

Lec 25.

Convergence of SD.

The diagram illustrates the iterative step in the SD algorithm. It shows a point x_k on a horizontal axis. An arrow points from x_k to a blue dot labeled x_{k+1} , representing the next iteration. Below x_k , the text "step size" is written in blue. To the right of the arrow, the text "direction" is written in black. Above the arrow, the equation $\alpha_k P_k = r_k = b - Ax_k$ is written in blue, where α_k is the step size, P_k is the direction, and r_k is the residual.

$$\alpha_k P_k = r_k = b - Ax_k$$

$$\varphi(x_{k+1}) = \frac{1}{2} x_{k+1}^T A x_{k+1} - x_{k+1}^T b$$

||

$$\inf_{\alpha} \varphi(x_k + \alpha_k P_k) \quad Ax_*$$

$$\varphi(x_{k+1}) + \frac{1}{2} x_*^T A x_*$$

$$= \frac{1}{2} (x_{k+1} - x_*)^T A (x_{k+1} - x_*)$$

$$:= \frac{1}{2} \|x_{k+1} - x_*\|_A^2 \quad A > 0.$$

$$\| \mathbf{x}_{k+1} - \mathbf{x}_* \|_A^2$$

$$= \| \mathbf{x}_k + \alpha_k (\mathbf{b} - A\mathbf{x}_k) - \mathbf{x}_* \|_A^2$$

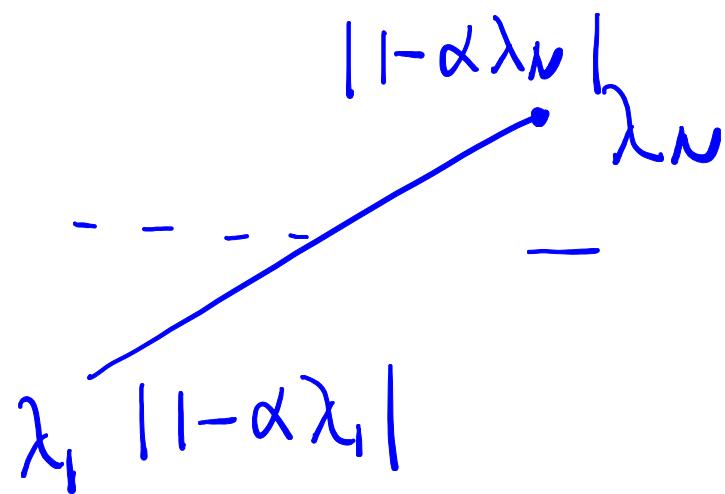
$$= \| (\mathbf{I} - \alpha_k \mathbf{A}) (\mathbf{x}_k - \mathbf{x}_*) \|_A^2$$

$$\left(= (\mathbf{x}_k - \mathbf{x}_*)^T (\mathbf{I} - \alpha_k \mathbf{A})^2 \mathbf{A} (\mathbf{x}_k - \mathbf{x}_*) \right)$$

$$\leq \left(\max_{\lambda_1 \leq \lambda \leq \lambda_N} |1 - \alpha_k \lambda|^2 \right) \|x_k - x_*\|_A^2$$

SD:

$$\min_{\alpha} \max_{\lambda_1 \leq \lambda \leq \lambda_N} |1 - \alpha \lambda|^2$$



$$1 - \alpha \lambda_1 = \alpha \lambda_N - 1$$

$$\Rightarrow \alpha = \frac{2}{\lambda_1 + \lambda_N}$$

$$\max_{\lambda_1 \leq \lambda \leq \lambda_N} |1 - \alpha \lambda|^2$$

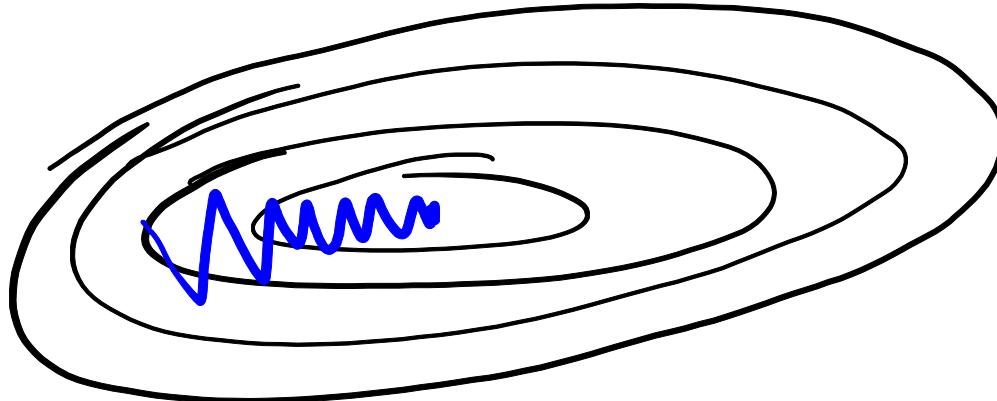
$$= \left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1} \right)^2 = \left(\frac{\frac{\lambda_N}{\lambda_1} - 1}{\frac{\lambda_N}{\lambda_1} + 1} \right)^2 = \left(\frac{K-1}{K+1} \right)^2$$

$$\|x_{k+1} - x_*\|_A \leq \left(\frac{k-1}{k+1} \right) \|x_k - x_*\|_A$$

$$\leq \left(\frac{k-1}{k+1} \right)^{k+1} \|x_0 - x_*\|_A$$

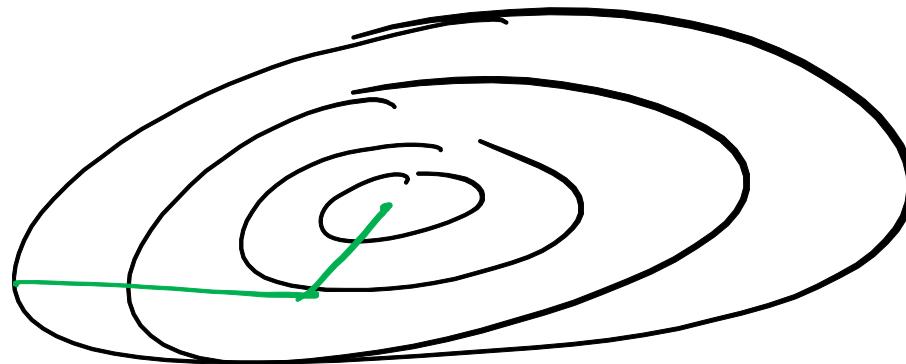
< \in

$$\Rightarrow k > C K \log(1/\epsilon).$$



SD

① 2D



CG

② arbitrary
large

$$\left(\frac{\sqrt{K} - 1}{\sqrt{K} + 1} \right)^k$$

$$k > C \sqrt{k} \log(\frac{1}{\epsilon}) .$$

$$\inf_{\alpha} \varphi(x_k + \alpha P_k)$$

$$\frac{1}{2} (x_k + \alpha P_k)^T A (x_k + \alpha P_k)$$

$$- b^T (x_k + \alpha P_k)$$

$$\Rightarrow \alpha = \frac{P_K^T r_K}{P_K^T A P_K}$$

2D.

1st step: 5D.

$$x_0 \Rightarrow r_0 \Rightarrow P_0 = r_0$$

$$\alpha_0 \Rightarrow x_1$$

2nd step. $P_1^T A P_0 = 0$. A -ortho
to P_0 .

(in 2D) uniquely fixes P_1 (up to scaling)

$$\Rightarrow d_1 \Rightarrow x_2 = x_* \Leftrightarrow r_2 = 0.$$

$$\Leftrightarrow P_0^T r_2 = 0, P_1^T r_2 = 0.$$

Now compute .

$$r_2 = b - Ax_2 = b - A(x_1 + \alpha_1 p_1)$$

$$= r_1 - \alpha_1 A p_1$$

$$= b - A(x_0 + \alpha_0 r_0) - \alpha_1 A p_1$$

$$= (I - \alpha_0 A) r_0 - \alpha_1 A p_1$$

$$P_0^T r_2 = \left(r_0^T r_0 - \alpha_0 r_0^T A r_0 \right) - \alpha_1 P_0^T A P_1$$

$$P_1^T r_2 = (P_1^T r_0 - \alpha_1 P_1^T A P_1) - \underbrace{\alpha_0 P_1^T A P_0}_{\parallel}$$

$$\begin{aligned} P_1^T (r_0 - r_1) &= P_1^T (r_0 - b + A(x_0 + \alpha_0 r_0)) \\ &= P_1^T A P_0 \cdot (+\alpha_0) = 0. \end{aligned}$$

$$\Rightarrow r_2 = 0 \quad \square.$$

Generally. $\{x_k, P_k, r_k\}$

$$x_{k+1} = x_k + \alpha_k P_k,$$

$$\alpha_k = \frac{P_k^T r_k}{P_k^T A P_k}$$

(minimization)

$$r_{k+1} = b - A x_{k+1}.$$

$$P_{k+1} = r_{k+1} + \beta_k P_k$$

$$\beta_k = \frac{P_k^T A r_{k+1}}{P_k^T A P_k}$$

(A-orthogonality)

$P_{k+1} : A\text{-ortho} \rightarrow P_k$.

(non-trivial fact):

$P_{k+1} \perp_A \text{span}\{P_0, \dots, P_k\}$.

