Lecture 16 : Iterative Solvers

obs $(x_i, y_i) \in \mathbb{R}^p \times \mathbb{R}$

or $\mathbb{R}^p \times \{-1, +1\}$ for $i = 1, \dots, n$

find $w \in \mathbb{R}^p$ $\quad \underset{\sim}{y} \approx X \underline{w}$ $\quad$ where $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$$\boxed{\hat{w} = (X^T X)^{-1} X^T y}$$

$$X = \begin{bmatrix} - & x_1^T & - \\ - & x_2^T & - \\ & \vdots & \\ - & x_n^T & - \end{bmatrix}$$

initialize: $\hat{w}^{(0)}$

for $k = 0, 1, 2, 3 \dots$

$$\hat{w}^{(k+1)} = \hat{w}^{(k)} - \tau \underline{\underline{X^T (X \hat{w}^{(k)} - y)}}$$

$\tau > 0$ step size

if $\| \hat{w}^{(k+1)} - \hat{w}^{(k)} \|_2 < \varepsilon$, then BREAK

end

Gradient descent

– or –

Landweber iterations

$$\hat{w} = \arg\min_{\underline{w}} \underbrace{\|X\underline{w} - y\|_2^2}_{f(\underline{w})}$$

$$\nabla_w f(w) = \nabla_w (Xw - y)^T (Xw - y)$$

$$= \nabla_w w^T X^T X w - 2w^T X^T y + y^T y$$

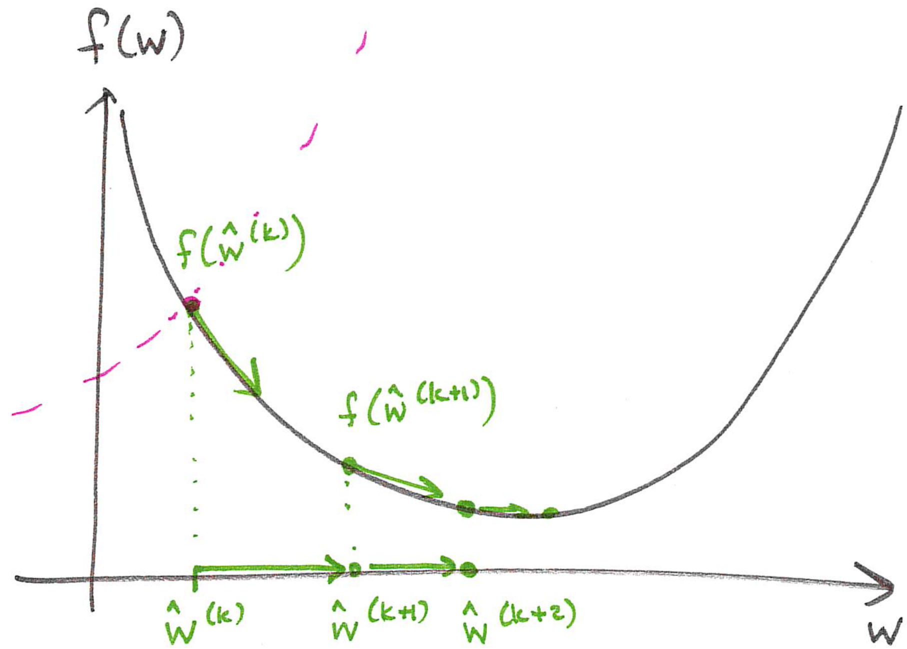
$$= 2X^T X w - 2X^T y$$

$$= 2X^T(Xw - y)$$

new iterate $\hat{w}^{(k+1)}$

= old iterate

+ update (step in direction of negative gradient)



Recall $\|X\|_{op} = \|X\|_2 = \sigma_{max}(X)$

$$Xw = U\Sigma V^T w$$

$$\|Xw\|_2 = \|\underset{\text{basis}}{U}\Sigma V^T w\|_2$$

$$= \|\Sigma V^T w\|_2$$

$$= \left(\sum_i \left(\sigma_i (V^T w)_i\right)^2\right)^{1/2}$$

$$\leq \sigma_{max}\left(\sum_i (V^T w)_i^2\right)^{1/2}$$

$$= \underline{\sigma_{max}}\ \|\underset{\text{orthogonal}}{V^T} w\|_2$$

$$= \sigma_{max} \|w\|_2 .$$

$$\boxed{\|Xw\|_2 \leq \|X\|_{op} \|w\|_2}$$

Convergence:

want $\|X\hat{w}^{(k+1)} - y\|_2^2 < \|X\hat{w}^{(k)} - y\|_2^2$

recall $\hat{w}^{(k+1)} = \hat{w}^{(k)} - \tau X^T(X\hat{w}^{(k)} - y)$

$$\|X\hat{w}^{(k+1)} - y\|^2 = \|X(\hat{w}^{(k)} - \tau X^T(X\hat{w}^{(k)} - y)) - y\|^2$$

$$= \|X\hat{w}^{(k)} - y - \tau XX^T(X\hat{w}^{(k)} - y)\|_2^2$$

$$= \|X\hat{w}^{(k)} - y\|_2^2 + \tau^2 \|XX^T(X\hat{w}^{(k)} - y)\|_2^2 - 2\tau (X\hat{w}^{(k)} - y)^T X X^T (X\hat{w}^{(k)} - y)$$

$\leq \|X\|_{op}^2 \cdot \|X^T(X\hat{w}^{(k)} - y)\|_2^2$

$$\leq \|X\hat{w}^{(k)} - y\|_2^2 + \tau\left(\|X\|_{op}^2 \|X^T(X\hat{w}^{(k)} - y)\|_2^2 \cdot \tau - 2\|X^T(X\hat{w}^{(k)} - y)\|_2^2\right)$$

$$= \|X\hat{w}^{(k)} - y\|_2^2 + \tau\|X^T(X\hat{w}^{(k)} - y)\|_2^2\left(\tau\|X\|_{op}^2 - 2\right)$$

if $\tau\|X\|_{op}^2 - 2 < 0$, then $\|X\hat{w}^{(k+1)} - y\|^2 < \|X\hat{w}^{(k)} - y\|^2$

if $\tau < 2/\|X\|_{op}^2$, then $\|X\hat{w}^{(k+1)} - y\|^2 < \|X\hat{w}^{(k)} - y\|^2$

if $\hat{w}^{(0)} = 0$ and $0 < \tau < 2/\|X\|_{op}^2$, then

$$\hat{w}^{(k)} \xrightarrow[k \to \infty]{} (X^T X)^{-1} X^T y$$