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

features / label (x_i, d_i) model $x_i^T w \approx d_i$

$$\min_w \|Aw - d\|_2^2 + \underbrace{\lambda \|w\|_1}_{}$$ → Encourages sparse solutions

no closed form solution

Proximal Gradient Descent Algorithm

a) $z^{(k)} = w^{(k)} - \tau A^T(Aw^{(k)} - d)$ Least-square GD

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

Regularization step involves scalar minimization.

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

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

case 1: $w_i \geq 0$

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

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

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

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

if $z_i < \frac{\lambda\tau}{2}$, $w_i = 0$

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

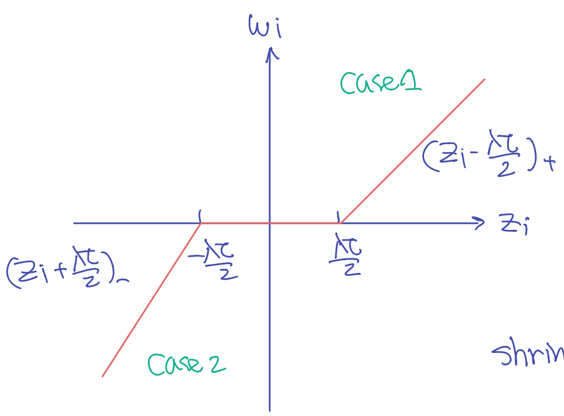
case 2: $w_i < 0$

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

$$\frac{d}{dw_i}\{(z_i - w_i)^2 - \lambda\tau w_i, w_i < 0\} = 0 \Longrightarrow -2(z_i - w_i) - \lambda\tau = 0, \quad w_i < 0$$

$$w_i = z_i + \frac{\lambda\tau}{2}, \quad w_i < 0.$$

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



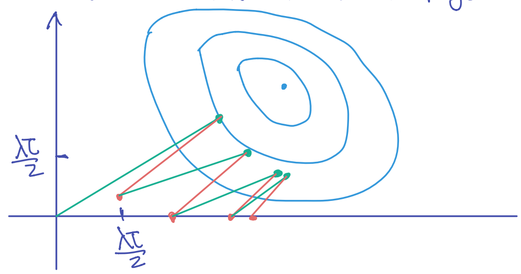
"soft threshold"

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

shrinkage

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

Algorithm alternates descent and shrinkage



gradient descent

regularization (shrinkage)