FYS4411/9411 MARCH 3, 2022

$\mathbb{E}[E_L(\vec{\alpha})]$

$$\vec{\nabla}_\alpha \mathbb{E}[E_L(\vec{\alpha})] = 0 \rightarrow 0 = f(\vec{x})$$

$$f(\vec{x}) \simeq f(\vec{x}_0) + (\vec{x}-\vec{x}_0)^T \nabla f(\vec{x}) + \frac{1}{2}(\vec{x}-\vec{x}_0)^T H (\vec{x}-\vec{x}_0)$$

$$\vec{x}_{k+1} = \vec{x}_k - \left[H(f(\vec{x}_k))\right]^{-1} \vec{\nabla} f(\vec{x}_k)$$

$$x_{k+1} = x_k - \gamma_k \vec{\nabla} f(\vec{x}_k)$$

Gradient descent, γ_k is a parameter to finetune.

— Steepest descent

$H \rightarrow A$

$A \in \mathbb{R}^{n \times n}$, positive definite

The problem of solving

$$\boxed{Ax = b}$$

is equivalent with the minimization of

$$q(x) = x^T A x - 2 x^T b$$

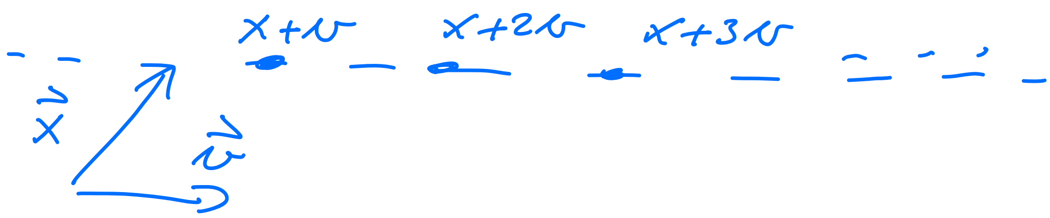
$$\left(\frac{1}{2} x^T A x - x^T b \right)$$

known

Example: 1-Dim ray

$$x + t \cdot v$$

$$v \in \mathbb{R}^n$$



$$q(x) = x^T A x - 2 x^T b$$

$$q(x+tv) = (x+tv)^T A (x+tv) - 2(x+tv)^T b$$

$$= q(x) + 2 t v^T (Ax - b) + t^2 v^T A v$$

optimal t

$$\frac{d\, q(x+tv)}{dt} = 0 = 2v^T(Ax-b) + 2tv^TAv$$

$$t^{opt} = \frac{v^T(Ax-b)}{v^TAv}$$

$$q(x+t^{opt}v) = q(x) + t^{opt}\left[2v^T(Ax-b) + v^T(b-Ax)\right]_2$$

$$= q(x) - \frac{\left[v^T(b-Ax)\right]^2}{v^TAv}$$

suggests an iterative scheme

$$x_{k+1} = x_k + t_k v_k$$

Link with steepest descent

$$Ax - b$$

Define residual

$$r = b - Ax$$

start with $x = x_0 + x^{final}$

$Ax_0 = b \quad (Ax_0 - b \neq 0)$

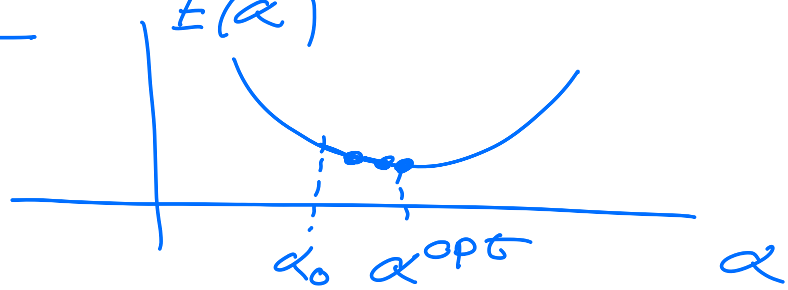
$$r_0 = b - Ax_0$$

$$r_{k+1} = b - Ax_{k+1}$$

continue $\quad \|r_{k+1} - r_k\|_2 \leq \varepsilon$

x_0 = guess (random value for example)

P1



$r_{k+1} = b - A\underline{x_{k+1}} \quad (= 0)$

$$x_{k+1} = x_k - t_k r_k$$

r_k = negative gradient of

$g(x)$ at $x = x_k$

$A =$ Hessian

$$\underbrace{b - Ax_k}_{r_k} - t_k A r_k = 0$$

$$r_k^T \,|\, r_k = t_k A r_k \Rightarrow$$

$$t_k = \boxed{\frac{r_k^T r_k}{r_k^T A r_k}}$$

requires knowledge of A $\left(= H = \nabla^2 \mathbb{E}(E_L)\right)$

$$x_{k+1} = x_k - t_k r_k$$

problematic

$t_k \rightarrow$ parameter γ_k

Project 1

Finding optimal $\vec{\alpha}$ for the large scale Monte Carlo calculation

(i) $\vec{\alpha}_{k+1} = \vec{\alpha}_k - \gamma_k \vec{\nabla} \mathbb{E}[E_L(\vec{\alpha}_k)]$

Gradient descent $\gamma_k \in [10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]$

$$\|\vec{\alpha}_{k+1} - \vec{\alpha}_k\|_2 \leq \varepsilon \sim 10^{-4} \sim 10^{-6}$$

(ii) quasi-Newton
Broyden et al

- need $\mathbb{E}[E_L(\vec{\alpha})]$, $\vec{\nabla}_\alpha \mathbb{E}[E_L(\vec{\alpha})]$

Conjugate gradient

- two vectors are conjugate if they are orthogonal wrt to the inner product

$$v_i^T A v_j = 0 \quad i \neq j$$

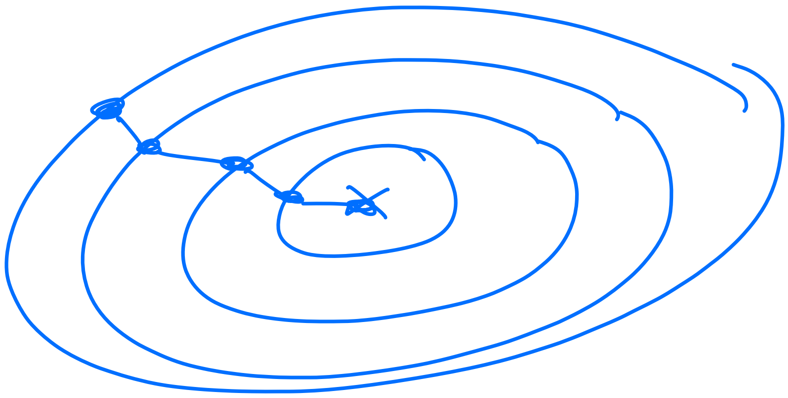
- Example: eigenvalue problem

$$v_i^T \Big| A v_j = \lambda_j v_j \Big|$$

$$v_i^T A v_i = \lambda_i v_i^T v_i =$$

$\lambda_i \delta_{ij}$

Steepest descent



p_i = a sequence of mutually conjugate direction $p \in \mathbb{R}^n$

$$x = \sum_{i=1}^{n} \alpha_i p_i$$

$$Ax = b \quad \left(\text{minimize } q(x) = \frac{1}{2} x^T A x - x^T b \right)$$

$$Ax = \sum_{i=1}^{n} \alpha_i A p_i = b$$

multiply with p_k^T

$$p_k^T A x = \sum_{i=1} \alpha_i \underbrace{p_k^T A p_i} = p_k^T b$$

$$\alpha_k p_k^T A p_k = p_k^T b \Rightarrow$$

$$\alpha_k = \frac{p_k^T b}{p_k^T A p_k}$$

$$Ax = b$$

$$r_k = b - A x_k$$

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

$$p_{k+1} = r_k - \frac{p_k^T A r_k}{p_k^T A p_k} p_k$$

$$b - A(x_k + \alpha p_k)$$

$$= (b - A x_k) - \alpha_k A p_k \Rightarrow$$

$$r_{k+1} = r_k - A p_k$$