OLS when dimension is $\ll n$:

Eg $\quad Y = \underset{n \times r}{X} \underset{r \times 1}{\beta^0} + \varepsilon$, ~~dim(X) =~~

$$E\left[\| X(\hat{\beta}_{OLS} - \beta^0) \|_2^2 / n \right] = \underbrace{\text{"Bias}^2\text{"}}_{=0} + \text{"Variance"}$$

$$= \sigma^2 \frac{r}{n}$$

$$\rightsquigarrow \| X(\hat{\beta}_{OLS} - \beta^0) \|_2^2 / n = O_P\left(\frac{r}{n}\right) \quad (*)$$

if oracle tells me S_0

where $Y = \underset{n\times p}{X}\underset{p\times 1}{\beta^0} + \varepsilon = X_{S_0}\beta^0_{S_0} + \varepsilon$

~~dim {S~~

run OLS on this model

since $\dim(S_0) = "r" = s_0$

$\overset{(*)}{\rightsquigarrow} \|X(\hat\beta^{\text{oracle}}_{\text{OLS}} - \beta^0)\|_2^2/n = O_P\left(\frac{s_0}{n}\right)$

$$\|\hat{\beta} - \beta^0\|_2 \xrightarrow{P} 0 \quad \text{if} \quad s_0 = o\left(n/\log(p)\right)$$

$$\|\hat{\beta} - \beta^0\|_1 \xrightarrow{P} 0 \quad \text{if} \quad s_0 = o\left(\sqrt{n/\log(p)}\right)$$

Lasso - OLS hybrid :

(1) Run Lasso $\rightsquigarrow \hat{S}$; $|\hat{S}| \leq \min(n,p)$

(2) Run OLS on $\hat{S} \longrightarrow \hat{\beta}^{\hat{S}}_{OLS}$

this is good if $\underline{\hat{S} \supseteq S_0}$

Why? Because OLS has no bias

goal: "something" which is computable
for general X and with lower bias than
Lasso

(1) Lasso-OLS hybrid
(2) Adaptive Lasso

$\rightsquigarrow$ I just look at (2) in orthonormal design
to illustrate what's going on

Adaptive Lasso:

if $\hat{\beta}_{init,j} = 0$

$\Longrightarrow \hat{\beta}_{adaptive,j}(\lambda) = 0$ because there is infinite penalty on j