$$Ax = b$$

① $x^* = A^{-1}b$ $\qquad \omega \leq 2.37\ldots$

→ computing A^{-1} $O(n^\omega)$

Gaussian elim. $O(n^3)$

→ numerical issues

→ lack of utilizing sparsity

② Iterative

$$x^0 \to x^1 \to x^2 \to \ldots \to x^t$$

→ fast*

→ interaction with A is only via matrix-vector product

$y \to Ay$ $\quad O(\mathrm{nnz}(A))$

exploits sparsity

→ A invertible $\quad \forall_i \ \lambda_i \neq 0$

→ WLOG A is PSD $\quad \forall_i \ \lambda_i > 0$

symmetric

$$A^T A x = A^T b$$

symmetric & PSD

$y \to A^T \underbrace{A y}$ $\leftarrow O(\mathrm{nnz}(A))$

Error measures:

→ $e(x) = \boxed{x - x^*}$ LHS error

→ $r(x) = b - Ax = \boxed{A(x^* - x)}$ RHS error

$$\|x\|_A = \sqrt{x^T A x} \qquad A = I \to \|\ \|_2$$

$$\min_x \frac{1}{2}\|e(x)\|_A^2 \leftarrow \qquad e(x) = x - x^*$$

$$\frac{1}{2}\|e(x)\|_A^2 = \frac{1}{2}(x-x^*)^T A (x-x^*) = \underbrace{\frac{1}{2} x^T A x - x^T \underbrace{A x^*}_{= b}} + \underbrace{\frac{1}{2}(x^*)^T A x^*}_{= \frac{1}{2}\|x^*\|_A^2}$$

$$\boxed{g(x) = \frac{1}{2} x^T A x - b^T x} \leftarrow$$

$$\min_x \frac{1}{2}\|g(x)\|^2$$

Solve via Gradient Descent

1. $x^0 = 0$
2. In each step t:

$$x^{t+1} \leftarrow x^t - \eta \nabla g(x^t)$$

$$\nabla g(x^t) = \overbrace{A x^t - b}^{= -r(x^t)}$$

$$\nabla g(x) = 0 \Rightarrow x = x^*$$

Hessian $\nabla^2 g(x) = A$ $\leftarrow$ convex

β-smoothness & α-strong conv. of g

$\nabla^2 g(x) = A$

β is the smallest eigenv.

α —"— largest eigenvalue

$$\boxed{\kappa = \frac{\lambda_n}{\lambda_1}}$$

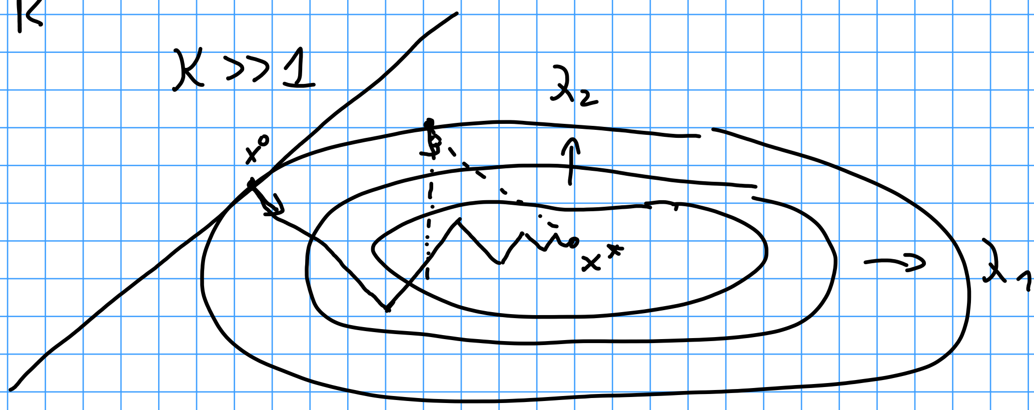
$\Rightarrow$ Runtime: $\boxed{O\left(\kappa \log \frac{1}{\varepsilon}\right)}$

$$\Rightarrow \boxed{\frac{1}{2}\|e(x^t)\|_A^2 \leq \varepsilon}$$

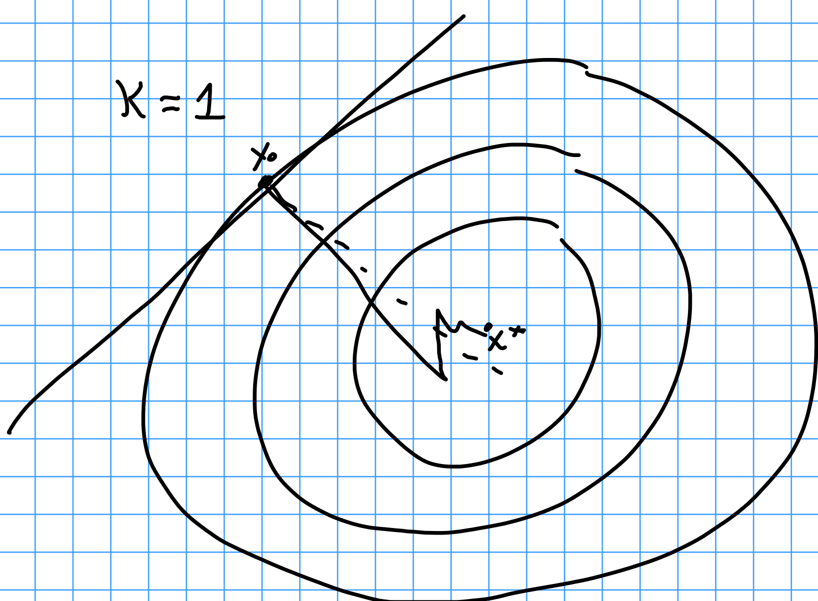
$g(x)$ Level sets

$$S_r = \{x \mid g(x) = r\} = \{x \mid \underbrace{\tfrac{1}{2}(x-x^*)^T A (x-x^*) = r}\}$$

$\mathbb{R}^2$



Ellipse centered at x^* and principal axis given by A



$O(\kappa \log \frac{1}{\varepsilon})$

$$x^{t+1} \in x^t - \eta \nabla g(x^t) = \underbrace{x^t + \eta\, r(x^t)}$$

$$\underbrace{r_{t+1}} = b - A x^{t+1} = \underbrace{b} - A(\underbrace{x^t} + \eta r_t) = \underbrace{(I - \eta A)\underbrace{r_t}} = \underbrace{(I-\eta A)^t b}$$

$$r_{t+1} = p_t(A) \cdot b \qquad p_t(y) = (1 - \eta y)^t$$

$$x_{t+1} = x_t + \eta r_t = \sum_{i=1}^{t} \eta (I - \eta A)^i b = q_{t,\eta}(A)\, b$$

$$\boxed{q_{t,\eta}(y) = \sum_{i=1}^{t} \eta (1 - \eta y)^i}$$

$$A = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \qquad p(A) = \begin{bmatrix} p(\lambda_1) & & & \\ & p(\lambda_2) & & \\ & & \ddots & \\ & & & p(\lambda_n) \end{bmatrix}$$

$$q_{t,\eta}(A) = \begin{bmatrix} q_{t,\eta}(\lambda_1) & & \\ & \ddots & \\ & & q_{t,\eta}(\lambda_n) \end{bmatrix} \to \begin{bmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n} \end{bmatrix} = A^{-1}$$

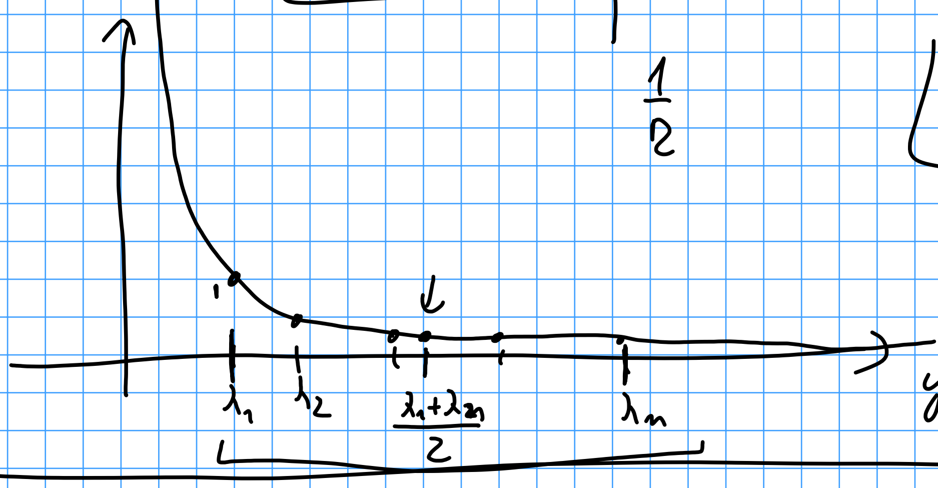
$$x^* = \underbrace{q^*(A)}_{A^{-1}} b \qquad q^*(y) = \frac{1}{y}$$

$$\forall i \qquad q_{t,\eta}(\lambda_i) \approx \frac{1}{\lambda_i} \qquad \text{degree } t$$

$$\boxed{q_{t,\eta}(y) = \sum_{i=1}^{t} \eta (1-\eta y)^i}$$

$$\frac{1}{\eta}$$

$$\boxed{\eta = \frac{2}{\lambda_1 + \lambda_n}}$$



Chebyshev polynomials $\Rightarrow O\left(\sqrt{\kappa} \log \frac{1}{\varepsilon}\right)$

$O\left(\sqrt{k}\,\log\frac{1}{\varepsilon}\right)$

$$A = \begin{pmatrix} 1000000 & 2500 & 0.1 \\ 2500 & 1000 & 0.2 \\ 0.1 & 0.2 & 1 \end{pmatrix} \qquad \kappa \approx 10^6$$

Preconditioning

$$\underbrace{D^{-1}A}_{\kappa \approx 1}\, x = D^{-1}b \qquad D = \begin{pmatrix} 10^6 & & 0 \\ & 10^3 & \\ 0 & & 1 \end{pmatrix}$$

① $P^{-1}A \approx I$ $\qquad P = A$

② multiply by P^{-1} is fast

$\Leftrightarrow \quad Py = z$

Laplacians $\qquad \kappa \approx n^2 \quad \to \quad O(n) \quad \Rightarrow \quad O(mn)$

$\to$ preconditioning Laplacians $\Rightarrow$ $\kappa \approx O(\log n)$

$\Rightarrow O(\log n)$