Proposition 55.1. *The limit of the matrix* $(X^\top X + KI_n)^{-1}X^\top$ *when* $K > 0$ *goes to zero is the pseudo-inverse* X^+ *of* X.

Proof. To show this let $X = V\Sigma U^\top$ be a SVD of X. Then

$$(X^\top X + KI_n) = U\Sigma^\top V^\top V\Sigma U^\top + KI_n = U(\Sigma^\top\Sigma + KI_n)U^\top,$$

so

$$(X^\top X + KI_n)^{-1}X^\top = U(\Sigma^\top\Sigma + KI_n)^{-1}U^\top U\Sigma^\top V^\top = U(\Sigma^\top\Sigma + KI_n)^{-1}\Sigma^\top V^\top.$$

The diagonal entries in the matrix $(\Sigma^\top\Sigma + KI_n)^{-1}\Sigma^\top$ are

$$\frac{\sigma_i}{\sigma_i^2 + K}, \quad \text{if } \sigma_i > 0,$$

and zero if $\sigma_i = 0$. All nondiagonal entries are zero. When $\sigma_i > 0$ and $K > 0$ goes to 0,

$$\lim_{K\mapsto 0} \frac{\sigma_i}{\sigma_i^2 + K} = \sigma_i^{-1},$$

so

$$\lim_{K\mapsto 0}(\Sigma^\top\Sigma + KI_n)^{-1}\Sigma^\top = \Sigma^+,$$

which implies that

$$\lim_{K\mapsto 0}(X^\top X + KI_n)^{-1}X^\top = X^+.$$

□

The dual function of the first formulation of our problem is a constant function (with value the minimum of J) so it is not useful, but the second formulation of our problem yields an interesting dual problem. The Lagrangian is

$$\begin{aligned} L(\xi, w, \lambda) &= \xi^\top\xi + Kw^\top w + (y - Xw - \xi)^\top\lambda \\ &= \xi^\top\xi + Kw^\top w - w^\top X^\top\lambda - \xi^\top\lambda + \lambda^\top y, \end{aligned}$$

with $\lambda, \xi, y \in \mathbb{R}^m$. The Lagrangian $L(\xi, w, \lambda)$, as a function of ξ and w with λ held fixed, is obviously convex, in fact strictly convex.

To derive the dual function $G(\lambda)$ we minimize $L(\xi, w, \lambda)$ with respect to ξ and w. Since $L(\xi, w, \lambda)$ is (strictly) convex as a function of ξ and w, by Theorem 40.13(4), it has a minimum iff its gradient $\nabla L_{\xi,w}$ is zero (in fact, by Theorem 40.13(2), a unique minimum since the function is strictly convex). Since

$$\nabla L_{\xi,w} = \begin{pmatrix} 2\xi - \lambda \\ 2Kw - X^\top\lambda \end{pmatrix},$$

we get

$$\begin{aligned} \lambda &= 2\xi \\ w &= \frac{1}{2K}X^\top\lambda = X^\top\frac{\xi}{K}. \end{aligned}$$