Logistics:

- A2 grades ~ Friday
- 1/2 class Friday for A2
  focus on coding and #4, 7
- A3 due next Friday

SURVEY ABOUT A2

Complete before class on Friday

Why does the lasso work?

Groups

1 person load notebook

$$\min_{\vec{\beta}} \|X\vec{\beta} - \vec{y}\|^2 + \lambda \|\vec{\beta}\|_1$$

What did you see? $\vec{y} = X\vec{\beta} + \varepsilon$

performance:

OLS < Ridge << Lasso

$$\vec{\beta} = \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

$40 = n$ data pts, $1000 = d$ dimension

$\alpha \|\vec{\beta}\|_1$ penalty

$\alpha \searrow 0 \implies$ more nonzeros in $\vec{\beta}_{lasso}$

$\alpha \nearrow \infty \implies$ many zeros, until all zeros

$\alpha \approx 1 \implies \vec{\beta} \approx \vec{\beta}_{lasso}$

Lasso going to work <u>IF</u>

$\boxed{\text{OLS needs } n \gtrsim 10 \cdot d}$

# of data pts $n \gtrsim 10 \cdot$ # nonzeros in true $\vec{\beta}$ $= 10 \cdot \|\vec{\beta}\|_0$

ex/ $n = 40$
$\|\vec{\beta}\|_0 = 4$ } synthetic example where we knew $\vec{\beta}$

Assumptions:
- true $\vec{\beta}$ is sparse
- choose correct penalty parameter λ
- conditions on X (incoherence)
  "not too many zeros in rows"

Geometric reasoning

$$\|\vec{\beta}\|_1 = \sum_{i=1}^{d} |\beta_i| \qquad \ell_1\text{-norm}$$

$$\|\vec{\beta}\|_2 = \sqrt{\sum_{i=1}^{d} \beta_i^2} \qquad \ell_2\text{-norm}$$

$$\|\vec{\beta}\|_0 = \#\text{ nonzeros in } \vec{\beta}$$

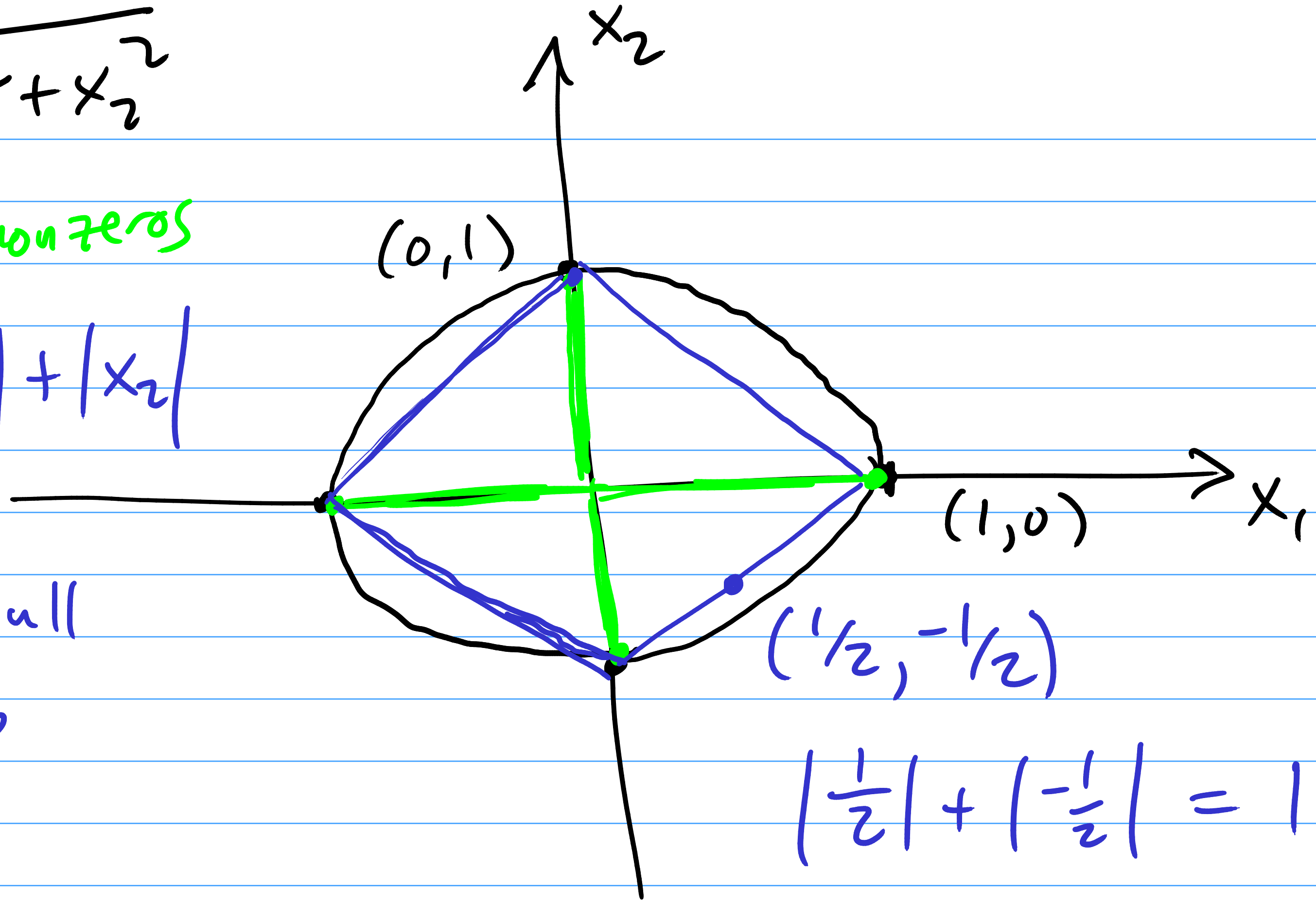
<u>Defn</u> The <u>unit ball</u> of a norm $\|\cdot\|$ is the set of points $\vec{x} \in \mathbb{R}^d$ where $\|\vec{x}\| = 1$.

$$\|\vec{x}\|_2 = \sqrt{x_1^2 + x_2^2}$$

$$\|\vec{x}\|_0 = \#\ \text{nonzeros}$$

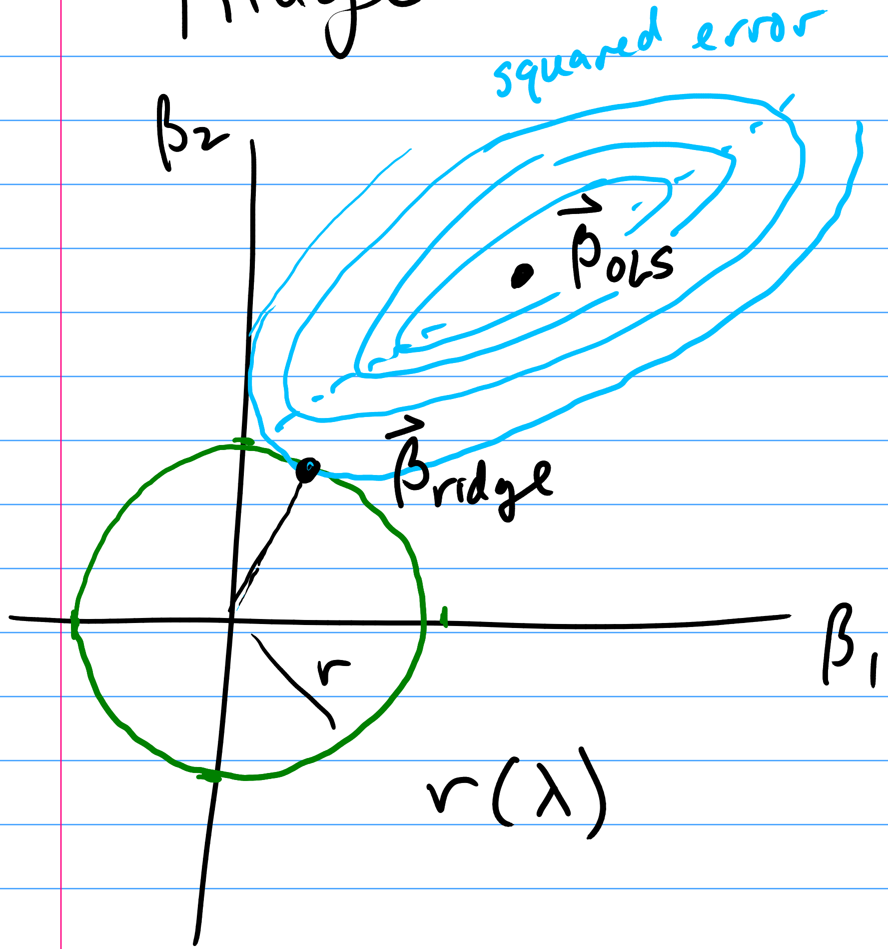
$$\|\vec{x}\|_1 = |x_1| + |x_2|$$

convex hull of $\|\cdot\|_0$

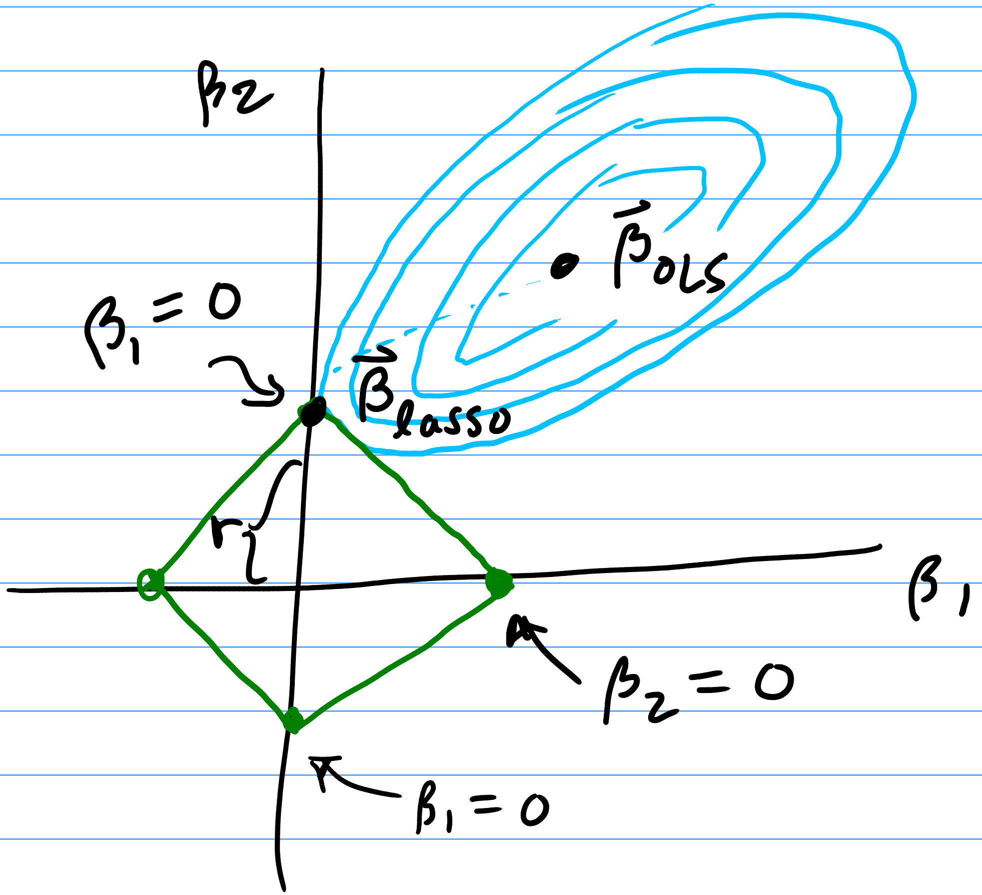


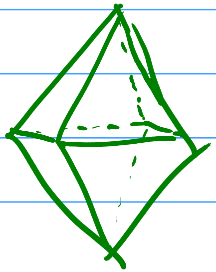
$$\left|\tfrac{1}{2}\right| + \left|-\tfrac{1}{2}\right| = 1$$

Ridge



Lasso





intersection points will be at sharp corners of ℓ_1 ball

A3 Help

eigenvalues : $A \in \mathbb{R}^{n \times n}$

$$A = \underbrace{W}_{\text{eigenvectors}} \overbrace{\Lambda}^{\text{eigenvalues}} W^{-1}$$

$$A\vec{w}_i = \lambda_i \vec{w}_i$$ $i^{\underline{th}}$ eigenvalue & eigenvector

$X \in \mathbb{R}^{n \times d}$

can use SVD

$$X = USV^T$$

$$\operatorname{Tr}(\underbrace{X^T X}_{\text{square}})$$

$$X^T X = (USV^T)^T (USV^T)$$

$$= (VSU^T)(USV^T)$$

$(U^T U = I)$ $$= VS^2V^T$$

\mbox{ text }

adding notes to equations