Recap

have developed a theory for analyzing prediction of Lasso $\leadsto$ with high probability:

$$\|X(\hat{\beta}(\lambda) - \beta^0)\|_2^2/n \leq \frac{3}{2}\lambda\|\beta^0\|_1$$

without requiring any condition on X (except scaled columns)

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:

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

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

for $p > n$: $\lambda_{\min}^2(\hat{\Sigma}) = 0 \rightsquigarrow$ bound above is "useless"

idea: restrict to small sub-matrices
$\rightsquigarrow$ sparse eigenvalues (Meinshausen & Yu, 2009)

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

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

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., $\|\theta\|_0 = s_\theta \leq \|\beta^0\|_0 = s_0$

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