Recap

if we want to analyze $\hat{\beta} - \beta^0$ (in a certain norm) we need conditions on X (e.g. $X^{(2)} = -X^{(1)}$ causes non-identifiability)

Sparse eigenvalues

suppose $X\theta = X\beta^0$
then:

$$0 = \|X(\theta - \beta^0)\|_2^2/n \geq \underbrace{\lambda^2_{\min}(\hat{\Sigma})}_{\text{min. eigenval. of } \hat{\Sigma}} \|\theta - \beta^0\|_2^2$$

$$\hat{\Sigma} = X^T X/n$$

for $p > n$: $\lambda^2_{\min}(\hat{\Sigma}) = 0 \rightsquigarrow$ bound above is “useless” idea: restrict to small sub-matrices

$\rightsquigarrow$ sparse eigenvalues (Meinshausen & Yu, 2009)

$$\phi^2_{\min}(m) = \min_{S \subseteq \{1,\ldots,p\}} \left(\lambda^2_{min}(\hat{\Sigma}_S);\, |S| \le m\right)$$

$$\iff \quad \phi^2_{\min}(m) = \min_{\beta \neq 0;\, \|\beta\|_0 \le m} \frac{\beta^T \hat{\Sigma} \beta}{\|\beta\|_2^2}$$

Then: if we require $\phi^2_{\min}(s_\theta + s_0) > 0$:

since $\|\theta - \beta^0\|_0 \le s_\theta + s_0$ we obtain

$$0 = \|X(\theta - \beta^0)\|_2^2/n \ge \phi^2_{\min}(s_\theta + s_0)\|\theta - \beta^0\|_2^2$$

$$\rightsquigarrow \quad \theta = \beta^0$$

Conclusion:
if we restrict to sparse vectors θ with at most the sparsity of β^0, i.e., $\|\theta\|_0 = s_\theta \le \|\beta^0\|_0 = s_0$

$\leadsto$ can identify the regression parameter vector if $\phi^2_{\min}(2s_0) > 0$

in addition: can show that under sparse eigenvalue condition and with high probability, for suitable λ

$$\|\hat{\beta}(\lambda)\|_0 \asymp \|\beta^0\|_0 = s_0$$

(non-trivial to show)

$\leadsto$ Lasso identifies β^0 with high probability
if $\phi^2(m) > 0$ for $m \gg s_0$

can also show (but this is non-trivial) that Lasso satisfies a cone condition with high probability

$$\text{(C)} \quad \|(\hat{\beta} - \beta^0)_{S_0^c}\|_1 \leq 3\|(\hat{\beta} - \beta^0)_{S_0}\|_1$$

consider sparse eigenvalues with the additional restriction that (C) is satisfied
$\leadsto$ restricted eigenvalues and compatibility constant which are larger (provide weaker assumptions) than sparse eigenvalues

compatibility condition: compatibility constant $\phi_0^2 > 0$
is the weakest assumption (among restricted and sparse eigenvalues which still allows to achieve (near) statistical optimality of Lasso