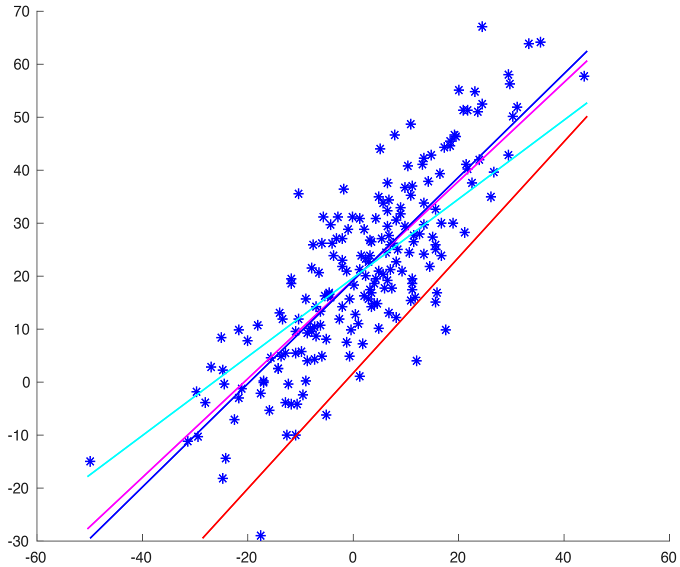

Figure 55.11: Comparison of the four methods with $K = 1000, \tau = 10000$.

55.6 Elastic Net Regression

The lasso method is unsatisfactory when n (the dimension of the data) is much larger than the number m of data, because it only selects m coordinates and sets the others to values close to zero. It also has problems with groups of highly correlated variables. A way to overcome this problem is to add a "ridge-like" term $(1/2)Kw^\top w$ to the objective function. This way we obtain a hybrid of lasso and ridge regression called the *elastic net method* and defined as follows:

Program (elastic net):

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\xi^\top\xi + \frac{1}{2}Kw^\top w + \tau \mathbf{1}_n^\top \epsilon \\ &\text{subject to} \\ &\qquad\qquad y - Xw - b\mathbf{1}_m = \xi \\ &\qquad\qquad w \leq \epsilon \\ &\qquad\qquad -w \leq \epsilon, \end{aligned}$$

where $K > 0$ and $\tau \geq 0$ are two constants controlling the influence of the ℓ^2-regularization and the ℓ^1-regularization.[1] Observe that as in the case of ridge regression, minimization is performed over ξ, w, ϵ and b, but b is not penalized in the objective function. The objective

[1]Some of the literature denotes K by λ_2 and τ by λ_1, but we prefer not to adopt this notation since we use λ, μ *etc.* to denote Lagrange multipliers.