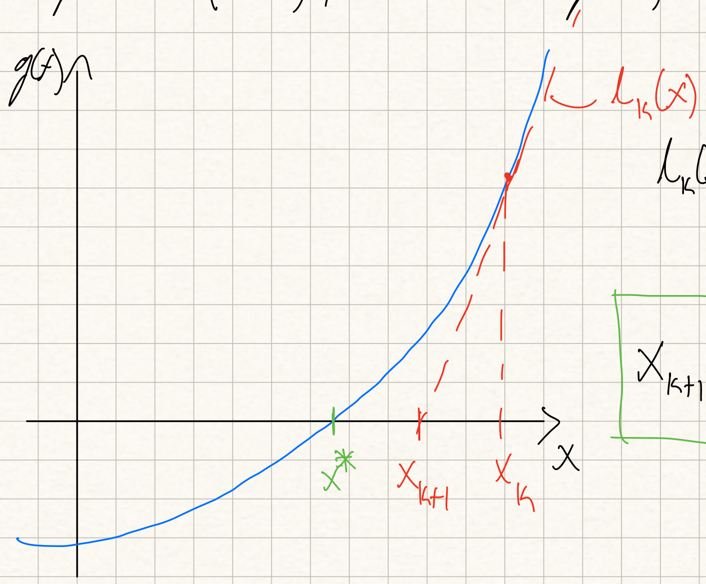
Lecture 6 - Newton and Quasi-Newton methods

Newton-Raphson method

a strictly convex function of one variable.

$g(x) = f'(x)$, find $g(x) = 0$



$$l_k(x) = g(x_k) + g'(x_k)(x - x_k)$$

$$\boxed{x_{k+1} = x_k - \frac{1}{g'(x_k)} g(x_k)}$$

In multivariable:

$$l_k(x) = g_k + J(x_k)(x - x_k)$$

Newton method for minimizing functions

a) Solve $\nabla f(x^*) = 0$

$$l_k(x) = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0 \Big|_{x = x_{k+1}}$$

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

b) Minimize a quadratic form of f

$$h_k(d) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla^2 f(x_k) d$$

$h_k(d)$ approximates $f(x_k + d)$ $\forall d \in \mathbb{R}^n$

Select $x_{k+1} = x_k + d_k$ with
$d_k = \arg\min h_k(d)$

$$\nabla h_k(d) = \nabla^2 f(x_k) d + \nabla f(x_k)$$

at d_k we have

$$\nabla h_k(d_k) = 0 \Rightarrow \nabla^2 f(x_k) d_k + \nabla f(x_k) = 0$$

$$\Rightarrow d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

$$\Rightarrow x_{k+1} = x_k + d_k = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

The local convergence of the Newton method is quadratic.

Observations about the Newton method:

1) Only good convergence rate locally
2) A costly method computationally:
   - have to calculate the gradient
   - have to calculate the Hessian

- have to take the inverse of the Hessia

Newton method:

$$x_{k+1} = x_k + d_k, \text{ where}$$

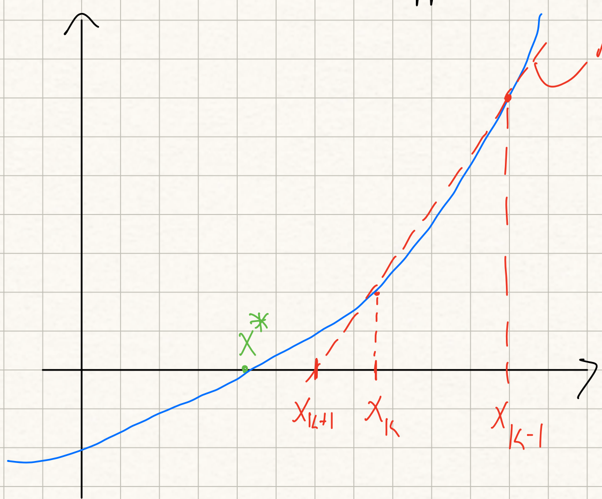
$$d_k = -\left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$$

Quasi-Newton Method

The Newton method is a tangent method (it uses the tangent of the graph $g(x)$ at point x_k).

Now, we will use a secant method.

We will use an approximation of $g'(x_k)$



$$s_k = \frac{g(x_{k-1}) - g(x_k)}{x_{k-1} - x_k}$$

$$l_k(x) = g(x_k) + s_k(x - x_k)$$

$$l_k(x) = g(x_k) + g'(x_k)(x - x_k)$$

Approximate $g'(x_k)$ with s_k.

s_k is the slope of the secant line connecting $(g(x_{k-1}), x_{k-1})$ and $(g(x_k), x_k)$

$$\boxed{x_{k+1} = x_k - \frac{1}{s_k} g(x_k)}$$

Mean value theorem;

f continous on $[a,b]$, differentiable on (a,b), there exists a $a<c<b$ such that

$$f'(c) = \frac{f(b)-f(a)}{b-a}$$

slope of tangent of graph at c = slope of secant line between b and a

$$f'(c)(b-a) = f(b)-f(a)$$

in our case, this becomes;

$$\nabla^2 f(c)(x_{k-1}-x_k) = \nabla f(x_{k-1}) - \nabla f(x_k)$$

$$c \in (x_{k-1}, x_k)$$

$$\nabla f(x_k) = \nabla f(x_{k-1}) - \nabla^2 f(c)(x_{k-1}-x_k)$$
$$= \nabla f(x_{k-1}) + \nabla^2 f(c)(x_k - x_{k-1})$$

If S_k is a good approximation of the Hessian $\nabla^2 f(x_k)$ we have

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

Define $\gamma_k = \nabla f(x_k) - \nabla f(x_{k-1})$, $\delta_k = x_k - x_{k-1}$, gives

$$\boxed{S_k \delta_k = \gamma_k}$$ Quasi-Newton equation

Usually, $S_0 = I$