FYS4411 FEB 25

$$\nabla_\alpha \mathbb{E}[E_L(\alpha)] = 0$$

$$\mathbb{E}[E_L(\alpha)] \to f(x) \quad x \in \mathbb{R}^n$$

Newton-Raphson:
Taylor expansion

$$f(x) \simeq \underline{f(x_0)} + \overbrace{(x-x_0)^T \nabla f(x_0)}^{x^T b}$$

$$+ \frac{1}{2} \overbrace{\underline{(x-x_0)^T H (x-x_0)}}^{x^T A x}$$

$$x_{k+1} = x_k - \underline{\left[\bar{H}(f(x_k))\right]^{-1}} \nabla f(x_k)$$

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

steepest descent and conjugate gradient -

$$H \to Ax = b \qquad A \in \mathbb{R}^{n \times n}$$

symmetric

$$x^T A x > 0$$

A positive definite

The problem of solving $\boxed{\underline{Ax = b}}$ is equivalent with the problem

by minimizing the quadra-
tic form

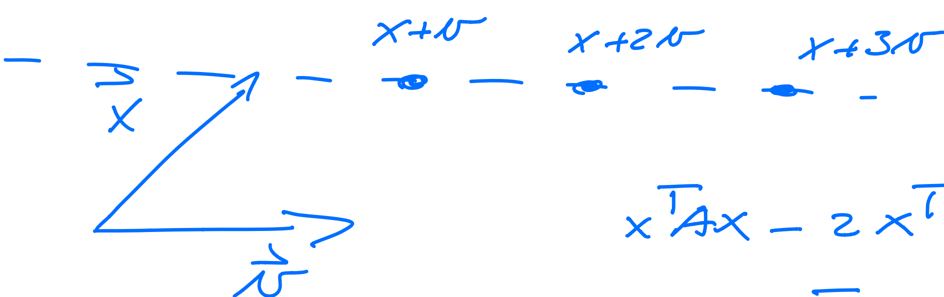
$$\boxed{q(x) = x^TAx - 2x^Tb}$$

$$\left(\frac{1}{2}x^TAx - x^Tb\right)$$

– one-dim ray $x + t\cdot v$

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

t is a scalar.



$$x^TAx - 2x^Tb = q(x)$$

$$q(x+tv) = (x+tv)^TA(x+tv) - 2(x+tv)^Tb$$

$$= q(x) + 2t\,v^T(Ax-b) + t^2 v^TAv$$

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

$$= 0$$

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

$$\overline{}$$
$$v^T A v$$

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

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

suggest an iterative scheme:

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

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

Steepest descent

$$Ax = b$$

Define a residual

$$r = b - Ax$$

$$A x_0 = b$$

$$r_0 = b - A x_0$$

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

continue iterating till

$$\| r_{k+1} - r_k \|_2 \leq \varepsilon \sim 10^{-10}$$

x_0 = guess, e.g. random values.

<u>Theorem</u> if A is positive definite the iterative process

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

converges to the exact x after a given number of iterations – irrespective of guess for x_0?–

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

$$x_{k+1} = x_k - t_1 r_k$$

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

$$\underline{r_k} - \underline{t_k} \underline{A r_k} = 0 \Rightarrow$$

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

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

Note: r_k is the negative

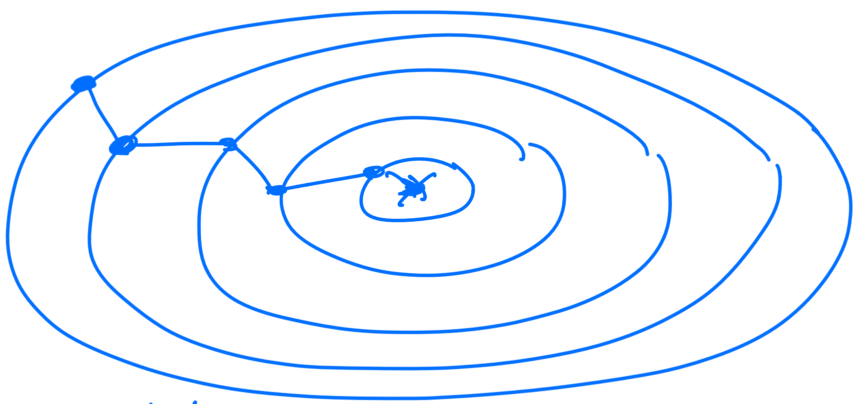
gradient of $q(x)$ at $x = x_k$

→ in our case A = Hessian

$t_k \to \gamma$ (learning rate)

$$\boxed{\begin{aligned} Ax &= b \\ q(x) &= x^T A x - 2x^T b \end{aligned}}$$

steepest descent (contour)



pretty slow.

Conjugate gradient

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

$$\boxed{v_i^T A v_j} = 0$$

- Example: eigenvalue

Example ... problem

$$\underline{v_i^T A v_j} = v_i \lambda_j v_j = \lambda_j v_i^T v_j$$
$$= \lambda_j \delta_{ij}$$

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

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

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

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

$$\alpha_k = \frac{p_k^T \underline{b}}{p_k^T A p_k} \qquad (Ax=b)$$

$Ax = b$; Define Residual r_k

$$\underline{r_k = \underline{b} - A x_k}$$

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

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

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

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

$$\Rightarrow \quad \underline{r_{k+1} = r_k - A p_k}$$