LASSO: Least-Absolute Shrinkage and Selection Operator

Ideally I can solve the problem

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\ L(\beta) \quad \text{s.t.} \quad \underbrace{\|\beta\|_0 \le t}_{\text{at most } t \text{ vars}}$$

SE loss

= building the best t-variable model

(best subset selection problem)

<u>Problem</u>: comp. intensive

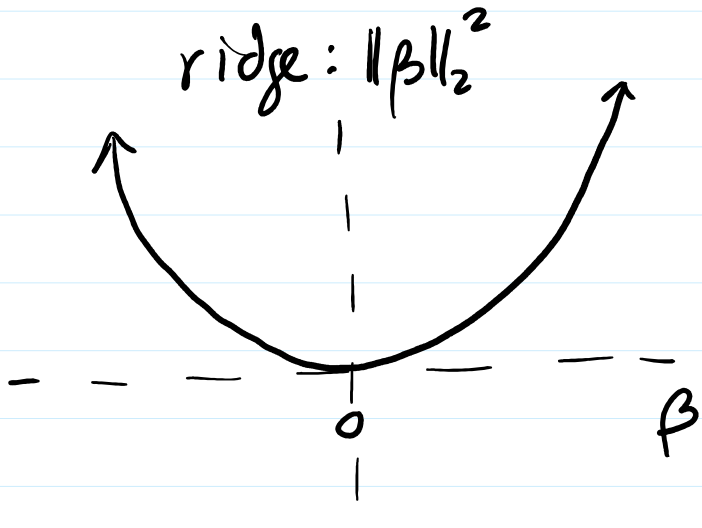
Optimizing under $\|\cdot\|_0$ constraint is difficult b/c $\|\beta\|_0$ is neither diff'able nor convex diff'able nor convex

LASSO : to make workable use the L1 norm instead – convex relaxation

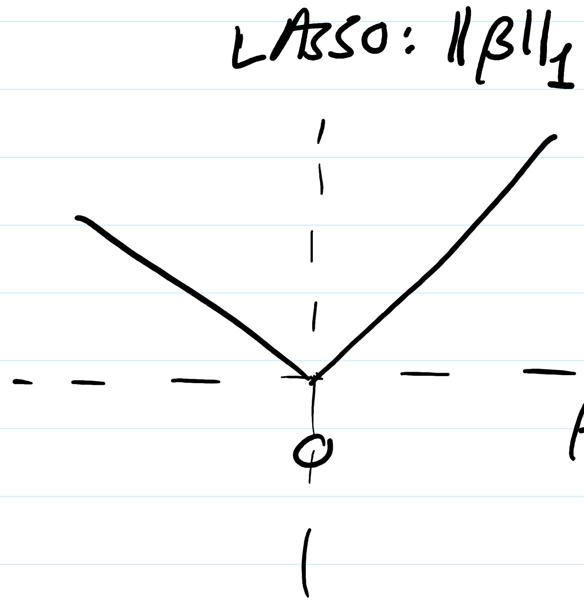
① $$\hat{\beta}^{(LASSO)} = \underset{\beta}{\operatorname{argmin}}\, L(\beta) \text{ s.t. } \|\beta\|_1 \le t$$

② $$\hat{\beta}^{(LASSO)} = \underset{\beta}{\operatorname{argmin}}\, L(\beta) + \lambda \|\beta\|_1$$

$\lambda \geq 0$

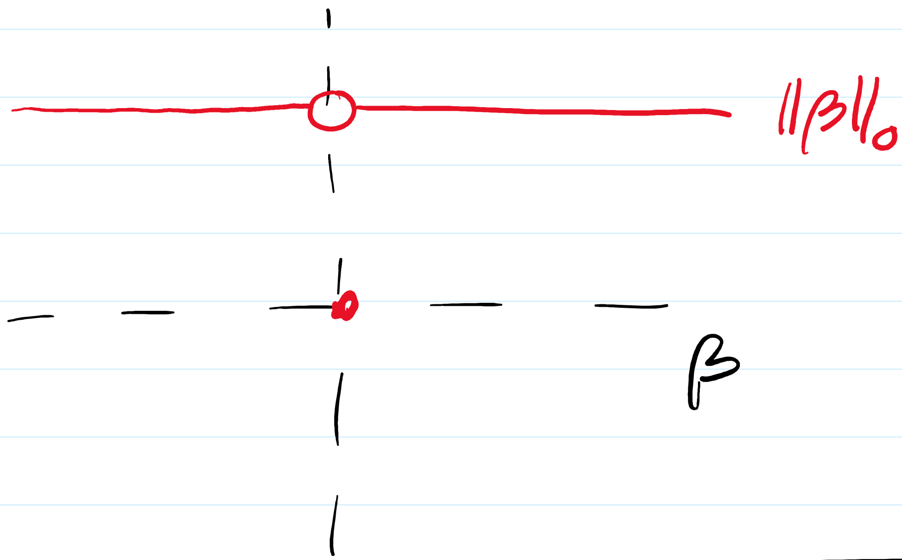


- convex
- diff'able



- convex
- not diff'able

Since $\|\cdot\|_1$ isn't diff'able, no closed form soln, need to use numerical methods.

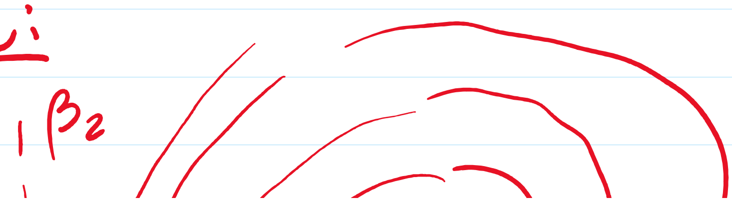


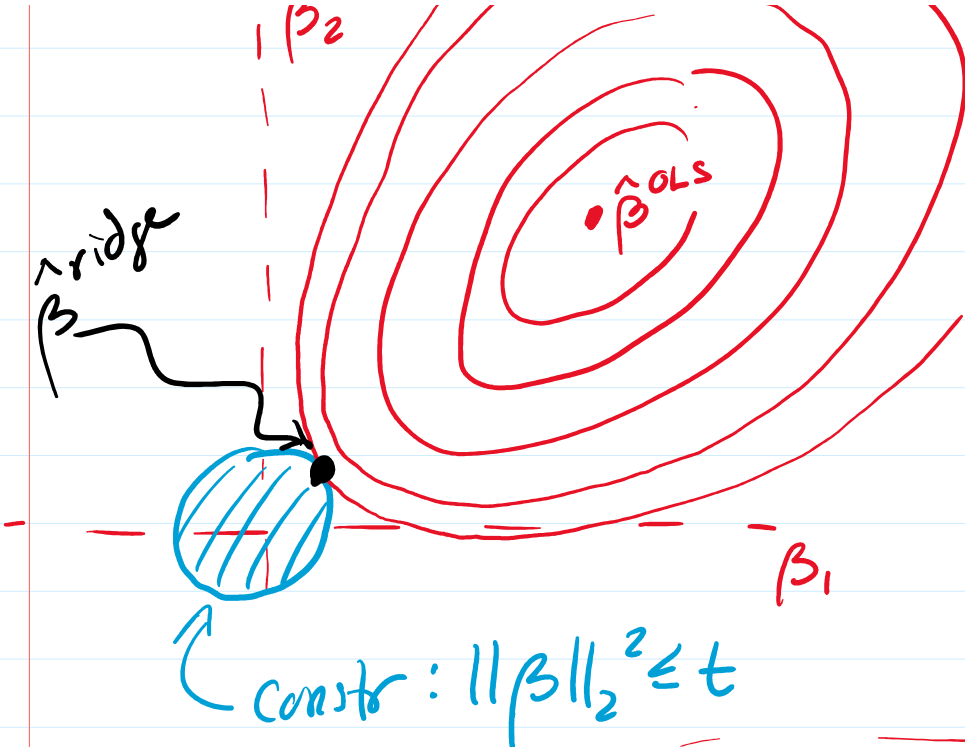
<u>Why use LASSO</u>?

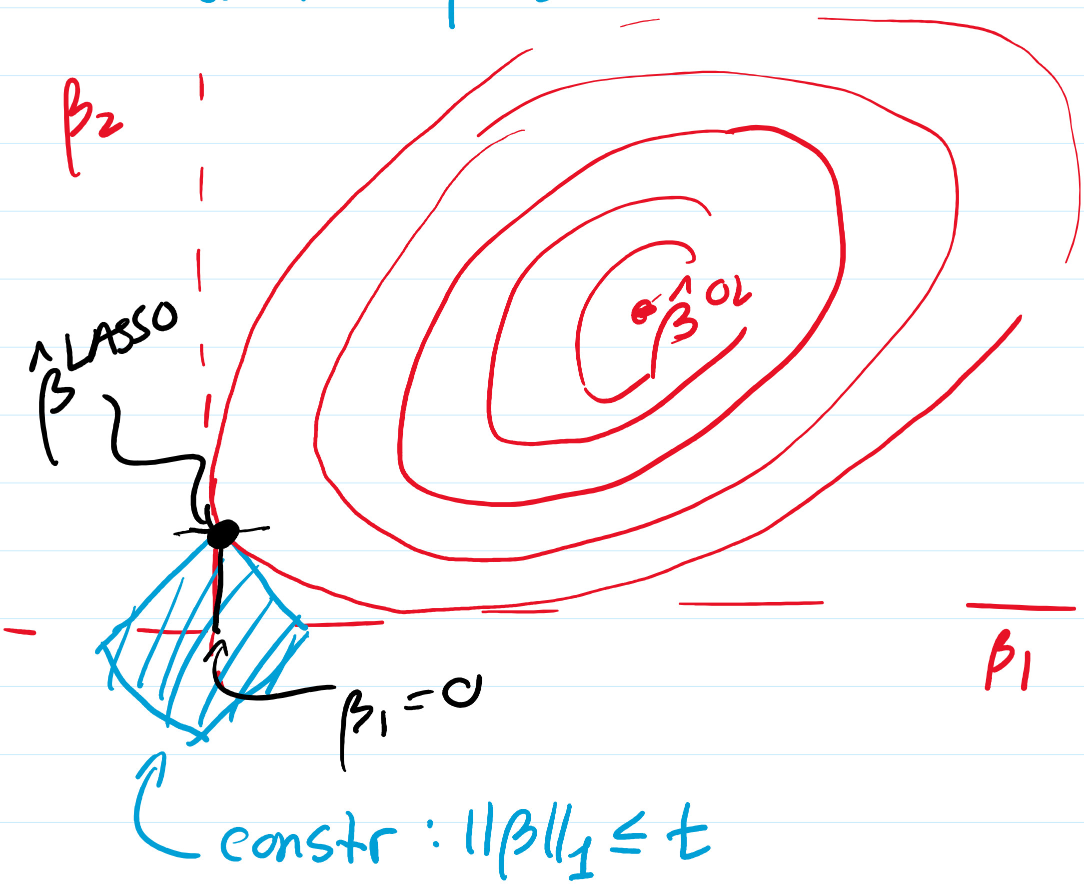
$\hat{\beta}^{(LASSO)}$ will exactly zero-out some elements of the coef vector for large enough λ - ridge doesn't do this

<u>Why</u>?

<u>Ridge:</u>







Comparison: assume X is orthogonal

(1) Variable Selection (Hard-thresholding)

$$\hat{\beta}_j^{HT} = \begin{cases} \hat{\beta}_j^{OLS} & \text{if } |\hat{\beta}_j^{OLS}| \geq t \\ 0 & \text{else} \end{cases}$$

(2) Ridge:

$$\hat{\beta}_j^{ridge} = \frac{\hat{\beta}_j^{OLS}}{1+\lambda} \quad \text{(proportional shrinkage)}$$

(3) LASSO:

$$\hat{\beta}_j^{LASSO} = \begin{cases} \hat{\beta}^{OLS} - \lambda, & \hat{\beta}^{OLS} \geq \lambda \\ \hat{\beta}^{OLS} + \lambda, & \hat{\beta}^{OLS} \leq -\lambda \\ 0, & |\hat{\beta}^{OLS}| \leq \lambda \end{cases}$$

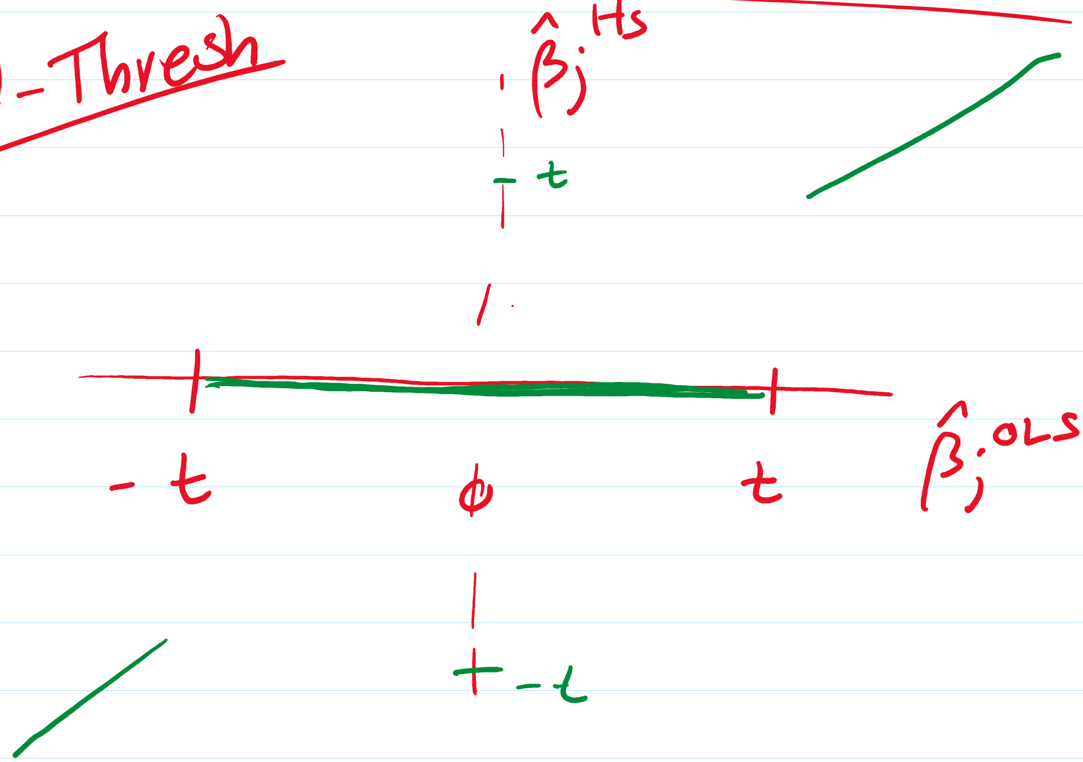
$$= \text{Sign}(\hat{\beta}_j^{OLS})\left(|\hat{\beta}_j^{OLS}| - \lambda\right)_+$$

$= \mathrm{sign}(\hat\beta_j)(|\hat\beta_j| - \lambda)_+$

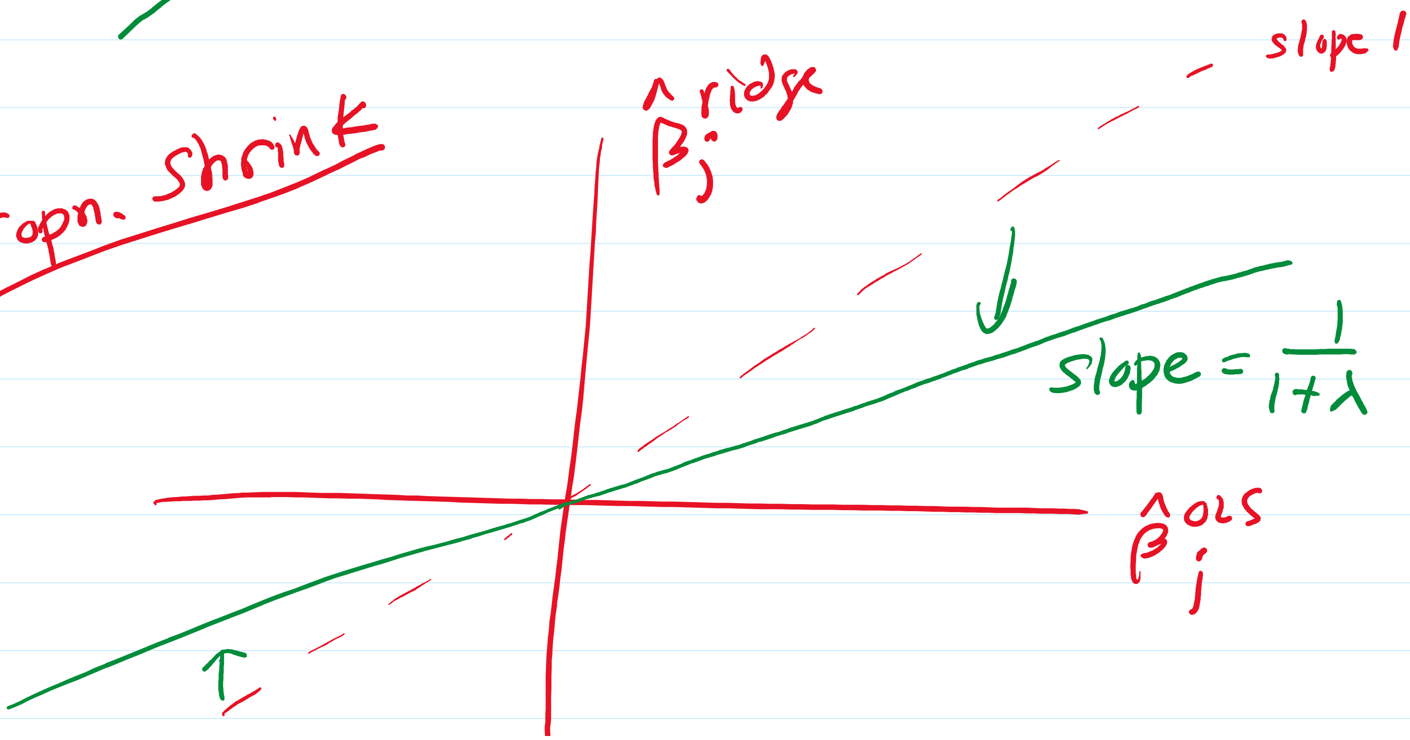
$(\cdot)_+ = \max(\cdot, 0)$

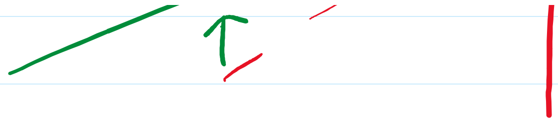
Soft-Thresholding.

Hard-Thresh

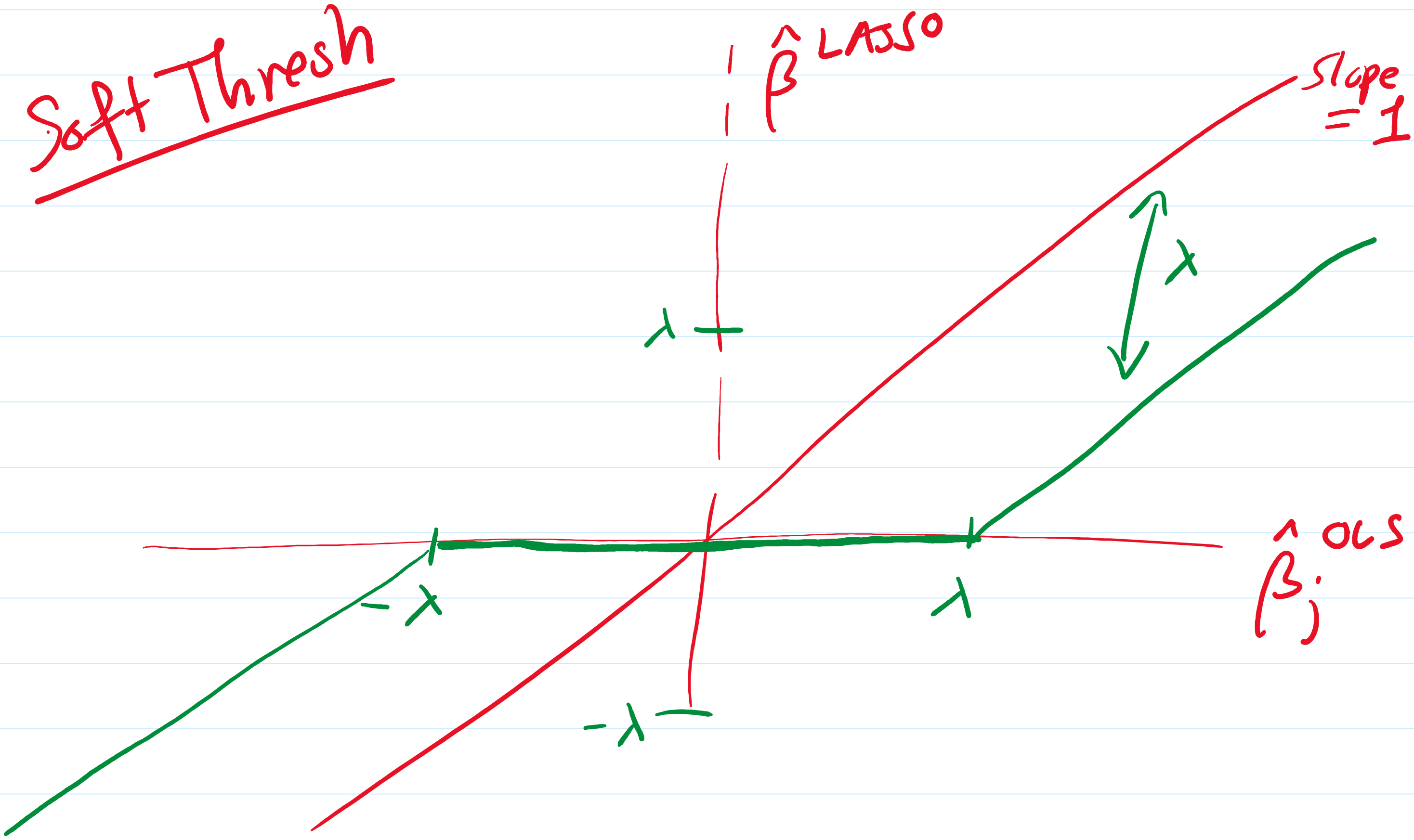


Propn. Shrink





Soft Thresh



Elastic Net

$$\hat{\beta}^{EN} = \underset{\beta}{\operatorname{argmin}}\ L(\beta) + \lambda\left[\frac{(1-\alpha)}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1\right]$$

$\alpha \in [0,1]$

= tradeoff between

... and L2 penalty

= tradeoff between
L1 and L2 penalty

$$\boxed{\begin{array}{l}\alpha = 0 \Rightarrow \text{ridge} \\ \alpha = 1 \Rightarrow \text{LASSO}\end{array}}$$

Can generally fit penalized methods

$$\hat{f} = \underset{f}{\arg\min}\ L(f)$$

<u>penalize</u>:

$$\hat{f} = \underset{f}{\arg\min}\ L(f) + \lambda J(f)$$

$J(f)$ = measure complexity of f

<u>Ex.</u> Penalized logistic regr

$$\hat{\beta} = \arg\min\ NLL(\beta) + \lambda \|\beta\|_2^2$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\; NLL(\beta) + \lambda \|\beta\|_2$$

$$\text{''} \qquad \text{'''} \quad + \lambda \|\beta\|_1$$