A USEFUL LEMMAS

The following two lemmas will be useful in deriving the bias and variance terms of the ensemble risk. Their proofs can be found in Section F.

Lemma A.1. *Let $S \subseteq [p]$ be a subset with corresponding selection matrix $\mathbf{S}$, and let $\mathbf{S}^c$ be the selection matrix corresponding to S^c. Then for a random matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ with rows independently drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ such that $n > |S|$, and for any random function $f : \mathbb{R}^{n\times|S|} \to \mathbb{R}^{n\times|S|}$ that $f(\mathbf{XS})$ and $\mathbf{XS}^c$ are independent,*

$$\mathbb{E}_{\mathbf{XS}^c}\left[\mathbf{S}^\top \mathbf{X}^\dagger\right] = (\mathbf{XS})^\dagger \tag{44}$$

and

$$\mathbb{E}_{\mathbf{XS}^c}\left[\mathbf{S}^{c\top}\mathbf{X}^\top f(\mathbf{XS})\mathbf{S}^\top \mathbf{X}^\dagger\right] = \mathbf{0}. \tag{45}$$

Lemma A.2. *Let $T_1, T_2 \subseteq [n]$ be independent random subsets with corresponding selection matrices $\mathbf{T}_1, \mathbf{T}_2$ such that $\mathbb{E}\left[\mathbf{T}_j\mathbf{T}_j^\top\right] = \frac{|T_j|}{n}\mathbf{I}_n$. Then for random matrices $\mathbf{X} \in \mathbb{R}^{n\times p_X}, \mathbf{Y} \in \mathbb{R}^{n\times p_Y}$ independent of T_1 and T_2 with independent and identically distributed rows such that $\mathbf{X}^\top\mathbf{T}_j\mathbf{T}_j^\top\mathbf{X}$ and $\mathbf{Y}^\top\mathbf{T}_j\mathbf{T}_j^\top\mathbf{Y}$ are invertible, and for any matrix $\mathbf{A} \in \mathbb{R}^{p_X\times p_Y}$,*

$$\mathbb{E}_{T_1,T_2}\left[\left(\mathbf{T}_1^\top\mathbf{X}\right)^\dagger \mathbf{T}_1^\top \left(\left(\mathbf{T}_2^\top\mathbf{X}\right)^\dagger \mathbf{T}_2^\top\right)^\top\right] = \left(\mathbf{X}^\top\mathbf{X}\right)^\dagger \tag{46}$$

and

$$\mathbb{E}_{T_1,T_2}\left[\left(\left(\mathbf{T}_1^\top\mathbf{X}\right)^\dagger \mathbf{T}_1^\top\right)^\top \mathbf{A}\left(\mathbf{T}_2^\top\mathbf{Y}\right)^\dagger \mathbf{T}_2^\top\right] = \left(\mathbf{X}^\dagger\right)^\top \mathbf{A}\mathbf{Y}^\dagger. \tag{47}$$

B PROOF OF LEMMA 3.2 (BIAS)

To compute the bias, we need to evaluate terms of the form

$$\mathbb{E}_{\mathbf{X},\mathcal{S},\mathcal{T}}\left\langle \boldsymbol{\beta}\boldsymbol{\beta}^\top, \left(\mathbf{I}_p - \mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger \mathbf{T}_i^\top\mathbf{X}\right)^\top \left(\mathbf{I}_p - \mathbf{S}_j\left(\mathbf{T}_j^\top\mathbf{X}\mathbf{S}_j\right)^\dagger \mathbf{T}_j^\top\mathbf{X}\right)\right\rangle. \tag{48}$$

First, we note that since $\mathbf{S}_i\mathbf{S}_i^\top + \mathbf{S}_i^c\mathbf{S}_i^{c\top} = \mathbf{I}_p$,

$$\begin{aligned}
\mathbf{I}_p - \mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger \mathbf{T}_i^\top\mathbf{X} &= \mathbf{I}_p - \mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger \mathbf{T}_i^\top\mathbf{X}\left(\mathbf{S}_i\mathbf{S}_i^\top + \mathbf{S}_i^c\mathbf{S}_i^{c\top}\right) && (49)\\
&= \mathbf{I}_p - \mathbf{S}_i\mathbf{S}_i^\top - \mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger \mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i^c\mathbf{S}_i^{c\top} && (50)\\
&= \left(\mathbf{I}_p - \mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger \mathbf{T}_i^\top\mathbf{X}\right)\mathbf{S}_i^c\mathbf{S}_i^{c\top}. && (51)
\end{aligned}$$

So, we can equivalently evaluate

$$\mathbb{E}_{\mathbf{X},\mathcal{S},\mathcal{T}}\left\langle \boldsymbol{\beta}\boldsymbol{\beta}^\top, \mathbf{S}_i^c\mathbf{S}_i^{c\top}\left[\mathbf{I}_p - \mathbf{X}^\top\mathbf{T}_i\left(\mathbf{S}_i^\top\mathbf{X}^\top\mathbf{T}_i\right)^\dagger \mathbf{S}_i^\top\right]\left[\mathbf{I}_p - \mathbf{S}_j\left(\mathbf{T}_j^\top\mathbf{X}\mathbf{S}_j\right)^\dagger \mathbf{T}_j^\top\mathbf{X}\right]\mathbf{S}_j^c\mathbf{S}_j^{c\top}\right\rangle. \tag{52}$$

It suffices to evaluate the expectation of the second argument of the inner product:

$$\begin{aligned}
&\mathbb{E}_{\mathbf{X},\mathcal{S},\mathcal{T}}\left[\mathbf{S}_i^c\mathbf{S}_i^{c\top}\left[\mathbf{I}_p - \mathbf{X}^\top\mathbf{T}_i\left(\mathbf{S}_i^\top\mathbf{X}^\top\mathbf{T}_i\right)^\dagger \mathbf{S}_i^\top\right]\left[\mathbf{I}_p - \mathbf{S}_j\left(\mathbf{T}_j^\top\mathbf{X}\mathbf{S}_j\right)^\dagger \mathbf{T}_j^\top\mathbf{X}\right]\mathbf{S}_j^c\mathbf{S}_j^{c\top}\right]\\
&= \mathbb{E}_{\mathbf{X},\mathcal{S},\mathcal{T}}\Big[\mathbf{S}_i^c\mathbf{S}_i^{c\top}\mathbf{X}^\top\mathbf{T}_i\left(\mathbf{S}_i^\top\mathbf{X}^\top\mathbf{T}_i\right)^\dagger \mathbf{S}_i^\top\mathbf{S}_j\left(\mathbf{T}_j^\top\mathbf{X}\mathbf{S}_j\right)^\dagger \mathbf{T}_j^\top\mathbf{X}\mathbf{S}_i^\top\mathbf{S}_j^c\mathbf{S}_j^{c\top}\\
&\quad - \mathbf{S}_i^c\mathbf{S}_i^{c\top}\mathbf{X}^\top\mathbf{T}_i\left(\mathbf{S}_i^\top\mathbf{X}^\top\mathbf{T}_i\right)^\dagger \mathbf{S}_i^\top - \mathbf{S}_j\left(\mathbf{T}_j^\top\mathbf{X}\mathbf{S}_j\right)^\dagger \mathbf{T}_j^\top\mathbf{X}\mathbf{S}_j^c\mathbf{S}_j^{c\top} + \mathbf{S}_i^c\mathbf{S}_i^{c\top}\mathbf{S}_j^c\mathbf{S}_j^{c\top}\Big].
\end{aligned} \tag{53}$$

The second and third terms are zero in expectation. To see this for the second term, observe that $\mathbf{S}_i^{c\top}\mathbf{X}^\top$ and $\mathbf{S}_i^\top\mathbf{X}^\top$ are independent and each zero-mean. An analogous argument applies to the third term. The fourth term is equal to

$$\frac{|S_i^c \cap S_j^c|}{p}\mathbf{I}_p. \tag{54}$$

We now consider the case where $i \neq j$. To evaluate the first term, we first apply Lemma A.2. This simplifies the expression to

$$\mathbb{E}_{\mathbf{X},\mathcal{S}}\left[\mathbf{S}_i^c\mathbf{S}_i^{c\top}\mathbf{X}^\top\left(\mathbf{S}_i^\top\mathbf{X}^\top\right)^\dagger\mathbf{S}_i^\top\mathbf{S}_j\left(\mathbf{X}\mathbf{S}_j\right)^\dagger\mathbf{X}\mathbf{S}_i^\top\mathbf{S}_j^c\mathbf{S}_j^{c\top}\right]. \tag{55}$$

Now let $\mathbf{S}_{i\cap j}$, $\mathbf{S}_{i\setminus j}$, $\mathbf{S}_{j\setminus i}$, and $\mathbf{S}_{i\cup j}^c$ denote the selection matrices corresponding to the sets $S_i \cap S_j$, $S_i \setminus S_j$, $S_j \setminus S_i$, and $S_i^c \cap S_j^c$, respectively. Without loss of generality, consider when $\mathbf{S}_i^c = \begin{bmatrix}\mathbf{S}_{j\setminus i} & \mathbf{S}_{i\cup j}^c\end{bmatrix}$ and $\mathbf{S}_j^c = \begin{bmatrix}\mathbf{S}_{i\setminus j} & \mathbf{S}_{i\cup j}^c\end{bmatrix}$. Then the matrix inside this expectation is of the form

$$\mathbf{S}_i^c\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}\mathbf{S}_j^{c\top}, \tag{56}$$

where

$$\mathbf{A} = \mathbf{S}_{j\setminus i}^\top\mathbf{X}^\top\left(\mathbf{S}_i^\top\mathbf{X}^\top\right)^\dagger\mathbf{S}_i^\top\mathbf{S}_j\left(\mathbf{X}\mathbf{S}_j\right)^\dagger\mathbf{X}\mathbf{S}_{i\setminus j} \tag{57}$$

$$\mathbf{B} = \mathbf{S}_{j\setminus i}^\top\mathbf{X}^\top\left(\mathbf{S}_i^\top\mathbf{X}^\top\right)^\dagger\mathbf{S}_i^\top\mathbf{S}_j\left(\mathbf{X}\mathbf{S}_j\right)^\dagger\mathbf{X}\mathbf{S}_{i\cup j}^c \tag{58}$$

$$\mathbf{C} = \mathbf{S}_{i\cup j}^{c\top}\mathbf{X}^\top\left(\mathbf{S}_i^\top\mathbf{X}^\top\right)^\dagger\mathbf{S}_i^\top\mathbf{S}_j\left(\mathbf{X}\mathbf{S}_j\right)^\dagger\mathbf{X}\mathbf{S}_{i\setminus j} \tag{59}$$

$$\mathbf{D} = \mathbf{S}_{i\cup j}^{c\top}\mathbf{X}^\top\left(\mathbf{S}_i^\top\mathbf{X}^\top\right)^\dagger\mathbf{S}_i^\top\mathbf{S}_j\left(\mathbf{X}\mathbf{S}_j\right)^\dagger\mathbf{X}\mathbf{S}_{i\cup j}^c. \tag{60}$$

In the case of $\mathbf{B}$ and $\mathbf{C}$, because $\mathbf{X}\mathbf{S}_{i\cup j}^c$ is independent of the remainder of the factors, $\mathbb{E}_{\mathbf{X}}[\mathbf{B}]$ and $\mathbb{E}_{\mathbf{X}}[\mathbf{C}]$ are equal to $\mathbf{0}$. By applying the second claim of Lemma A.1, we observe that $\mathbb{E}_{\mathbf{X}}[\mathbf{A}]$ is also equal to $\mathbf{0}$. This leaves

$$\mathbb{E}_{\mathbf{X}}[\mathbf{D}] = \mathbb{E}_{\mathbf{X}}\left[\mathbf{S}_{i\cup j}^{c\top}\mathbf{X}^\top\mathbb{E}_{\mathbf{X}\mathbf{S}_{j\setminus i}}\left[\left(\mathbf{S}_i^\top\mathbf{X}^\top\right)^\dagger\mathbf{S}_i^\top\mathbf{S}_{i\cap j}\right]\mathbb{E}_{\mathbf{X}\mathbf{S}_{j\setminus i}}\left[\mathbf{S}_{i\cap j}^\top\mathbf{S}_j\left(\mathbf{X}\mathbf{S}_j\right)^\dagger\right]\mathbf{X}\mathbf{S}_{i\cup j}^c\right] \tag{61}$$

$$= \mathbb{E}_{\mathbf{X}}\left[\mathbf{S}_{i\cup j}^{c\top}\mathbf{X}^\top\left(\mathbf{X}\mathbf{S}_{i\cap j}\mathbf{S}_{i\cap j}^\top\mathbf{X}^\top\right)^\dagger\mathbf{X}\mathbf{S}_{i\cup j}^c\right]. \tag{62}$$

We can evaluate the expectation of the pseudoinverse on its own since $\mathbf{X}\mathbf{S}_{i\cap j}$ and $\mathbf{X}\mathbf{S}_{i\cup j}^c$ are independent. This matrix has a generalized inverse Wishart distribution with scale matrix $\mathbf{I}_n$ and $|S_i \cap S_j|$ degrees of freedom, which yields

$$\mathbb{E}_{\mathbf{X}}\left[\left(\mathbf{X}\mathbf{S}_{i\cap j}\mathbf{S}_{i\cap j}^\top\mathbf{X}^\top\right)^\dagger\right] = \frac{|S_i \cap S_j|}{n(n-|S_i \cap S_j|-1)}\mathbf{I}_n. \tag{63}$$

This leaves

$$\mathbb{E}_{\mathbf{X}}\left[\mathbf{S}_{i\cup j}^{c\top}\mathbf{X}^\top\left(\frac{|S_i \cap S_j|}{n(n-|S_i \cap S_j|-1)}\mathbf{I}_n\right)\mathbf{X}\mathbf{S}_{i\cup j}^c\right] = \frac{|S_i \cap S_j|}{n-|S_i \cap S_j|-1}\mathbf{I}_{|S_i^c \cap S_j^c|}. \tag{64}$$

Then the expectation in (55) becomes

$$\mathbb{E}_{\mathcal{S}}\left[\frac{|S_i \cap S_j|}{(n-|S_i \cap S_j|-1)}\mathbf{S}_i^c\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{|S_i^c \cap S_j^c|}\end{bmatrix}\mathbf{S}_j^{c\top}\right] = \frac{|S_i \cap S_j||S_i^c \cap S_j^c|}{p(n-|S_i \cap S_j|-1)}\mathbf{I}_p, \tag{65}$$

and combing with (54), we have that the bias is equal to

$$\frac{|S_i^c \cap S_j^c|}{p}\left(1+\frac{|S_i \cap S_j|}{n-|S_i \cap S_j|-1}\right)\|\boldsymbol{\beta}\|_2^2. \tag{66}$$

When $i = j$, by a similar argument, without the need to apply Lemma A.2, it follows that the bias is equal to

$$\frac{|S_i^c|}{p}\left(1+\frac{|S_i|}{|T_i|-|S_i|-1}\right)\|\boldsymbol{\beta}\|_2^2. \tag{67}$$

C PROOF OF LEMMA 3.3 (VARIANCE)

To compute the variance, we need to evaluate the terms of the form

$$\mathbb{E}_{\mathbf{X},\mathcal{T}}\left\langle \mathbf{S}_i \left(\mathbf{T}_i^\top \mathbf{X}\mathbf{S}_i\right)^\dagger \mathbf{T}_i^\top, \mathbf{S}_j \left(\mathbf{T}_j^\top \mathbf{X}\mathbf{S}_j\right)^\dagger \mathbf{T}_j^\top \right\rangle. \tag{68}$$

Let $\mathbf{S}$ be the selection matrix corresponding to the set $S_i \cap S_j$. Then

$$\begin{aligned}
&\mathbb{E}\left\langle \mathbf{S}_i \left(\mathbf{T}_i^\top \mathbf{X}\mathbf{S}_i\right)^\dagger \mathbf{T}_i^\top, \mathbf{S}_j \left(\mathbf{T}_j^\top \mathbf{X}\mathbf{S}_j\right)^\dagger \mathbf{T}_j^\top \right\rangle \\
&= \mathbb{E}\left\langle \mathbf{S}^\top\mathbf{S}_i \left(\mathbf{T}_i^\top \mathbf{X}\mathbf{S}_i\right)^\dagger \mathbf{T}_i^\top, \mathbf{S}^\top\mathbf{S}_j \left(\mathbf{T}_j^\top \mathbf{X}\mathbf{S}_j\right)^\dagger \mathbf{T}_j^\top \right\rangle && (69)\\
&= \mathbb{E}\left\langle \mathbb{E}_{\mathbf{X}^{S_i\setminus S_j}}\left[\mathbf{S}^\top\mathbf{S}_i \left(\mathbf{T}_i^\top \mathbf{X}\mathbf{S}_i\right)^\dagger\right] \mathbf{T}_i^\top, \mathbb{E}_{\mathbf{X}^{S_j\setminus S_i}}\left[\mathbf{S}^\top\mathbf{S}_j \left(\mathbf{T}_j^\top \mathbf{X}\mathbf{S}_j\right)^\dagger\right] \mathbf{T}_j^\top \right\rangle && (70)\\
&= \mathbb{E}\left\langle \left(\mathbf{T}_i^\top \mathbf{X}\mathbf{S}\right)^\dagger \mathbf{T}_i^\top, \left(\mathbf{T}_j^\top \mathbf{X}\mathbf{S}\right)^\dagger \mathbf{T}_j^\top \right\rangle. && (71)
\end{aligned}$$

The equality (71) is the result of two applications of Lemma A.1.

In the case that $i \neq j$, an application of Lemma A.2 simplifies the above to

$$\operatorname{tr}\left(\mathbb{E}_{\mathbf{X}}\left[\left(\mathbf{S}^\top\mathbf{X}^\top\mathbf{X}\mathbf{S}\right)^{-1}\right]\right) = \frac{|S_i \cap S_j|}{n - |S_i \cap S_j| - 1}. \tag{72}$$

The equality comes from $\left(\mathbf{S}^\top\mathbf{X}^\top\mathbf{X}\mathbf{S}\right)^{-1}$ having an inverse Wishart distribution with scale matrix $\mathbf{I}_{|S_i\cap S_j|}$ and n degrees of freedom.

When $i = j$, we obtain a similar result directly without needing Lemma A.2. The above simplifies to

$$\operatorname{tr}\left(\mathbb{E}_{\mathbf{X}}\left[\left(\mathbf{S}_i^\top\mathbf{X}^\top\mathbf{T}_i\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^{-1}\right]\right) = \frac{|S_i|}{|T_i| - |S_i| - 1}. \tag{73}$$

D PROOF OF THEOREM 3.6

We first introduce the result due to Dobriban and Wager (2018). We note again, as we noted in the main text, that in the setting of $\mathbf{\Sigma} = \mathbf{I}_p$, where the optimal ridge regression risk is equal to the estimation error of the minimum mean squared error (MMSE) estimator, results on the value of this quantity predate the result of Dobriban and Wager (2018). We refer the reader, for example, to the wireless communication literature (see, e.g., Tulino and Verdú, 2004). However, Dobriban and Wager (2018) have developed the first results on ridge regression risk for general $\mathbf{\Sigma}$, and their clean theorem statement is simple and straightforward to use, even in the $\mathbf{\Sigma} = \mathbf{I}_p$ case.

Proposition D.1 (from Dobriban and Wager, 2018, Theorem 2.1)**.** *Assume that $\Sigma = \mathbf{I}_p$ and $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, p^{-1}\mathbf{I}_p)$. Then in the limit as $n, p \to \infty$ with $p/n \to \gamma$, we have almost surely that*

$$\inf_\lambda R(\widehat{\boldsymbol{\beta}}_\lambda^{\text{ridge}}) = \frac{1}{2}\left(\frac{\gamma-1}{\gamma} - \sigma^2 + \sqrt{\left(\sigma^2 - \frac{\gamma-1}{\gamma}\right)^2 + 4\sigma^2}\right). \tag{74}$$

We note that this expression is equal to $\sigma^2(R^*(1/\sigma^2, \gamma) - 1)$ in the notation of Dobriban and Wager (2018), where this transformation is necessary because we assume $\|\boldsymbol{\beta}\|_2 = 1$ rather than $\sigma = 1$ and because we evaluate the noise-free risk.

The minimizer of the large ensemble risk should satisfy the first-order optimality condition, so we begin by taking its derivative.

$$\begin{aligned}
\frac{dR_\alpha^{\text{ens}}}{d\alpha} &= \frac{(-2(1-\alpha) + 2\sigma^2\alpha\gamma)(1-\alpha^2\gamma) - ((1-\alpha)^2 + \sigma^2\alpha^2\gamma)(-2\alpha\gamma)}{(1-\alpha^2\gamma)^2} && (75)\\
&= \frac{-\alpha^2\gamma + (\gamma(\sigma^2+1)+1) - 1}{(1-\alpha^2\gamma)^2}. && (76)
\end{aligned}$$

Thus the minimizer α_* should satisfy

$$\alpha_*^2\gamma - \alpha_*(\gamma(\sigma^2+1)+1)+1 = 0. \tag{77}$$

From here, it is simply a matter of cumbersome algebra to show that the choice

$$\alpha_* = \frac{\gamma(\sigma^2+1)+1-\sqrt{(\gamma(\sigma^2+1)+1)^2-4\gamma}}{2\gamma} \tag{78}$$

is the valid root of this quadratic expression and is such that $R^{\text{ens}}_{\alpha_*} = \inf_\lambda R(\widehat{\boldsymbol{\beta}}^{\text{ridge}}_\lambda)$. We here show a slightly more interesting approach, leading to Corollary 3.7. First, we start from (77) and add a root of $\alpha_* = 0$, and then we proceed to manipulate the resulting equation.

$$\alpha_*(\alpha_*^2\gamma - \alpha_*(\gamma(\sigma^2+1)+1)+1) = 0 \tag{79}$$

$$\alpha_* - \alpha_*^2(\gamma(\sigma^2+1)+1) = -\alpha_*^3\gamma \tag{80}$$

$$2\alpha_* - \alpha_*^2(\gamma(\sigma^2+1)+1) = \alpha_*(1-\alpha_*^2\gamma) \tag{81}$$

$$\frac{2\alpha_* - \alpha_*^2(\gamma(\sigma^2+1)+1)}{1-\alpha_*^2\gamma} = \alpha_*. \tag{82}$$

Continuing from this last equation,

$$\alpha_* = \frac{2\alpha_* - \alpha_*^2(\gamma(\sigma^2+1)+1)}{1-\alpha_*^2\gamma} \tag{83}$$

$$= \frac{2\alpha_* - \sigma^2\alpha_*^2\gamma - \alpha^2\gamma - \alpha^2 + 1 - 1}{1-\alpha_*^2\gamma} \tag{84}$$

$$= \frac{1-\alpha^2\gamma - (1-2\alpha_*+\alpha_*^2) - \sigma^2\alpha_*^2\gamma}{1-\alpha_*^2\gamma} \tag{85}$$

$$= 1 - \frac{(1-\alpha_*)^2 + \sigma^2\alpha_*^2\gamma}{1-\alpha_*^2\gamma} \tag{86}$$

$$= 1 - R^{\text{ens}}_{\alpha_*}. \tag{87}$$

Thus, if α_* is a root of (77) or $\alpha_* = 0$, then $R^{\text{ens}}_{\alpha_*} = 1-\alpha_*$. We proceed by checking the larger root of (77), but before doing so, we derive the following equality:

$$(\gamma(\sigma^2+1)+1)^2 - 4\gamma = (\gamma(\sigma^2+1)+1)^2 - (4\gamma^2(\sigma^2+1)+4\gamma) + 4\gamma^2 + 4\sigma^2\gamma^2 \tag{88}$$

$$= (\gamma(\sigma^2+1)+1-2\gamma)^2 + 4\sigma^2\gamma^2 \tag{89}$$

$$= (\gamma(\sigma^2-1)+1)^2 + 4\sigma^2\gamma^2. \tag{90}$$

Now, we observe for the larger root (which we denote as α'_*) that

$$\alpha'_* = \frac{\gamma(\sigma^2+1)+1+\sqrt{(\gamma(\sigma^2-1)+1)^2+4\sigma^2\gamma^2}}{2\gamma} \tag{91}$$

$$\geq \frac{1}{2}\left(\sigma^2+1+\frac{1}{\gamma}+\left|\sigma^2-1+\frac{1}{\gamma}\right|\right) \tag{92}$$

$$= \begin{cases} \sigma^2+\frac{1}{\gamma} & \text{if } \frac{1}{\gamma} > 1-\sigma^2 \\ 1 & \text{if } \frac{1}{\gamma} \leq 1-\sigma^2. \end{cases} \tag{93}$$

Thus the only case where α'_* is a valid hyperparameter choice (that is, $\alpha_* \leq \min\left\{1, \gamma^{-1}\right\}$) is when $\sigma^2 = 0$ and $\gamma = 1$, in which case $\alpha_* = 1$ is a double root of (77). So it suffices to evaluate the smaller root even in that case. Now that we know that α'_* is not conatined in $[0, \min\left\{1, \gamma^{-1}\right\}]$ (except in the aforementioned special case) and that by inspection of R^{ens}_α the asymptote at $\alpha = \gamma^{-1/2}$ is not contained in this interval, if we can show that the smaller root (which we denote simply as α_*) of (77) is contained in this interval, then we know that it is the minimizer of R^{ens}_α.

For the smaller root, it is clear from (78) that $\alpha_* \geq 0$. We show by a series of equivalences that $\alpha_* \leq 1/\gamma$:

$$\alpha_* = \frac{1}{2}\left(\sigma^2 + 1 + \frac{1}{\gamma} - \sqrt{\left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2}\right) \leq \frac{1}{\gamma} \tag{94}$$

$$\iff \sigma^2 + 1 - \frac{1}{\gamma} \leq \sqrt{\left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2} \tag{95}$$

$$\iff \left(\sigma^2 + 1 - \frac{1}{\gamma}\right)^2 \leq \left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2 \tag{96}$$

$$\iff \left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2 - 4\frac{\sigma^2}{\gamma} \leq \left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2 \tag{97}$$

$$\iff \left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2 - 4\frac{\sigma^2}{\gamma} \leq \left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2 \tag{98}$$

$$\iff 0 \leq \frac{\sigma^2}{\gamma}. \tag{99}$$

The last inequality is always true. Further, we note that every equivalence here still holds under strict inequalities, so for $\sigma > 0$, we have that $\alpha_* < \gamma^{-1}$. By a similar argument, we can show that $\alpha_* \leq 1$ and that $\alpha_* < 1$ if and only if $\sigma > 0$. By the form of the derivative in (76), we know that α_*, as the smaller root, is a local minimum, and therefore it must be the minimum of R_α^{ens} on $[0, \min\left\{1, \gamma^{-1}\right\}]$. Evaluating the risk at α_*, we have

$$R_{\alpha_*}^{\text{ens}} = 1 - \alpha_* \tag{100}$$

$$= 1 - \frac{1}{2}\left(\sigma^2 + 1 + \frac{1}{\gamma} - \sqrt{\left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2}\right) \tag{101}$$

$$= \frac{1}{2}\left(1 - \sigma^2 - \frac{1}{\gamma} + \sqrt{\left(\sigma^2 - 1 + \frac{1}{\gamma}\right)^2 + 4\sigma^2}\right) \tag{102}$$

$$= \frac{1}{2}\left(\frac{\gamma - 1}{\gamma} - \sigma^2 + \sqrt{\left(\sigma^2 - \frac{\gamma - 1}{\gamma}\right)^2 + 4\sigma^2}\right) \tag{103}$$

$$= \inf_\lambda R(\widehat{\boldsymbol{\beta}}_\lambda^{\text{ridge}}). \tag{104}$$

E PROOFS OF DISCUSSION RESULTS

E.1 Proof of Equation (35) (μ-scaled Risk)

Under the assumption that $\boldsymbol{\Sigma} = \mathbf{I}_p$, the μ-scaled risk is given by

$$R(\mu\widehat{\boldsymbol{\beta}}^{\text{ens}}) = \left\|\boldsymbol{\beta} - \mu\widehat{\boldsymbol{\beta}}^{\text{ens}}\right\|_2^2 \tag{105}$$

$$= \left\|(1-\mu)\boldsymbol{\beta} + \mu(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{\text{ens}})\right\|_2^2 \tag{106}$$

$$= (1-\mu)^2\|\boldsymbol{\beta}\|_2^2 + 2(1-\mu)\mu\left\langle\boldsymbol{\beta}, \boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{\text{ens}}\right\rangle + \mu^2\left\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{\text{ens}}\right\|_2^2 \tag{107}$$

Examining the inner product, we find that

$$\mathbb{E}_{\mathbf{X},\mathbf{z},\mathcal{S},\mathcal{T}}\left[\left\langle\boldsymbol{\beta}, \boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{\text{ens}}\right\rangle\right] = \left\langle\boldsymbol{\beta}, \mathbb{E}_{\mathbf{X},\mathbf{z},\mathcal{S},\mathcal{T}}\left[\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{\text{ens}}\right]\right\rangle \tag{108}$$

$$= \left\langle\boldsymbol{\beta}, \mathbb{E}_{\mathbf{X},\mathcal{S},\mathcal{T}}\left[\mathbf{I}_p - \frac{1}{k}\sum_{i=1}^k \mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger\mathbf{T}_i^\top\mathbf{X}\right]\boldsymbol{\beta}\right\rangle, \tag{109}$$

where the equation (109) holds because $\mathbb{E}[\mathbf{z}] = \mathbf{0}$. Because the subsamplings are identically distributed, we have

$$\mathbb{E}_{\mathbf{X},\mathcal{S},\mathcal{T}}\left[\mathbf{I}_p - \frac{1}{k}\sum_{i=1}^{k}\mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger\mathbf{T}_i^\top\mathbf{X}\right] = \mathbf{I}_p - \mathbb{E}_{\mathbf{X},\mathcal{S},\mathcal{T}}\left[\mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger\mathbf{T}_i^\top\mathbf{X}\right] \tag{110}$$

$$= \mathbf{I}_p - \mathbb{E}_{\mathbf{X},\mathcal{S},\mathcal{T}}\left[\mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger\mathbf{T}_i^\top\mathbf{X}\left(\mathbf{S}_i\mathbf{S}_i^\top + \mathbf{S}_i^c\mathbf{S}_i^{c\top}\right)\right] \tag{111}$$

$$= \mathbf{I}_p - \mathbb{E}_{\mathcal{S}}\left[\mathbf{S}_i\left(\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\right)^\dagger\mathbf{T}_i^\top\mathbf{X}\mathbf{S}_i\mathbf{S}_i^\top\right] \tag{112}$$

$$= \mathbf{I}_p - \mathbb{E}_{\mathcal{S}}\left[\mathbf{S}_i\mathbf{S}_i^\top\right] \tag{113}$$

$$= (1-\alpha)\mathbf{I}_p, \tag{114}$$

where the equation (112) holds because $\mathbb{E}[\mathbf{X}\mathbf{S}_i^c] = \mathbf{0}$. Thus

$$\mathbb{E}_{\mathbf{X},\mathbf{z},\mathcal{S},\mathcal{T}}[R(\mu\widehat{\boldsymbol{\beta}}^{\text{ens}})] = (1-\mu)^2\|\boldsymbol{\beta}\|_2^2 + 2(1-\mu)\mu\mathbb{E}_{\mathbf{X},\mathbf{z},\mathcal{S},\mathcal{T}}\left[\left\langle\boldsymbol{\beta}, \boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{\text{ens}}\right\rangle\right] + \mu^2\mathbb{E}_{\mathbf{X},\mathbf{z},\mathcal{S},\mathcal{T}}\left[\left\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{\text{ens}}\right\|_2^2\right] \tag{115}$$

$$= (1-\mu)^2 + 2(1-\mu)\mu(1-\alpha) + \mu^2 R_\alpha^{\text{ens}}, \tag{116}$$

where the last equality holds because $\langle\boldsymbol{\beta},\boldsymbol{\beta}\rangle = \|\boldsymbol{\beta}\|_2^2 = 1$.

E.2 Proof of Equation (38) (Generalized Dropout)

For $k \to \infty$, dropout minimizes the expected loss:

$$\mathbb{E}_{S_i}\left[\ell_i(\boldsymbol{\beta}')\right] = \mathbb{E}_{S_i}\left[\left\|\mathbf{X}\mathbf{S}_i\mathbf{S}_i^\top\boldsymbol{\beta}' - \mathbf{y}\right\|_2^2\right]. \tag{117}$$

The expected loss is convex in $\boldsymbol{\beta}'$, so we can find its minimizer by the first order optimality condition:

$$\nabla_{\boldsymbol{\beta}'}\mathbb{E}_{S_i}\left[\ell_i(\boldsymbol{\beta}')\right] = \mathbb{E}_{S_i}\left[\mathbf{S}_i\mathbf{S}_i^\top\mathbf{X}^\top\left(\mathbf{X}\mathbf{S}_i\mathbf{S}_i^\top\boldsymbol{\beta}' - \mathbf{y}\right)\right] = 0 \tag{118}$$

Thus,

$$\widehat{\boldsymbol{\beta}} = \left(\mathbb{E}_{S_i}\left[\mathbf{S}_i\mathbf{S}_i^\top\mathbf{X}^\top\mathbf{X}\mathbf{S}_i\mathbf{S}_i^\top\right]\right)^{-1}\mathbb{E}_{S_i}\left[\mathbf{S}_i\mathbf{S}_i^\top\right]\mathbf{X}^\top\mathbf{y}. \tag{119}$$

Turning first to the inverse, consider that

$$\left[\mathbb{E}_{S_i}\left[\mathbf{S}_i\mathbf{S}_i^\top\mathbf{X}^\top\mathbf{X}\mathbf{S}_i\mathbf{S}_i^\top\right]\right]_{j\ell} = \left[\mathbf{X}^\top\mathbf{X}\right]_{j\ell}\Pr(j \in S_i, \ell \in S_i), \tag{120}$$

and that

$$\Pr(j \in S_i, \ell \in S_i) = \begin{cases}\alpha_j & \text{if } j = \ell,\\ \alpha_j\alpha_\ell & \text{otherwise.}\end{cases} \tag{121}$$

This gives us

$$\mathbb{E}_{S_i}\left[\mathbf{S}_i\mathbf{S}_i^\top\mathbf{X}^\top\mathbf{X}\mathbf{S}_i\mathbf{S}_i^\top\right] = \mathbf{A}\mathbf{X}^\top\mathbf{X}\mathbf{A} + \text{diag}(\mathbf{X}^\top\mathbf{X})(\mathbf{A} - \mathbf{A}^2), \tag{122}$$

where $\mathbf{A} = \text{diag}(\boldsymbol{\alpha})$. By a similar and simpler argument,

$$\mathbb{E}_{S_i}\left[\mathbf{S}_i\mathbf{S}_i^\top\right] = \mathbf{A}, \tag{123}$$

which all together yields

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{A}\mathbf{X}^\top\mathbf{X}\mathbf{A} + \text{diag}(\mathbf{X}^\top\mathbf{X})(\mathbf{I}_p - \mathbf{A})\mathbf{A}\right)^{-1}\mathbf{A}\mathbf{X}^\top\mathbf{y} \tag{124}$$

$$= \mathbf{A}^{-1}\left(\mathbf{X}^\top\mathbf{X} + \mathbf{A}^{-1}\text{diag}(\mathbf{X}^\top\mathbf{X})(\mathbf{I}_p - \mathbf{A})\right)^{-1}\mathbf{X}^\top\mathbf{y}. \tag{125}$$

F PROOFS OF LEMMAS A.1 and A.2

F.1 Proof of Lemma A.1

Without loss of generality, let $\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix} = \mathbf{X}$, such that $\mathbf{X}_2 = \mathbf{X}\mathbf{S}$. Let $\begin{bmatrix}\mathbf{Y}_1 \\ \mathbf{Y}_2\end{bmatrix} = \mathbf{X}^\dagger$ be a partitioning of the pseudo-inverse of $\mathbf{X}$ in the same manner, such that $\mathbf{Y}_2 = \mathbf{S}^\top\mathbf{X}^\dagger = \mathbf{S}^\top\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$. Then the Gram matrix can be written as

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix}\mathbf{X}_1^\top\mathbf{X}_1 & \mathbf{X}_1^\top\mathbf{X}_2 \\ \mathbf{X}_2^\top\mathbf{X}_1 & \mathbf{X}_2^\top\mathbf{X}_2\end{bmatrix}, \tag{126}$$

and using block matrix inversion, the inverse admits the form $\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} = \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix}$. The relevant quantities are

$$\mathbf{C} = -\mathbf{D}\mathbf{X}_2^\top\mathbf{X}_1\left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1} \tag{127}$$

$$\mathbf{D} = \left(\mathbf{X}_2^\top\mathbf{X}_2 - \mathbf{X}_2^\top\mathbf{X}_1\left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top\mathbf{X}_2\right)^{-1} \tag{128}$$

$$= \left(\mathbf{X}_2^\top\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_1^\top)}\mathbf{X}_2\right)^{-1}, \tag{129}$$

Where $\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_1^\top)} \triangleq \mathbf{I}_n - \left(\mathbf{X}_1^\top\right)^\dagger\mathbf{X}_1^\top$ denotes the projection onto the column space of $\mathbf{X}_1$. This gives

$$\mathbf{Y}_2 = \mathbf{C}\mathbf{X}_1^\top + \mathbf{D}\mathbf{X}_2^\top \tag{130}$$

$$= \mathbf{D}\mathbf{X}_2^\top\left(\mathbf{I}_n - \mathbf{X}_1\left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top\right) \tag{131}$$

$$= \mathbf{D}\mathbf{X}_2^\top\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_1^\top)}\left(\mathbf{X}_2\mathbf{X}_2^\dagger + \mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_2^\top)}\right) \tag{132}$$

$$= \mathbf{X}_2^\dagger + \mathbf{D}\mathbf{X}_2^\top\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_1^\top)}\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_2^\top)}. \tag{133}$$

Let $\mathbf{U}$, $\mathbf{U}_*$, and $\mathbf{V}$ be the matrices containing the left singular vectors of $\mathbf{X}_2$, $\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_2^\top)}$, and $\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_1^\top)}$, respectively. Because the rows of $\mathbf{X}$ are independently drawn from a spherical Gaussian distribution, $\mathbf{V}$ has a uniform distribution over orthogonal matrices in $\mathbb{R}^{n\times|S^c|}$. As such, $\mathbb{E}_{\mathbf{V}}\left[\mathbf{V}^\top\mathbf{U}_*|\mathbf{V}^\top\mathbf{U}\right] = \mathbf{0}$. Then

$$\mathbb{E}_{\mathbf{X}_1}\left[\mathbf{D}\mathbf{X}_2^\top\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_1^\top)}\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_2^\top)}\right] = \mathbb{E}_{\mathbf{V}}\left[\left(\mathbf{X}_2^\top\mathbf{V}\mathbf{V}^\top\mathbf{X}_2\right)^{-1}\mathbf{X}_2^\top\mathbf{V}\mathbf{V}^\top\mathbf{U}_*\mathbf{U}_*^\top\right] \tag{134}$$

$$= \mathbb{E}_{\mathbf{V}}\left[\left(\mathbf{X}_2^\top\mathbf{V}\mathbf{V}^\top\mathbf{X}_2\right)^{-1}\mathbf{X}_2^\top\mathbf{V}\mathbb{E}_{\mathbf{V}}\left[\mathbf{V}^\top\mathbf{U}_*\middle|\mathbf{V}^\top\mathbf{U}\right]\mathbf{U}_*^\top\right] \tag{135}$$

$$= \mathbf{0}, \tag{136}$$

which combined with (133) yields the first claim.

For the second claim, let $\mathbf{V}_*$ denote the left singular vectors of $\mathbf{X}_1$, and observe that $\mathbb{E}_{\mathbf{V}}\left[\mathbf{V}^\top\mathbf{U}_*|\mathbf{V}^\top\mathbf{U}, \mathbf{V}_*\right] = \mathbf{0}$. Then using similar arguments,

$$\mathbb{E}_{\mathbf{X}_1}\left[\mathbf{X}_1^\top f(\mathbf{X}_2)\mathbf{S}^\top\mathbf{X}^\dagger\right]$$

$$= \mathbb{E}_{\mathbf{X}_1}\left[\mathbf{X}_1^\top f(\mathbf{X}_2)\mathbf{S}^\top\left(\mathbf{X}_2^\dagger + \mathbf{D}\mathbf{X}_2^\top\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_1^\top)}\mathbf{\Pi}_{\mathrm{Null}(\mathbf{X}_2^\top)}\right)\right] \tag{137}$$

$$= \mathbb{E}_{\mathbf{X}_1}\left[\mathbf{X}_1^\top f(\mathbf{X}_2)\mathbf{S}^\top\left(\mathbf{X}_2^\dagger + \left(\mathbf{X}_2^\top\mathbf{V}\mathbf{V}^\top\mathbf{X}_2\right)^{-1}\mathbf{X}_2^\top\mathbf{V}\mathbb{E}_{\mathbf{V}}\left[\mathbf{V}^\top\mathbf{U}_*\middle|\mathbf{V}^\top\mathbf{U}, \mathbf{V}_*\right]\mathbf{U}_*^\top\right)\right] \tag{138}$$

$$= \mathbb{E}_{\mathbf{X}_1}\left[\mathbf{X}_1^\top f(\mathbf{X}_2)\mathbf{S}^\top\mathbf{X}_2^\dagger\right] \tag{139}$$

$$= \mathbf{0}. \tag{140}$$

F.2 Proof of Lemma A.2

Define $\mathbf{\Pi}_{\text{Null}(\mathbf{X}^\top)} \triangleq \mathbf{I}_n - \left(\mathbf{X}^\top\right)^\dagger \mathbf{X}^\top$, the projection operator onto the null space of $\mathbf{X}^\top$. Then for the first claim,

$$
\begin{aligned}
&\mathbb{E}_{T_1,T_2}\left[\left(\mathbf{T}_1^\top\mathbf{X}\right)^\dagger \mathbf{T}_1^\top \left(\left(\mathbf{T}_2^\top\mathbf{X}\right)^\dagger \mathbf{T}_2^\top\right)^\top\right] \\
&= \mathbb{E}_{T_1,T_2}\left[\left(\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{X}\right)^{-1}\right] && (141)\\
&= \mathbb{E}_{T_1,T_2}\left[\left(\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\left(\mathbf{X}\left(\mathbf{X}\mathbf{X}^\top\right)^\dagger\mathbf{X}^\top + \mathbf{\Pi}_{\text{Null}(\mathbf{X}^\top)}\right)\mathbf{T}_2\mathbf{T}_2^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{X}\right)^{-1}\right] && (142)\\
&= \left(\mathbf{X}\mathbf{X}^\top\right)^\dagger + \mathbb{E}_{T_1,T_2}\left[\left(\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\mathbf{\Pi}_{\text{Null}(\mathbf{X}^\top)}\mathbf{T}_2\mathbf{T}_2^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{X}\right)^{-1}\right] && (143)\\
&= \left(\mathbf{X}\mathbf{X}^\top\right)^\dagger + \frac{|T_1||T_2|}{n^2}\left(\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{\Pi}_{\text{Null}(\mathbf{X}^\top)}\mathbf{X}\left(\mathbf{X}^\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{X}\right)^{-1} && (144)\\
&= \left(\mathbf{X}^\top\mathbf{X}\right)^\dagger. && (145)
\end{aligned}
$$

The equality (144) follows due the fact that, because of the distributional assumption on the rows of $\mathbf{X}$, $\mathbf{X}^\top\mathbf{T}_j\mathbf{T}_j^\top\mathbf{X}$ and $\mathbf{T}_j\mathbf{T}_j^\top$ are conditionally independent given $|T_j|$. The equality (145) follows because $\mathbf{X}^\top\mathbf{\Pi}_{\text{Null}(\mathbf{X}^\top)} = \mathbf{0}$.

For the second claim,

$$
\begin{aligned}
&\mathbb{E}_{T_1,T_2}\left[\left(\left(\mathbf{T}_1^\top\mathbf{X}\right)^\dagger\mathbf{T}_1^\top\right)^\top\mathbf{A}\left(\mathbf{T}_2^\top\mathbf{Y}\right)^\dagger\mathbf{T}_2^\top\right] \\
&= \mathbb{E}_{T_1,T_2}\left[\mathbf{T}_1\mathbf{T}_1^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\mathbf{X}\right)^{-1}\mathbf{A}\left(\mathbf{Y}^\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{Y}\right)^{-1}\mathbf{Y}^\top\mathbf{T}_2\mathbf{T}_2^\top\right] && (146)\\
&= \mathbb{E}_{T_1,T_2}\left[\left(\left(\mathbf{X}^\top\right)^\dagger\mathbf{X}^\top + \mathbf{\Pi}_{\text{Null}(\mathbf{X}^\top)}\right)\mathbf{T}_1\mathbf{T}_1^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{T}_1\mathbf{T}_1^\top\mathbf{X}\right)^{-1}\mathbf{A}\left(\mathbf{Y}\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{Y}\right)^{-1}\mathbf{Y}^\top\mathbf{T}_2\mathbf{T}_2^\top\right] && (147)\\
&= \mathbb{E}_{\Pi(T_2)}\left[\left(\mathbf{X}^\dagger\right)^\top\mathbf{A}\left(\mathbf{Y}^\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{Y}\right)^{-1}\mathbf{Y}^\top\mathbf{T}_2\mathbf{T}_2^\top\right] && (148)\\
&= \mathbb{E}_{\Pi(T_2)}\left[\left(\mathbf{X}^\dagger\right)^\top\mathbf{A}\left(\mathbf{Y}^\top\mathbf{T}_2\mathbf{T}_2^\top\mathbf{Y}\right)^{-1}\mathbf{Y}^\top\mathbf{T}_2\mathbf{T}_2^\top\left(\mathbf{Y}\mathbf{Y}^\dagger + \mathbf{\Pi}_{\text{Null}(\mathbf{Y}^\top)}\right)\right] && (149)\\
&= \left(\mathbf{X}^\dagger\right)^\top\mathbf{A}\mathbf{Y}^\dagger, && (150)
\end{aligned}
$$

where the equations (148) and (150) follow by similar arguments to those used to show the first claim.