Structural or analytic derivation

Projection removes directions invisible to all observations. A quadratic norm is orthogonally invariant, which is precisely why this reduction is exact. An ℓ_1 penalty has no corresponding rotation invariance; substituting it invalidates the argument.

Comparative cost and preferred method

The parameter dimension drops from p to $r = \operatorname{rank} X \leq N$. The projection costs $O(pN^2)$ for a dense thin factorization; subsequent loss evaluations and gradients depend on r, not p. The theorem certifies equivalence for any supplied loss, but it does not make a nonconvex reduced loss globally easy to optimize.