Orthogonality converts the objective into independent scalar problems. Hard selection, linear shrinkage and soft thresholding follow from three different feasible-set/penalty contracts; none can be substituted for another without changing the optimization problem.

Comparative cost and preferred method

After b is formed, ridge and lasso require $O(p)$ scalar operations. Subset selection can use a linear-time order-statistic algorithm plus a pass, or sorting at $O(p \log p)$, instead of enumerating $\binom{p}{M}$ supports. The orthogonal-design assumption is indispensable; general best-subset regression does not decompose this way.