into the least squares estimate of β in the linear model $y = X\beta + \epsilon$, for example. It is a factor of the projection matrix of y onto $\hat{y}$, that is, $X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$, the "hat" matrix. We recall that a matrix is a projection matrix if and only if it is symmetric and idempotent. (That means a projection matrix is necessarily either the identity or it is singular.)

On page 100 of Harrell there is an interesting fact about a matrix $(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$, and the related matrix $(X_1^{\mathrm{T}}X_1)^{-1}X_1^{\mathrm{T}}$, where X_1 is a matrix formed by deleting some of the columns of X; that is,

$$X = [X_1 \mid X_2].$$

The fact is

$$X_1^{\mathrm{T}} = X_1^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}; \tag{1}$$

that is, the hat matrix projects X_1 onto itself. This comes from the last sentence in the second paragraph on page 100, where my X_1 is Harrell's T. (Note that there is a typo in this sentence in Harrell: "$(T^{\mathrm{T}}T)$" should be "$(T^{\mathrm{T}}T)^{-1}$".)

Thinking about this in terms of projections, we first note the very general fact that if P is a projection matrix, P projects P onto P (because P is idempotent). Therefore, a projection matrix projects any columns of itself onto themselves. Equation (1) can be derived by using this fact and partitioning the matrices appropriately. I am not sure how intuitive this is, or how wellknown is this fact.

Charles explained the equation involving the full projection matrices, that is, the equation formed by premultiplying both sides of equation (1) by $(X_1^{\mathrm{T}}X_1)^{-1}$ (that is, by $(T^{\mathrm{T}}T)^{-1}$), in terms of a projection of y onto $\hat{y}$ using a projection matrix formed from the full X followed by a projection of $\hat{y}$ using a projection matrix formed from X_1 (that is, from T). Again, I am not sure how many people had the intuition to see all that. It seems to me more complicated that the argument in the paragraph above, in which there is only one projection matrix, namely $X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$. (In any event, clearly there would have been a problem with the typo if you were paying attention.)

For those who like a more algebraic proof, we can use the famous Schur decomposition of the inverse of a partitioned matrix,

$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array}\right],$$

$$A^{-1} = \left[\begin{array}{c|c} A_{11}^{-1} + A_{11}^{-1}A_{12}Z^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}Z^{-1} \\ & \\ -Z^{-1}A_{21}A_{11}^{-1} & Z^{-1} \end{array}\right],$$

where $Z = A_{22} - A_{21}A_{11}^{-1}A_{12}$.

With $X = [X_1 \mid X_2]$ and $Z = X_2^{\mathrm{T}}X_2 - X_2^{\mathrm{T}}X_1(X_1^{\mathrm{T}}X_1)^{-1}X_1^{\mathrm{T}}X_2$ and remembering that $X_2^{\mathrm{T}}X_1 = X_1^{\mathrm{T}}X_2$, we have

$$\begin{aligned}
&X_1^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} \\
&\quad= X_1^{\mathrm{T}}[X_1 \mid X_2]\left[\begin{array}{c|c} X_1^{\mathrm{T}}X_1 & X_1^{\mathrm{T}}X_2 \\ X_2^{\mathrm{T}}X_1 & X_2^{\mathrm{T}}X_2 \end{array}\right]^{-1}[X_1 \mid X_2]^{\mathrm{T}} \\
&\quad= [X_1^{\mathrm{T}}X_1 \mid X_1^{\mathrm{T}}X_2] \\
&\qquad\left[\begin{array}{c|c} (X_1^{\mathrm{T}}X_1)^{-1} - (X_1^{\mathrm{T}}X_1)^{-1}(X_2^{\mathrm{T}}X_1)Z^{-1}(X_1^{\mathrm{T}}X_2)(X_1^{\mathrm{T}}X_1)^{-1} & -(X_1^{\mathrm{T}}X_1)^{-1}(X_1^{\mathrm{T}}X_2)Z^{-1} \\ -Z^{-1}(X_2^{\mathrm{T}}X_1)(X_1^{\mathrm{T}}X_1)^{-1} & Z^{-1} \end{array}\right] \\
&\qquad\left[\begin{array}{c} X_1^{\mathrm{T}} \\ X_2^{\mathrm{T}} \end{array}\right] \\
&\quad= [I - (X_2^{\mathrm{T}}X_1)Z^{-1}(X_1^{\mathrm{T}}X_2)(X_1^{\mathrm{T}}X_1)^{-1} - X_1^{\mathrm{T}}X_2Z^{-1}(X_2^{\mathrm{T}}X_1)(X_1^{\mathrm{T}}X_1)^{-1} \quad | \quad - X_1^{\mathrm{T}}X_2Z^{-1} + X_1^{\mathrm{T}}X_2Z^{-1}] \\
&\qquad\left[\begin{array}{c} X_1^{\mathrm{T}} \\ X_2^{\mathrm{T}} \end{array}\right] \\
&\quad= X_1^{\mathrm{T}},
\end{aligned}$$

which is equation (1).