GRADIENT METHOD

"To iterate is human."   Anon.

Consider a linear system $Ax = b$.
The solution minimises the objective function

$$\frac{1}{2} x^T A x - x^T b \qquad (!)$$

QUADRATIC FORMS ; HESSIANS

Recall multivariate Taylor expansion:

$$f(a+h) \simeq \sum_{j=0}^{m} \frac{(h^T\nabla)^j f(a)}{j!}$$

At a critical point $\nabla f(x) = 0$ leads to

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

EXAMPLE $f = f(a,b) : \mathbb{R}^2 \to \mathbb{R}$; $\underline{h} = (h, k)^T$

We get: $\frac{1}{2}\Big(h^2 f_{11}(a,b) + hk\, f_{12}(a,b) +$
$\qquad k h f_{21}(a,b) + k^2 f_{22}(a,b) \Big)$

$$= \frac{1}{2} \begin{pmatrix} h \\ k \end{pmatrix}^T \underbrace{\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}}_{\text{Hessian : } H_f(x)} \begin{pmatrix} h \\ k \end{pmatrix}$$

PROPERTIES OF HESSIANS

H_f is symmetric and real $\Rightarrow$ all eigenvalues are real

<u>Definition</u> $A \in \mathbb{R}^{n\times n}$; $A = A^T$

A is

(i) positive definite : $\lambda_i > 0$

(ii) negative definite : $\lambda_i < 0$

(iii) indefinite : $\lambda_i > 0$, $\lambda_j < 0$, $i \neq j$

If A is pos. def, than $x^T A x > 0$ for all $x \in \mathbb{R}^n$.

$A = Q \Lambda Q^T$; Q orthogonal , $Q = (v_1 \; v_2 \dots v_n)$

Any $y = \sum_{i=1}^{n} \alpha_i v_i \Rightarrow y^T A y = \alpha_1^2 \lambda_1 + \alpha_2^2 \lambda_2 + \dots + \alpha_n^2 \lambda_n$

> 0 only if $\lambda_i > 0$.

SYLVESTER CRITERION

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2n} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \cdots & \vdots \\ \vdots & & & \ddots & \\ \alpha_{n1} & & \cdots & & \alpha_{nn} \end{pmatrix}$$

$$\Delta_1 = |\alpha_{11}|$$

$$\Delta_2 = \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix}$$

$$\Delta_3 = \cdots$$

If for all $k = 1, \dots, n$
$\Delta_k > 0$, then A is pos. def.

If the signs alterente, then
A is neg. def.

Why: $\det(-A) = (-1)^n |A|$

GRADIENT METHOD

First we must choose our norm: $\|v\|_A = \sqrt{v^T A v}$

$$\phi(x) = \frac{1}{2} x^T A x - x^T b \quad ; \quad \text{Let } A x_* = b.$$

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

$$= \underbrace{\frac{1}{2}\|x - x_*\|_A^2}_{\text{If we can find } x \to x_*, \text{ then the squared norm } \to 0.} + \phi(x_*)$$

Idea: Take steps in the negative gradient direction!

ITERATION

$$x_{k+1} = x_k - \mu_k g_k$$

$$g_k = A x_k - b \quad \text{(gradient)}$$

$$\mu_k = \frac{g_k^T g_k}{g_k^T A g_k} \in \mathbb{R},$$ step length that minimises $\phi(x_{k+1})$.

Direct computation:

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