we use

Woodbury matrix identity

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

and

$$(A^T)^{-1} = (A^{-1})^T, \quad (AB)^T = B^T A^T, \quad (A+B)^T = A^T + B^T$$

to proove the Lemma.

left side

$$(\tau \mathbb{1}_D + X^T X)^{-1} = (\tau \mathbb{1}_D + X^T \mathbb{1}_N X)^{-1} \qquad X \in \mathbb{R}^{N \times D}, \quad X^T \in \mathbb{R}^{D \times N}$$

Woodbury matrix identity

$$= \frac{1}{\tau}\mathbb{1}_D - \frac{1}{\tau}\mathbb{1}_D X^T \left(\mathbb{1}_N + X \frac{1}{\tau}\mathbb{1}_D X^T\right)^{-1} X \frac{1}{\tau}\mathbb{1}_D$$

Complete:

$$(\tau \mathbb{1}_D + X^T X)^{-1} X^T = \frac{1}{\tau} X^T - \frac{1}{\tau}\mathbb{1}_D X^T \left(\mathbb{1}_N + X \frac{1}{\tau}\mathbb{1}_D X^T\right)^{-1} \frac{1}{\tau} X X^T$$

Right side

$$\left(XX^T + \tau \mathbb{1}_N\right)^{-1} = \left(\tau \mathbb{1}_N + X \mathbb{1}_D X^T\right)^{-1}$$

$$\overset{\text{Woodbury matrix identity}}{=} \frac{1}{\tau}\mathbb{1}_N - \frac{1}{\tau}\mathbb{1}_N X \left(\mathbb{1}_D + X^T \frac{1}{\tau}\mathbb{1}_N X\right)^{-1} X^T \frac{1}{\tau}\mathbb{1}_N$$

Complete:

$$X^T\left(XX^T + \tau \mathbb{1}_N\right)^{-1}$$

$$= \frac{1}{\tau}X^T - \frac{1}{\tau}X^T X \left(\mathbb{1}_D + X^T \frac{1}{\tau}\mathbb{1}_N X\right)^{-1} X^T \frac{1}{\tau}$$

together

$$\cancel{\frac{1}{\tau}X^T} + \cancel{\frac{1}{\tau^2}} X^T \left(\mathbb{1}_N + X \frac{1}{\tau}\mathbb{1}_D X^T\right)^{-1} X \cancel{X^T} = \cancel{\frac{1}{\tau}} X^T X \left(\mathbb{1}_D + X^T \frac{1}{\tau}\mathbb{1}_N X\right)^{-1} \cancel{X^T} \cancel{- \frac{1}{\tau}X^T}$$

$$\left(\mathbb{1}_N + \frac{1}{\tau}XX^T\right)^{-1} X = X\left(\mathbb{1}_D + \frac{1}{\tau}X^T X\right)^{-1}$$

$$\cancel{\frac{1}{\tau}}\left(\tau \mathbb{1}_N + XX^T\right)^{-1} X = X\left(\tau \mathbb{1}_D + X^T X\right)^{-1} \cancel{\frac{1}{\tau}}$$

Now we use the [illegible] identities $(A^{-1})^T = (A^T)^{-1}$; $(AB)^T = B^T A^T$

$$\left[(\tau \mathbb{1}_N + X X^T)^{-1} X\right]^T = \left[X(\tau \mathbb{1}_D + X^T X)^{-1}\right]^T$$

$$X^T\left[(\tau \mathbb{1}_N + X X^T)^T\right]^{-1} = \left[(\tau \mathbb{1}_D + X^T X)^T\right]^{-1} X^T$$

This gives with the identities $(A+B)^T = A^T + B^T$

$\mathbb{1}^T = \mathbb{1}$

$$X^T(\tau \mathbb{1}_N + (X^T)^T X^T)^{-1} = (\tau \mathbb{1}_D + X^T (X^T)^T)^{-1} X^T$$

$\square$

The optimal $\hat{a}$ of the dual formulation of the ridge regression problem is given by

$$\hat{a} = (XX^T + \tau \mathbb{1}_N)^{-1} \vec{y}$$

To which there exist a β such that

$$\beta = X^T a = X^T (XX^T + \tau \mathbb{1}_N)^{-1} \vec{y}$$

Using the Lemma above we can write β as

$$\beta = (\tau \mathbb{1}_D + X^T X) X^T \vec{y}$$

which is just $\hat{\beta}$ of the primal formulation

$\square$