Source task and conventions

Derive equation (18.15) with $X = RV^\top$ and $V^\top V = I$.

Conventional solution and proof

The normal equation is $(X^\top X + \lambda I)\widehat{\beta} = X^\top y$. Projecting onto the orthogonal complement of $\operatorname{ran} V$ gives $\lambda\widehat{\beta}_\perp = 0$, so for $\lambda > 0$, $\widehat{\beta} = V\theta$. Substitution gives $V(R^\top R + \lambda I)\theta = VR^\top y$, and multiplication by $V^\top$ yields

$$\boxed{\widehat{\beta} = V(R^\top R + \lambda I)^{-1} R^\top y.}$$

Equivalently, the original first-order equation implies $\widehat{\beta} = \lambda^{-1} X^\top (y - X\widehat{\beta})$, proving the source hint directly. The reduced shifted matrix is positive definite, so its solution is unique.

Structural or analytic derivation

The data rows occupy at most N dimensions of $\mathbb{R}^p$. Every orthogonal coefficient component changes only the penalty, so it is zero at the optimum. This is the finite-dimensional form of the representer argument.

Comparative cost and preferred method

A thin factorization costs $O(pN^2)$ classically when $p \gg N$, followed by an N-dimensional solve. This replaces the unnecessary $p \times p$ normal matrix and its $O(p^3)$ factorization. Constructing the final p-vector still costs $O(pN)$ and cannot be omitted when it is the requested output.