$$f(w) = \frac{1}{2} w^T A w, \quad w \in \mathbb{R}^d, \quad A = \mathrm{diag}(\lambda_1, \dots, \lambda_d) \succ 0$$

$$w^* = \underline{0} \in \mathbb{R}^d$$

$$\nabla_w f(w) = A w \implies w^{k+1} = w^k - \alpha A \cdot w^k$$

$$i \left\{ \begin{pmatrix} \lambda_1 \cdot w_1^k \\ \vdots \\ \lambda_d \cdot w_d^k \end{pmatrix} \right.$$

$$w_i^{k+1} = w_i^k - \alpha \cdot \lambda_i w_i^k$$

$$= w_i^k (1 - \alpha \lambda_i)$$

$$w_i^k = \cancel{k}\, w_i^{k-1}(1 - \alpha\lambda_i)^2$$

$$= w_i^0 (1 - \alpha\lambda_i)^{k+1}$$

$w_i^k \longrightarrow w_i^* \quad \text{as} \quad k \to \infty$

$$w_i^k - w_i^* = w_i^0 (1 - \alpha\lambda_i)^k$$

$0 < |1 - \alpha\lambda_i| < 1$

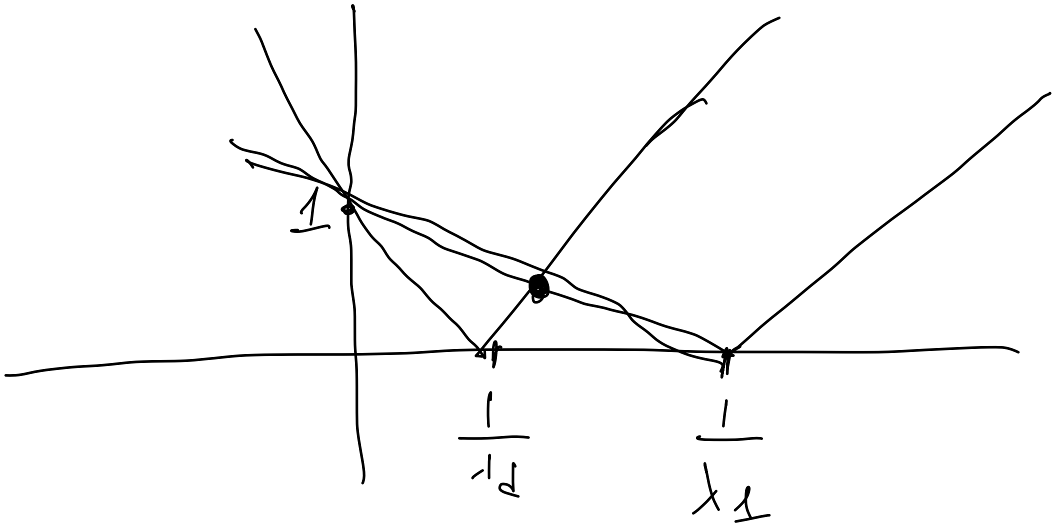
$\iff \alpha\lambda_i < 2$

if $\boxed{\alpha < \frac{2}{\lambda_d}} \Rightarrow w^k - w^* \to 0$ as $k \to \infty$

$$\text{rate}(\alpha) = \max_i |1 - \alpha\lambda_i|$$

$$= \max\{|1 - \alpha\lambda_1|, |1 - \alpha\lambda_d|\}$$

$$\min_{\alpha} \max \{ \overbrace{|1-\alpha\lambda_1|}^{:=f_1}, \overbrace{|1-\alpha\lambda_d|}^{:=f_2} \}$$



$$|1-\alpha\lambda_1| = |1-\alpha\lambda_d|$$

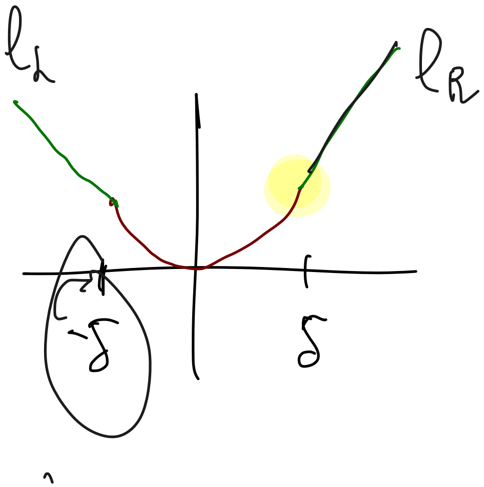
$$1-\alpha\lambda_1 = \alpha\lambda_d - 1$$

$$\alpha^* = \frac{2}{\lambda_1+\lambda_d}$$

learning rate α
which achieves the faster rate

$$\text{rate}(\alpha^*) = 1 - \underbrace{\frac{2}{\lambda_1+\lambda_d}}_{\alpha^*} \cdot \lambda_1 = \frac{\lambda_d-\lambda_1}{\lambda_d+\lambda_1} = \frac{\frac{\lambda_d}{\lambda_1}-1}{\underbrace{\frac{\lambda_d}{\lambda_1}}_{:=\kappa \text{ condition number}}+1}$$

$$\text{rate}(d^*) = \frac{k-1}{k+1}$$



$$l_{huber}(x) = \begin{cases} x^2 & \text{for } |x| \leq \delta \\ l_R & \text{for } x > \delta \\ l_L & \text{for } x < -\delta \end{cases}$$

Conditions to satisfy:

① $l_R(x) = x^2$ at $x = \delta$

② $\left.\frac{d}{dx} l_R(x)\right|_{x=\delta} = \left.\frac{d}{dx} x^2\right|_{x=\delta} = 2\delta$

③ $l_R(x) = Kx + c$

From ② $k = 2\delta \Rightarrow l_R(x) = 2\delta x + c$

From (i) $l_R(\delta) = \delta^2 \Rightarrow 2\delta^2 + c = \delta^2$

$$c = -\delta^2$$

$$\Rightarrow l_R(x) = 2\delta x - \delta^2$$

$$l_h(x) = l_R(-x)$$

$x = y - f(x)$

$$l_{huber} = \begin{cases} x^2 & \text{for } |x| \leq \delta \\ 2\delta|x| - \delta^2 & \text{for } |x| > \delta \end{cases}$$

$$\ell'_{huber}(x) = \begin{cases} 2x & \text{if } |x| \leq \delta \\ 2\delta \operatorname{sgn}(x) & \text{else} \end{cases}$$

$$\ell''_{huber}(x) = \begin{cases} 2 & \text{if } |x| \leq \delta \\ 0 & \text{else} \end{cases} \implies \text{convexity}$$