Conventional solution and proof

Assume the source covariance convention $\beta \sim N(0, \tau I)$, $\tau > 0$, and $y \mid \beta \sim N(X\beta, \sigma^2 I)$, $\sigma^2 > 0$. Up to a constant, minus twice the posterior log density is

$$\sigma^{-2}\|y - X\beta\|^2 + \tau^{-1}\|\beta\|^2.$$

Let $A = X^\top X/\sigma^2 + I/\tau \succ 0$ and $b = X^\top y/\sigma^2$. Completing the square gives $\beta^\top A\beta - 2b^\top\beta = (\beta - A^{-1}b)^\top A(\beta - A^{-1}b) - b^\top A^{-1}b$. Thus

$$\beta \mid y \sim N(A^{-1}b, A^{-1}), \qquad E(\beta \mid y) = \text{mode}(\beta \mid y) = (X^\top X + \lambda I)^{-1}X^\top y, \quad \lambda = \sigma^2/\tau.$$

The mean equals the mode because the Gaussian density is symmetric about its centre and its positive-definite quadratic exponent is uniquely minimized there. If the prior is parameterized with covariance $\tau^2 I$, the denominator is τ^2 instead; that is not the source convention.

Structural or analytic derivation

The posterior precision is the sum of data precision and prior precision. This makes the regularizing effect transparent even when X is rank deficient, because I/τ remains positive definite.

Comparative cost and preferred method

One factorization gives the posterior mean and permits covariance actions. Forming the entire posterior covariance has additional $O(p^3)$ classical work and $O(p^2)$ output storage. Completing the square is already the standard optimal derivation; an analytic reinterpretation supplies no separate computational gain.