starting from zero. Hence $\|\beta_{k+1}\|^2 - \|\beta_k\|^2 = 2a_k\beta_k^\top p_k + a_k^2\|p_k\|^2 > 0$ until convergence. The norm property therefore holds for this standard PLS sequence, although its individual spectral factors need not all be shrinkage factors.

Structural or analytic derivation

Ridge diagonalization proves monotonicity termwise; lasso needs its own norm and active-set analysis; PLS uses Krylov orthogonality. The three estimators cannot be placed under one unproved scalar-shrinkage argument.

Comparative cost and preferred method

The ridge path can be evaluated after one SVD, with $O(r)$ spectral work per parameter and $O(pr)$ work if full coefficient vectors are required. The lasso counterexample is exact rational arithmetic. The PLS proof concerns iteration count and norm, not a guarantee that every further step improves prediction. No universal preference among the three statistical procedures follows.