Solving ℓ_1 Regularized Least Squares via Proximal Gradient Descent

Objectives

- apply proximal gradient approach to solve ℓ_1-regularized least squares
- derive solution to regularization phase
- explore alternating gradient and soft thresholding steps

The ℓ_1-regularized least-squares problem can be solved via proximal gradient descent

features/labels $(\underline{x}_i, d_i)$ model $\underline{x}_i^T \underline{w} \approx d_i$

$$\min_{\underline{w}} \|\underline{A}\underline{w} - \underline{d}\|_2^2 + \lambda \underbrace{\|\underline{w}\|_1}$$

encourages sparse solutions

☆ This is seperable, perfect for PGD

no closed form solution

Proximal Gradient Descent Algorithm

step size

a) $$\underline{z}^{(k)} = \underline{w}^{(k)} - \tau \underline{A}^T(\underline{A}\underline{w}^{(k)} - \underline{d})$$

least squares gradient descent

b) $$\underline{w}^{(k+1)} = \arg\min_{\underline{w}} \|\underline{z}^{(k)} - \underline{w}\|_2^2 + \tau\lambda\|\underline{w}\|_1$$

Regularization step involves scalar minimization[3]

$$\min_{\underline{w}} \|\underline{z}^{(k)} - \underline{w}\|_2^2 + \tau\lambda \|\underline{w}\|_1 \Rightarrow \min_{w_i, i=1,\dots,M} \sum_{i=1}^{M} (z_i^{(k)} - w_i)^2 + \lambda\tau |w_i|$$

Can be rewritten as a sum of scalar cost functions

Consider $\min_{w_i} (z_i^{(k)} - w_i)^2 + \lambda\tau |w_i|, \quad \lambda, \tau > 0$

case 1: $w_i \geq 0$

$$\min_{w_i} (z_i - w_i)^2 + \lambda\tau w_i, \quad w_i \geq 0$$

$$\frac{d}{dw_i}\left\{(z_i - w_i)^2 + \lambda\tau w_i\right\} = 0, \quad w_i \geq 0$$

$$-2(z_i - w_i) + \lambda\tau = 0, \quad w_i \geq 0$$

$$w_i = z_i - \frac{\lambda\tau}{2}, \quad w_i \geq 0$$

if $z_i > \frac{\lambda\tau}{2}$,

$$w_i = z_i - \frac{\lambda\tau}{2}$$

if $z_i < \frac{\lambda\tau}{2}$,

$$w_i = 0$$

$$w_i = \left(z_i - \frac{\lambda\tau}{2}\right)_+$$

compact form, if this is negative we set it to 0

Case 2: $w_i \le 0$

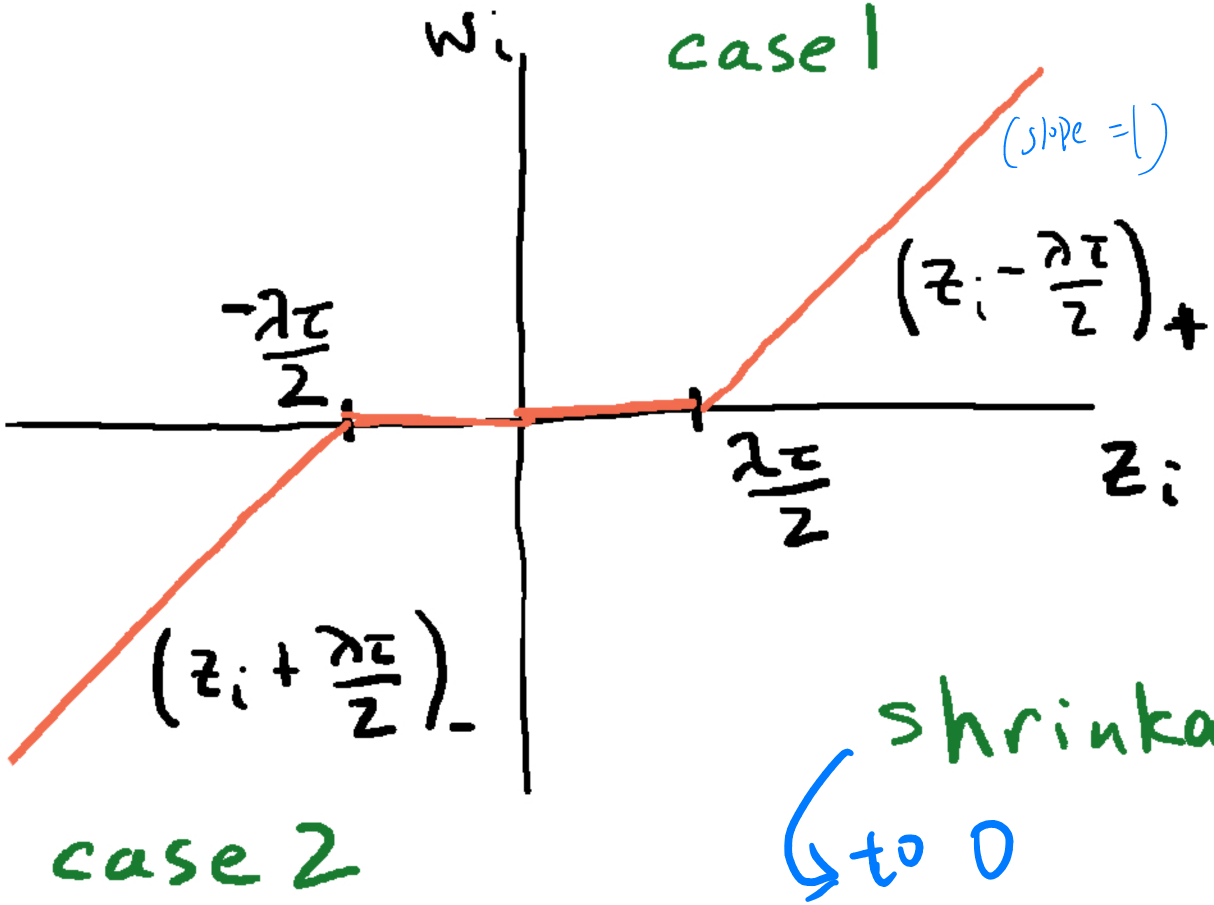
$$\min_{w_i} (z_i - w_i)^2 - \lambda\tau w_i \Rightarrow \frac{d}{dw_i}\left\{(z_i - w_i)^2 - \lambda\tau w_i\right\} = 0$$

$|w_i| = -w_i$ for negative case

★ If positive, we set to 0

$$-2(z_i - w_i) - \lambda\tau = 0, \quad w_i \le 0$$

$$w_i = \left(z_i + \frac{\lambda\tau}{2}\right)_-$$



"Soft threshold"

$$w_i = \begin{cases} 0, & -\frac{\lambda\tau}{2} < z_i < \frac{\lambda\tau}{2} \\ z_i - \frac{\lambda\tau}{2}, & z_i > \frac{\lambda\tau}{2} \\ z_i + \frac{\lambda\tau}{2}, & z_i < -\frac{\lambda\tau}{2} \end{cases}$$

shrinkage

to 0

★ $$w_i = \left(|z_i| - \frac{\lambda\tau}{2}\right)_+ \operatorname{sign}(z_i)$$

Algorithm alternates descent and shrinkage