**Problem A1.4.** The vector space $\mathscr{L}^2(\Omega, \mathcal{B}, \mathbb{P})$ is endowed with the inner product $\langle \boldsymbol{Z}, \boldsymbol{Y} \rangle_{\mathbb{P}} = \boldsymbol{\Sigma}_{\boldsymbol{ZY}}$. To be explicit, let's adopt the Euclidean heuristic and suppose our sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\}$ is finite, with corresponding probabilities given by the vector $\mathbb{P} = \{p_1, p_2, \ldots, p_{|\Omega|}\}$. (Yes, this is sort of an abuse of notation.)

Then we can think of the random variables $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_{p_{\boldsymbol{Z}}})^T$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_{p_{\boldsymbol{Y}}})^T$ as $p_{\boldsymbol{Z}} \times |\Omega|$ and $p_{\boldsymbol{Y}} \times |\Omega|$ matrices, respectively:

$$\boldsymbol{Z} \;=\; (\boldsymbol{Z}(\omega_1), \boldsymbol{Z}(\omega_2), \ldots, \boldsymbol{Z}(\omega_{|\Omega|})) \;=\; \begin{pmatrix} Z_1(\omega_1) & Z_1(\omega_2) & \cdots & Z_1(\omega_{|\Omega|}) \\ Z_2(\omega_1) & Z_2(\omega_2) & \cdots & Z_2(\omega_{|\Omega|}) \\ \vdots & \vdots & \vdots & \vdots \\ Z_{p_{\boldsymbol{Z}}}(\omega_1) & Z_{p_{\boldsymbol{Z}}}(\omega_2) & \cdots & Z_{p_{\boldsymbol{Z}}}(\omega_{|\Omega|}) \end{pmatrix}$$

$$\boldsymbol{Y} \;=\; (\boldsymbol{Y}(\omega_1), \boldsymbol{Y}(\omega_2), \ldots, \boldsymbol{Y}(\omega_{|\Omega|})) \;=\; \begin{pmatrix} Y_1(\omega_1) & Y_1(\omega_2) & \cdots & Y_1(\omega_{|\Omega|}) \\ Y_2(\omega_1) & Y_2(\omega_2) & \cdots & Y_2(\omega_{|\Omega|}) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{p_{\boldsymbol{Y}}}(\omega_1) & Y_{p_{\boldsymbol{Y}}}(\omega_2) & \cdots & Y_{p_{\boldsymbol{Y}}}(\omega_{|\Omega|}) \end{pmatrix}$$

Consider the "covariance inner product" given by:

$$\begin{aligned} \langle \boldsymbol{Z}, \boldsymbol{Y} \rangle_{\mathbb{P}} &= \boldsymbol{Z}\ \mathsf{diag}(\mathbb{P})\, \boldsymbol{Y}^T \\ &= \sum_{\omega_i \in \Omega} p_i\, \boldsymbol{Z}(\omega_i)\, \boldsymbol{Y}(\omega_i)^T \\ &= \begin{pmatrix} Z_1(\omega_1) & Z_1(\omega_2) & \cdots & Z_1(\omega_{|\Omega|}) \\ Z_2(\omega_1) & Z_2(\omega_2) & \cdots & Z_2(\omega_{|\Omega|}) \\ \vdots & \vdots & \vdots & \vdots \\ Z_{p_{\boldsymbol{Z}}}(\omega_1) & Z_{p_{\boldsymbol{Z}}}(\omega_2) & \cdots & Z_{p_{\boldsymbol{Z}}}(\omega_{|\Omega|}) \end{pmatrix} \begin{pmatrix} p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & p_{|\Omega|} \end{pmatrix} \begin{pmatrix} Y_1(\omega_1) & Y_2(\omega_1) & \cdots & Y_{p_{\boldsymbol{Y}}}(\omega_1) \\ Y_1(\omega_2) & Y_2(\omega_2) & \cdots & Y_{p_{\boldsymbol{Y}}}(\omega_2) \\ \vdots & \vdots & \vdots & \vdots \\ Y_1(\omega_{|\Omega|}) & Y_2(\omega_{|\Omega|}) & \cdots & Y_{p_{\boldsymbol{Y}}}(\omega_{|\Omega|}) \end{pmatrix} \\ &= \boldsymbol{\Sigma}_{\boldsymbol{ZY}}, \end{aligned}$$

which, of course, is itself a $p_{\boldsymbol{Z}} \times p_{\boldsymbol{Y}}$ matrix. Then, $\boldsymbol{\Sigma}_{\boldsymbol{ZY}} = \mathbf{0}$ if and only if $\boldsymbol{Z}$ is uncorrelated with $\boldsymbol{Y}$. With this interpretation, we have that $\mathcal{L}^{\perp}(\boldsymbol{Z}) = \{\boldsymbol{X} : \langle \boldsymbol{Z}, \boldsymbol{X} \rangle_{\mathbb{P}}) = \mathbf{0}\}$ is the set of all random variables that are uncorrelated with $\boldsymbol{Z}$.

**Important Remark:** Note that $\langle \cdot, \cdot \rangle_{\mathbb{P}}$ satisfies all of the usual axioms of an inner product (i.e., *bilinearity*) when $p_{\boldsymbol{Z}} = p_{\boldsymbol{Y}}$, with the generalization that $\succeq \mathbf{0}$ means positive semi-definite. In particular, the "squared norm" of any $\boldsymbol{Z} \in \mathscr{L}^2(\Omega, \mathcal{B}, \mathbb{P})$ is non-negative, since $\langle \boldsymbol{Z}, \boldsymbol{Z} \rangle_{\mathbb{P}} \succeq \mathbf{0}$. ♦

**Problem A2.3.** Let $\tilde{\mathbb{P}}_I$ denote the "identity" projection in the transformed space. That is:

$$\begin{aligned}\tilde{\mathbb{P}}_I &= \tilde{V}(\tilde{V}^T\tilde{V})^{-1}\tilde{V}^T \\ &= A^{1/2}V[(V^TA^{1/2})(A^{1/2}V)]^{-1}V^TA^{1/2} \\ &= A^{1/2}V[(V^TAV)]^{-1}V^TA^{1/2}.\end{aligned}$$

Then for any $\tilde{y} = A^{1/2}y$ in the new space:

$$\begin{aligned}\tilde{\mathbb{P}}_I(\tilde{y}) &= A^{1/2}V[(V^TAV)]^{-1}V^TA^{1/2}(A^{1/2}y) \\ &= A^{1/2}[V(V^TAV)^{-1}V^TAy] \\ &= A^{1/2}\mathbb{P}_A(y) \\ &= \widetilde{\mathbb{P}_A(y)},\end{aligned}$$

which gives the coordinates of the projection $\mathbb{P}_A(y) = \hat{y}$ in terms of the new space. ♦

**Problem A2.7.** Define $\boldsymbol{Y}_2^{\perp} = \boldsymbol{Y}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1$. In this problem, we assume that $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are mean zero. So $\mathbb{E}(\boldsymbol{Y}_2^{\perp}) = \mathbf{0}$.

**(i.)** The covariance of $\boldsymbol{Y}_2^{\perp}$ is then:

$$\begin{aligned}\boldsymbol{\Sigma}_{22}^{\perp} &= \mathbb{E}\{[\boldsymbol{Y}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1][\boldsymbol{Y}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1]^T\} \\ &= \mathbb{E}\{\boldsymbol{Y}_2\boldsymbol{Y}_2^T - \boldsymbol{Y}_2\boldsymbol{Y}_1^T(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1})^T - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1\boldsymbol{Y}_2^T(-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1)(-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1)^T\} \\ &= \mathbb{E}\{\boldsymbol{Y}_2\boldsymbol{Y}_2^T - \boldsymbol{Y}_2\boldsymbol{Y}_1^T(\boldsymbol{\Sigma}_{11}^{-1})^T\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1\boldsymbol{Y}_2^T + (\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1)\boldsymbol{Y}_1^T(\boldsymbol{\Sigma}_{11}^{-1})^T\boldsymbol{\Sigma}_{21}^T\} \\ &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}(\boldsymbol{\Sigma}_{11}^{-1})^T\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} + (\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1})\boldsymbol{\Sigma}_{11}(\boldsymbol{\Sigma}_{11}^{-1})^T\boldsymbol{\Sigma}_{21}^T \\ &\overset{(*)}{=} \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12},\end{aligned}$$

where the identity in $(*)$ follows from the fact that $\boldsymbol{\Sigma}_{11}^{-1}$ is symmetric and that $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T$.

**(ii.)** The Grammian in an $\mathscr{L}^2(\Omega, \mathcal{B}, \mathbb{P})$ space is the covariance matrix. So in terms of the Theorem A.2 (the *Extended Pythagorean Theorem*), $\boldsymbol{\Sigma}_{22}^{\perp}$ is the residual variance/unexplained part/Grammian of $\boldsymbol{Y}_2$ after regressing on $\boldsymbol{Y}_1$ ("$\mathcal{L}_{\mathsf{col}}(\boldsymbol{Y}_1)$"). ♦

**Problem A3.2.** Let $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ be vectors in $\mathbb{R}^2$. Then from the property (iv) of determinants found on page A3.1, we have:

$$\left|\boldsymbol{a}_1,\, \boldsymbol{a}_1 + \boldsymbol{a}_2\right| = \left|\boldsymbol{a}_1,\, \boldsymbol{a}_1\right| + \left|\boldsymbol{a}_1,\, \boldsymbol{a}_2\right| = \left|\boldsymbol{a}_1,\, \boldsymbol{a}_2\right|.$$

Geometrically, this says that the area of the parallelogram in $\mathbb{R}^2$ formed by $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ is the same as the area formed by the vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_1 + \boldsymbol{a}_2\}$. This can be visualized by noting that the resultant vector $\boldsymbol{a}_1 + \boldsymbol{a}_2$ bisects the parallelogram formed by $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$.

Furthermore, note that for any $c \in \mathbb{R}$, this is the same as the area formed by $\{\boldsymbol{a}_1, c\boldsymbol{a}_1 + \boldsymbol{a}_2\}$. ♦

**Problem A3.5.** Suppose that the given matrix is $n \times n$ and define $\mathbf{1}_n = (1, 1, \cdots, 1)^T$ to be a column vector of $n$ ones. So $\mathbf{1}_n \mathbf{1}_n^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$ is an $n \times n$ matrix consisting of all 1's.

$$\begin{aligned}
\begin{vmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & \ddots & \rho \\ \rho & \rho & \rho & 1 \end{vmatrix} &= \left|(1-\rho)\boldsymbol{I}_n + (1-\rho)\frac{\rho}{(1-\rho)}\mathbf{1}_n\mathbf{1}_n^T\right| \\
&\overset{[\text{since } |c\boldsymbol{A}|=c^n|\boldsymbol{A}|]}{=} (1-\rho)^n \left|\boldsymbol{I}_n + \frac{\rho}{(1-\rho)}\mathbf{1}_n\mathbf{1}_n^T\right| \\
&\overset{[\text{Fact(ix)}]}{=} (1-\rho)^n \left|\boldsymbol{I}_1 + \frac{\rho}{(1-\rho)}\mathbf{1}_n^T\mathbf{1}_n\right| \\
&\overset{[\text{since } \mathbf{1}_n^T\mathbf{1}_n = n]}{=} (1-\rho)^n \left(1 + \frac{n\rho}{1-\rho}\right). \quad \blacklozenge
\end{aligned}$$

**Problem A3.6.** We are asked to show that $J(x \to \rho, \theta_1, \ldots, \theta_{n-1}) = \rho^{n-1}(\sin\theta_1)^{n-2}(\sin\theta_2)^{n-3}\cdots(\sin\theta_{n-2})$. This result can be proved inductively or by directly exploiting the almost triangular structure of the Jacobian (i.e, expand the determinant along the last column).

The surface area of the unit sphere is then found by setting $\rho = 1$ and integrating out the angles:

$$\int_{0\le\theta_1\le\pi}\int_{0\le\theta_2\le\pi}\cdots\int_{0\le\theta_{n-2}\le\pi}\int_{0\le\theta_{n-1}\le 2\pi}(\sin\theta_1)^{n-2}\cdots(\sin\theta_{n-2})d\theta_1 d\theta_2,\cdots d\theta_{n-2}d\theta_{n-1}$$

$$= \underbrace{2\pi}_{\text{from }\theta_{n-1}}\prod_{i=1}^{n-2} 2\int_0^{\pi/2}\sin^i\theta\, d\theta = (2\pi)\prod_{i=1}^{n-2} 2\left[\frac{\sqrt{\pi}}{2}\Gamma\left(\frac{i+1}{2}\right)\Big/\Gamma\left(\frac{i+2}{2}\right)\right] = \boxed{2\pi^{n/2}\Big/\Gamma\left(\frac{n}{2}\right)} \quad \blacklozenge$$

**Problem A3.10.** Let $\Omega_{\boldsymbol{X}} = \mathbb{R}^n$ and $S = \sum_{i=1}^n |X_i|$. To compute the integral Jacobian $J(\boldsymbol{x} \to s)$, it suffices to find *any* situation such that probability distribution for $S$ can be expressed as a product of the joint distribution of $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$.

Motivated by the given example, let's suppose that the $X_i$'s are i.i.d.Laplacian. That is, $X_i \overset{\text{i.i.d.}}{\sim} f^{X_i}(x_i) = \frac{1}{2}e^{-|x_i|}$. Then the joint distribution of the $\boldsymbol{X}$ vector is:

$$f^{\boldsymbol{X}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^n \frac{1}{2}e^{-|x_i|} = 2^{-n}e^{-\sum|x_i|} = 2^{-n}e^{-s}$$

But since the $|X_i|$'s are exponential, the sum $S = \sum_{i=1}^{n} |X_i|$ follows a Gamma distribution with density

$$f^S(s) = \frac{s^{n-1} e^{-s}}{\Gamma(n)},$$

and so the integral Jacobian $J(\boldsymbol{x} \to s)$ can be computed by noting:

$$J(\boldsymbol{x} \to s) = \frac{f^S(s)}{f^{\boldsymbol{X}}(x_1, x_2, \ldots, x_n)} = \frac{s^{n-1} e^{-s}/\Gamma(n)}{2^{-n} e^{-s}} = \boxed{\frac{2^n s^{n-1}}{\Gamma(n)}}. \quad \blacklozenge$$

**Problem A4.3.** Let $\overset{\perp}{\boldsymbol{Y}} = (\boldsymbol{Y} - \widehat{\boldsymbol{Y}}) = [(\boldsymbol{Y} - \boldsymbol{\mu_Y}) - \boldsymbol{M}(\boldsymbol{X} - \boldsymbol{\mu_X})]$. Note that $\mathbb{E}(\overset{\perp}{\boldsymbol{Y}}) = \boldsymbol{0}$.

Let $\boldsymbol{\Sigma}^{\perp}$ denote the covariance matrix of $\overset{\perp}{\boldsymbol{Y}}$; that is,

$$\begin{aligned}
\boldsymbol{\Sigma}^{\perp} &\overset{\text{def}}{=} \mathbb{E}[\overset{\perp}{\boldsymbol{Y}} \overset{\perp}{\boldsymbol{Y}}{}^T] \\
&= \mathbb{E}\Big\{[(\boldsymbol{Y} - \boldsymbol{\mu_Y}) - \boldsymbol{M}(\boldsymbol{X} - \boldsymbol{\mu_X})][(\boldsymbol{Y} - \boldsymbol{\mu_Y}) - \boldsymbol{M}(\boldsymbol{X} - \boldsymbol{\mu_X})]^T\Big\} \\
&= \mathbb{E}\Big\{(\boldsymbol{Y} - \boldsymbol{\mu_Y})(\boldsymbol{Y} - \boldsymbol{\mu_Y})^T - (\boldsymbol{Y} - \boldsymbol{\mu_Y})(\boldsymbol{X} - \boldsymbol{\mu_X})^T \boldsymbol{M}^T - \boldsymbol{M}(\boldsymbol{X} - \boldsymbol{\mu_X})(\boldsymbol{Y} - \boldsymbol{\mu_Y})^T\Big\} \\
&\quad + \mathbb{E}\Big\{\boldsymbol{M}(\boldsymbol{X} - \boldsymbol{\mu_X})(\boldsymbol{X} - \boldsymbol{\mu_X})^T \boldsymbol{M}^T\Big\} \\
&= \boldsymbol{\Sigma_{YY}} - \boldsymbol{\Sigma_{YX}} \boldsymbol{M}^T - \boldsymbol{M} \boldsymbol{\Sigma_{XY}} + \boldsymbol{M} \boldsymbol{\Sigma_{XX}} \boldsymbol{M}^T.
\end{aligned}$$

Now take derivatives of the $\mathsf{trace}$ w.r.t. $\boldsymbol{M}$ and set it to $\boldsymbol{0}$.
(The site `http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html` provides a reference on matrix calculus.)

$$\begin{aligned}
\frac{\partial\, \mathsf{trace}(\boldsymbol{\Sigma}^{\perp})}{\partial \boldsymbol{M}} &= \frac{\partial}{\partial \boldsymbol{M}} \mathsf{trace}[\boldsymbol{\Sigma_{YY}} - \boldsymbol{\Sigma_{YX}} \boldsymbol{M}^T - \boldsymbol{M} \boldsymbol{\Sigma_{XY}} + \boldsymbol{M} \boldsymbol{\Sigma_{XX}} \boldsymbol{M}^T] \\
&= -\boldsymbol{\Sigma_{YX}} - \boldsymbol{\Sigma}_{\boldsymbol{XY}}^T + \boldsymbol{M}(\boldsymbol{\Sigma_{XX}} + \boldsymbol{\Sigma}_{\boldsymbol{XX}}^T) \\
&= -\boldsymbol{\Sigma_{YX}} - \boldsymbol{\Sigma_{YX}} + 2\boldsymbol{M} \boldsymbol{\Sigma_{XX}} \\
&= \boldsymbol{0}.
\end{aligned}$$

Solving the equation gives $\boldsymbol{M} = \boldsymbol{\Sigma_{YX}} \boldsymbol{\Sigma}_{\boldsymbol{XX}}^{-1}$. This is the unique solution, since taking the second derivative gives $\frac{\partial^2 \mathsf{trace}(\boldsymbol{\Sigma}^{\perp})}{\partial \boldsymbol{M} \partial \boldsymbol{M}^T} = \boldsymbol{\Sigma_{XX}} \succ \boldsymbol{0}$. This proves that $\widehat{\boldsymbol{Y}} = \boldsymbol{\mu_Y} + \boldsymbol{\Sigma_{YX}} \boldsymbol{\Sigma}_{\boldsymbol{XX}}^{-1}(\boldsymbol{X} - \boldsymbol{\mu_X})$ is the best linear predictor in the sense of minimizing $\mathsf{trace}(\boldsymbol{\Sigma}^{\perp})$. $\blacklozenge$

**Problem A4.3.5** Suppose $\mathsf{rank}(\boldsymbol{X}) = r$ and let $\boldsymbol{X}_{n\times p} = \boldsymbol{U}_{n\times r}\boldsymbol{D}_{r\times r}\boldsymbol{V}^T_{r\times p}$ be the singular value decomposition of $\boldsymbol{X}$.

Then:

**(i.)** $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T \implies \boldsymbol{X}\boldsymbol{V} = (\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T)\boldsymbol{V} \implies \boxed{\boldsymbol{X}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{D} = \boldsymbol{C}_{\mathsf{row}}}$.
That is, $\boldsymbol{C}_{\mathsf{row}}$ gives the mapping $\boldsymbol{X} \mapsto \boldsymbol{X}\boldsymbol{V}$. Let $\vec{\boldsymbol{x}}_j$ for $1 \le j \le n$ denote the $j$th row vector of $\boldsymbol{X}$ (the "$\vec{\boldsymbol{x}}_j$" arrow notation is used to emphasize that it is a row vector). We can then write:

$$\boldsymbol{X}_{n\times p} = \begin{pmatrix} \vec{\boldsymbol{x}}_1 \\ \vec{\boldsymbol{x}}_2 \\ \vdots \\ \vec{\boldsymbol{x}}_n \end{pmatrix} \quad ; \quad \boldsymbol{V}_{p\times r} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r) \quad ; \quad \boldsymbol{C}_{\mathsf{row}} = \boldsymbol{X}\boldsymbol{V} = \begin{pmatrix} \vec{\boldsymbol{x}}_1\boldsymbol{v}_1 & \vec{\boldsymbol{x}}_1\boldsymbol{v}_2 & \cdots & \vec{\boldsymbol{x}}_1\boldsymbol{v}_r \\ \vec{\boldsymbol{x}}_2\boldsymbol{v}_1 & \vec{\boldsymbol{x}}_2\boldsymbol{v}_2 & \cdots & \vec{\boldsymbol{x}}_2\boldsymbol{v}_r \\ \vdots & \vdots & \vdots & \vdots \\ \vec{\boldsymbol{x}}_n\boldsymbol{v}_1 & \vec{\boldsymbol{x}}_n\boldsymbol{v}_2 & \cdots & \vec{\boldsymbol{x}}_n\boldsymbol{v}_r \end{pmatrix},$$

which shows that $\boldsymbol{C}_{\mathsf{row}}$ gives the coordinates of $\boldsymbol{X}$'s rows in terms of the orthogonal basis $\boldsymbol{V}$.

**(ii.)** $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T \implies \boldsymbol{U}^T\boldsymbol{X} = (\boldsymbol{U}^T)\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T \implies \boxed{\boldsymbol{U}^T\boldsymbol{X} = \boldsymbol{D}\boldsymbol{V}^T = \boldsymbol{C}_{\mathsf{col}}}$. Now let $\boldsymbol{x}_k$ for $1 \le k \le p$ denote the $k$th column of $\boldsymbol{X}$ (no arrow notation ("$\rightarrow$") this time). We can then write:

$$\boldsymbol{U}^T_{r\times n} = \begin{pmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_r^T \end{pmatrix} \quad ; \quad \boldsymbol{X}_{n\times p} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p) \quad ;$$

$$\boldsymbol{C}_{\mathsf{col}} = \boldsymbol{U}^T\boldsymbol{X} = \begin{pmatrix} \boldsymbol{u}_1^T\boldsymbol{x}_1 & \boldsymbol{u}_1^T\boldsymbol{x}_2 & \cdots & \boldsymbol{u}_1^T\boldsymbol{x}_p \\ \boldsymbol{u}_2^T\boldsymbol{x}_1 & \boldsymbol{u}_2^T\boldsymbol{x}_2 & \cdots & \boldsymbol{u}_2^T\boldsymbol{x}_p \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{u}_r^T\boldsymbol{x}_1 & \boldsymbol{u}_r^T\boldsymbol{x}_2 & \cdots & \boldsymbol{u}_r^T\boldsymbol{x}_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^T\boldsymbol{u}_1 & \boldsymbol{x}_2^T\boldsymbol{u}_1 & \cdots & \boldsymbol{x}_p^T\boldsymbol{u}_1 \\ \boldsymbol{x}_1^T\boldsymbol{u}_2 & \boldsymbol{x}_2^T\boldsymbol{u}_2 & \cdots & \boldsymbol{x}_p^T\boldsymbol{u}_2 \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{x}_1^T\boldsymbol{u}_r & \boldsymbol{x}_2^T\boldsymbol{u}_r & \cdots & \boldsymbol{x}_p^T\boldsymbol{u}_r \end{pmatrix},$$

which gives the coordinates of $\boldsymbol{X}$'s columns in the orthogonal basis $\boldsymbol{U}$. ♦

**Problem A4.10.** Assume that $\mathsf{rank}(\boldsymbol{X}) = p$ so that $\boldsymbol{G} = \boldsymbol{X}^T\boldsymbol{X}$ is invertible. Now let $\boldsymbol{X}_{n\times p} = \boldsymbol{U}_{n\times p}\boldsymbol{D}_{p\times p}\boldsymbol{V}^T_{p\times p}$ be the singular value decomposition of $\boldsymbol{X}$. Then by our definition of the pseudo-inverse,

$$\boldsymbol{X}^-_{p\times n} = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^T.$$

To prove the identity $\boldsymbol{G}^{-1} = \boldsymbol{X}^-(\boldsymbol{X}^-)^T$, we show that $\boldsymbol{G}[\boldsymbol{X}^-(\boldsymbol{X}^-)^T] = \boldsymbol{I}_p$.

Noting that $\boldsymbol{G} = (\boldsymbol{X}^T\boldsymbol{X}) = \boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}^T$, we have:

$$\boldsymbol{G}[\boldsymbol{X}^-(\boldsymbol{X}^-)^T] = [\boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}^T][(\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^T)(\boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{V}^T)] = \boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}_p,$$

which proves the identity.

**<u>Remark.</u>** Remember that it is *not* always true that $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}$. But since $\boldsymbol{V}$ is itself a $p \times p$ matrix in this problem, we do have that $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}_p$.

Property (vii) can be verified by noting that:

$$||x_j^-||^2 = [\boldsymbol{X}^-(\boldsymbol{X}^-)^T]_{jj} = (\boldsymbol{G}^{-1})_{jj} = (\overset{\perp}{\boldsymbol{X}}{}_j^T\overset{\perp}{\boldsymbol{X}}_j)^{-1} = 1/||\overset{\perp}{\boldsymbol{x}}_j||^2. \quad ♦$$