Jensen's Inequality

$f: \mathbb{R} \to \mathbb{R}$, convex, $a_i > 0$, $i = 1, 2 \cdots m$

$f\left(\frac{\sum a_i x_i}{\sum a_i}\right) \le \frac{\sum a_i f(x_i)}{\sum a_i}$ "=" holds if all x_i are the same or f is linear.

Ex: $m = 2$ $a_1 = \alpha$ $a_2 = 1-\alpha$

$\Rightarrow$ convex definition: $f(\alpha x_1 + (1-\alpha) x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2)$

① $f = x^2$ $a_i = 1$, $\forall i$

$\left(\frac{\sum x_i}{m}\right)^2 \le \frac{\sum x_i^2}{m} \Rightarrow (\sum x_i)^2 \le m \sum x_i^2$

② $f(x) = -\ln x$

$a_i = 1$, $\forall i$

$\Rightarrow \frac{\sum x_i}{n} \ge (\prod x_i)^{\frac{1}{n}}$

$f(a, b)$ convex

$\Rightarrow \nabla f = 0 \Rightarrow \begin{cases} a = \\ b = \end{cases} \cdots$

$X = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}$ $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$

$f(a, b) = \frac{1}{2} |X\theta - y|^2 = \frac{1}{2}(X\theta - y)^T (X\theta - y) = \frac{1}{2}(\theta^T X^T X\theta - 2y^T X\theta + y^T y)$

$\nabla f(\theta) = (X^T X\theta)^T - y^T X$

$\nabla^2 f(\theta) = X^T X$

$y^T \nabla^2 f(\theta) y = (Xy)^T Xy = \|Xy\|^2 \ge 0$, so $\nabla^2 f(\theta)$ always semidefinite.

$\nabla f(\theta) = 0 \Rightarrow (X^T X\theta)^T = y^T X$ $X^T X\theta = X^T y$ $\theta = (X^T X)^{-1} X^T y$: global minimum. (unique)

Unconstrained optimization in 1D

$\min f(x) \quad x \in [x_s, x_t]$

Ex: $\min f(x) = 12x^6 + 3x^4 - 12x + 7$

s.t. $x \in [0, 1]$

$f'(x) = 72x^5 + 12x^3 - 12$

$f''(x) = 360x^4 + 36x^2 \geq 0 \Rightarrow f$ convex

$f'(x) = 0 \Rightarrow x = 0.654$

Derivative-free methods / 0th order method

$f: [a,b] \to \mathbb{R}$ is unimodal if f has a unique global minimum $x^* \in (a,b)$.

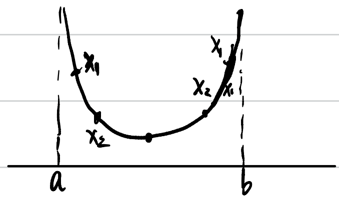
Thm: If f is unimodal in $[a,b]$ with min point x^* in (a,b)

let x_1, x_2 $a < x_1 < x_2 < b$ Then

① $f(x_1) < f(x_2) \Rightarrow x^* \in (a, x_2)$

② $f(x_1) > f(x_2) \Rightarrow x^* \in (x_1, b)$.

③ $f(x_1) = f(x_2) \Rightarrow x^* \in (x_1, x_2)$



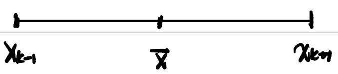
proof: ① $f(x_1) < f(x_2)$. Suppose $x^* \geq x_2$, $f(x_2) \leq f(x^*) < f(x_1)$.

Exhaustive search method

f is unimodal in $[a,b]$, let $a < x_1 < x_2 \cdots < x_N < b$

let $x_k = \underset{1 \leq i \leq N}{\operatorname{argmin}} f(x_i)$, The $x^* \in \boxed{(x_{k-1}, x_{k+1})}$ interval of uncertainty.

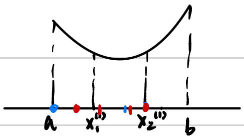
$L_0 = b - a \quad h = \frac{b-a}{N+1} = \frac{L_0}{N+1} \qquad \frac{2L_0}{N+1} < \varepsilon \qquad N > \frac{2L_0}{\varepsilon} - 1$

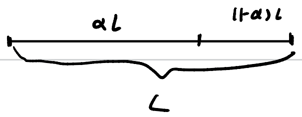


$\bar{x} = \frac{x_{k-1} + x_{k+1}}{2} \quad \Rightarrow |\bar{x} - x^*| \leq \frac{x_{k+1} - x_{k-1}}{2}$

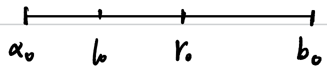
Computational Cost.

Golden section method





$\frac{1}{\alpha} = \frac{\alpha}{1-\alpha}$ $\alpha = 0.618$



$r_0 = a_0 + \alpha L_0$

$l_0 = a_0 + (1-\alpha)L_0$

① if $f(l_0) < f(r_0)$

$a_1 = a_0$ $b_1 = r_0$

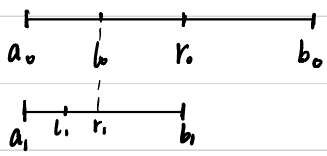
② otherwise

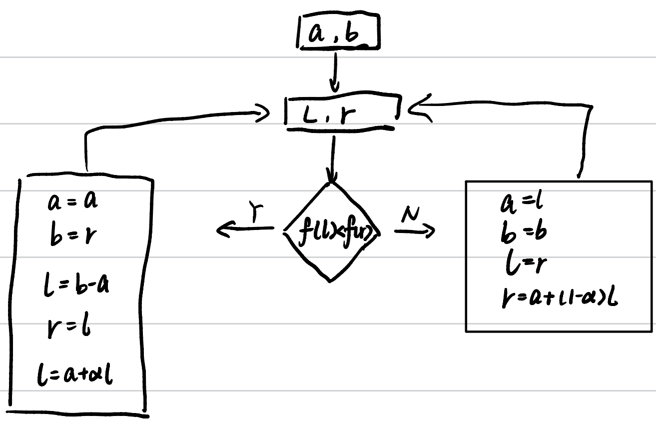
$a_1 = l_0$, $b_1 = b_0$

$L_1 = \alpha L_0$

Suppose case ① $r_1 = a_1 + \alpha L_1$

$l_1 = a_1 + (1-\alpha)L_1$





$L_0 = b - a$

$L_1 = \alpha L_0$

$L_2 = \alpha^2 L_0$

$\vdots$

$L_N = \alpha^N L_0 < \varepsilon$

(a_N, b_N)

$\Rightarrow$ $\alpha^N < \frac{\varepsilon}{L_0}$

$N > \log_\alpha \frac{\varepsilon}{L_0}$

function evaluations $= N+2$.

$N = O(10^d)$.

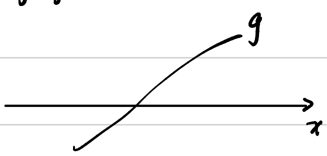
$N > \log_\alpha \varepsilon$.

1st order method.

bisection method.

f is differentiable, x^* is local minium, $f'(x^*)=0$.

let $g=f'$



intermediate value theorem,

$$\left.\begin{array}{l} \text{if } g(a)>0,\ g(b)<0 \\ \text{or } g(a)<0,\ g(b)>0 \end{array}\right\} \Longleftrightarrow g(a)g(b)<0$$

$\exists\ x^* \in (a,b)$ s.t. $g(x^*)=0$.

let $\bar{x}=\frac{a+b}{2}$

① if $g(a)g(\bar{x})<0$, then let $b=\bar{x}$

② if $g(\bar{x})g(b)<0$, then let $a=\bar{x}$.

③ if $g(\bar{x})=0$, then take $x^*=\bar{x}$

$L_n=b_n-a_n=2^{-n}L_0$

$2^{-n}L_0<\varepsilon$

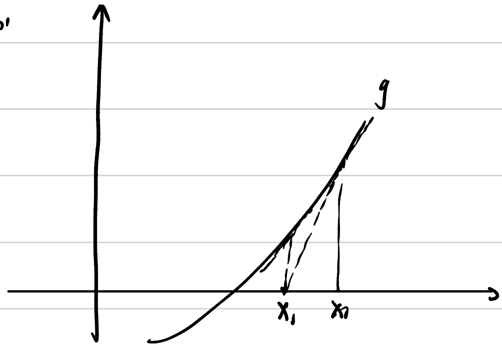
$N>\log_2\frac{L_0}{\varepsilon}$

$\log_{\frac{1}{2}}\frac{L_0}{\varepsilon}$

2nd order method

Newton's method

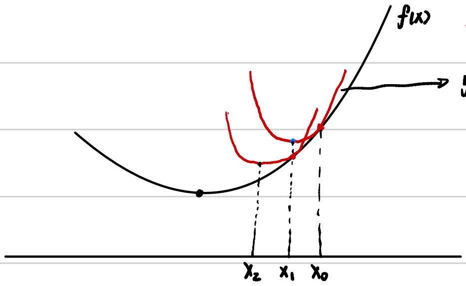
$f'(x^*)=0 \qquad g=f'$



$y=g(x_0)+g'(x_0)(x-x_0)=0$

$x_1=x_0-\frac{g(x_0)}{g'(x_0)}$.

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



$y=f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2$

$x_1=-\frac{f'(x_0)-f''(x_0)x_0}{f''(x_0)}$

So $x_{k+1}=-\frac{f'(x_k)-f''(x_k)x_k}{f''(x_k)}$

Secant method (1st order method)

$x_{k+1}=x_k-\left[\frac{x_k-x_{k-1}}{f'(x_k)-f'(x_{k-1})}\right]f'(x_k)$

Speed of convergence

$x_k \to x^* \qquad k\to\infty$

$0<\lim_{k\to\infty}\frac{|x_{k+1}-x^*|}{|x_k-x^*|^p}=\beta<\infty$ $\qquad$ p: order of convergence.

Ex: $x_k=\frac{1}{k}\to 0=x^*$

$\frac{|x_{k+1}-x^*|}{|x_k-x^*|}=\frac{\frac{1}{k+1}}{\frac{1}{k}}=\frac{k}{k+1}\to 1.$

$x_k=a^k \qquad a\in(0,1) \qquad x_k\to 0.$

$\frac{|x_{k+1}-x^*|}{|x_k-x^*|}=\frac{a^{k+1}}{a^k}=a.$

$x_k = a^{2^k} \to 0 = x^*$

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \frac{a^{2^k}}{a^{2^{k-1}}} = a^{2^k - 2^{k-1}} = a^{2^{k-1}} \to 0$$

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} = \frac{a^{2^k}}{a^{2^k}} = 1$$

① if $p=1$. we say $x_k \to x^*$ linearly, β is called convergence ratio

② if $p=2$, we say $x_k \to x^*$ quadratically

③ if $1<p<2$ or $(p=1, \beta=0)$ we say $x_k \to x^*$ superlinearly

Theorem: Suppose $f \in C^2(\mathbb{R})$ $f'(x^*)=0$ $f''(x^*) \neq 0$ x_0 is sufficiently close to x^*, x_k: from newton's method, then $x^k \to x^*$

convergence order ≥ 2.

secant method. order ≈ 1.618

Multi-demensional algorithms.

steepest Descent method

deriictional derivative

$f \in C^1$ $D_u f(x) = \lim_{\alpha \to 0} \frac{f(x+\alpha u) - f(x)}{\alpha}$

$g(\alpha) = f(x+\alpha u)$, $g'(0) = D_u f(x)$ $g'(0) = \nabla f(x) u$: by the chain rule.

So $D_u f(x) = \nabla f(x) u$

given $\|u\| = 1$, then $D_u f(x)$ is max when $u = \frac{\nabla f(x)}{|\nabla f(x)|}$

$\nabla f(x)$: direction of steepest ascent.

① select initial guess $x_0 \in \mathbb{R}^n$

set $k=0$

② let $g_k = [\nabla f(x_k)]^T$

$\varphi_k(\alpha) = f(x_k - \alpha g_k)$

$\min_{\alpha \in [0,\infty)} \varphi_k(\alpha) \Rightarrow \alpha_k^*$

③ $x_{k+1} = x_k - \alpha_k^* g_k$

④ let $k=k+1$, go to ②.

Stopping Criterion

① $k > k$

② $|\nabla f(x_k)| < \varepsilon$

③ $|x_{k+1} - x_k| < \delta_1$

④ $|f(x_{k+1}) - f(x_k)| < \delta_2$

$\varphi_k(\alpha) = f(x_k - \alpha g_k)$

$\varphi_k'(\alpha) = \nabla f(x_k - \alpha g_k)(-g_k)$

$\varphi_k'(\alpha_k) = 0 \Rightarrow \nabla f(x_k - \alpha_k g_k) g_k = 0 \quad \nabla f(x_{k+1}) g_k = 0 \qquad g_{k+1}^T g_k = 0$

Quadratic Case

$f(x) = \frac{1}{2} x^T A x - b^T x \qquad$ A: spd (symmetric positive definite).

$(Ax)_i = \sum_j a_{ij} x_j$

$x^T(Ax) = \sum_i x_i (Ax_i)$

$= \sum_i \sum_j a_{ij} x_i x_j$

$\nabla^2 f(x) = A \qquad \nabla f(x) = (Ax-b)^T = 0 \qquad x^* = A^{-1} b.$ is the unique global minium.

① $x_0 = 0 \quad k=0$

② $g_k = [\nabla f(x_k)]^T = Ax_k - b$

③ $\varphi_k(\alpha) = f(x_k - \alpha g_k) = \frac{1}{2}(x_k - \alpha g_k)^T A (x_k - \alpha g_k) - b^T(x_k - \alpha g_k).$

$= \frac{1}{2}[x_k^T A x_k - (2 g_k^T A x_k)\alpha + (g_k^T A g_k)\alpha^2] - b^T x_k + (b^T g_k)\alpha$

$= \frac{1}{2}(g_k^T A g_k)\alpha^2 + (b^T g_k - g_k^T A x_k)\alpha + (\frac{1}{2} x_k^T A x_k - b^T x_k).$

if $g_k = 0$, then $x_k = x^*$, stop.

if $g_k \neq 0$. $\quad g_k^T A g_k > 0 \qquad \varphi'(\alpha) = (g_k^T A g_k)\alpha + (b^T g_k - g_k^T A x_k) = 0$

$$\alpha_k = \frac{g_k^T A x_k - b^T g_k}{g_k^T A g_k} = \frac{g_k^T (Ax_k - b)}{g_k^T A g_k} = \frac{g_k^T g_k}{g_k^T A g_k}$$

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

Stopping Criterion

Convergent analysis

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

let $E(x) = f(x) - f(x^*)$

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

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

$E(x_k) = \frac{1}{2}(x_k - x^*)^T A (x_k - x^*)$, $E(x_k) \to 0$? How fast?

lemma

$E(x_{k+1}) = \left[1 - \frac{(g_k^T g_k)^2}{(g_k^T A g_k)(g_k^T A^{-1} g_k)}\right] E(x_k)$

proof: $E(x_k) - E(x_{k+1}) = E(x_k) - E(x_k - \alpha_k g_k) = \frac{1}{2}(x_k - x^*)^T A (x_k - x^*) - \frac{1}{2}(x_k - x^* - \alpha_k g_k)^T A (x_k - x_k^* - \alpha_k g_k)$.

let $y = x_k - x^*$

$= \frac{1}{2}(y_k^T A g_k) - \frac{1}{2}(y - \alpha_k g_k)^T A (y_k - \alpha_k g_k)$.

$= \alpha_k (g_k^T A y_k) - \frac{1}{2}\alpha_k^2 g_k^T A g_k$

$= \alpha_k (g_k^T g_k) - \frac{1}{2}\alpha_k^2 g_k^T A g_k$

$A y_k = A x_k - A x^*$, $g_k = A x_k - b$, $A x^* = b$

$= g_k + b - b$, back substitute.

$\alpha_k (g_k^T g_k) - \frac{1}{2}\alpha_k^2 g_k^T A g_k = \frac{(g_k^T g_k)^2}{g_k^T A g_k} - \frac{1}{2}\frac{(g_k^T g_k)^2}{g_k^T A g_k} = \frac{1}{2}\frac{(g_k^T g_k)^2}{g_k^T A g_k} > 0$

$E(x_k) = \frac{1}{2} y_k^T A y_k = \frac{1}{2} y_k^T \cdot g_k = \frac{1}{2}(A^{-1} g_k)^T g_k = \frac{1}{2} g_k^T (A^{-1})^T g_k = \frac{1}{2} g_k^T A^{-1} g_k$

$\frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{(g_k^T g_k)^2}{(g_k^T A g_k)(g_k^T A^{-1} g_k)} \Rightarrow E(x_{k+1}) = \left[1 - \frac{(g_k^T g_k)^2}{(g_k^T A g_k)(g_k^T A^{-1} g_k)}\right] E(x_k)$.

lemma: kantorovich inequality

let $0 < a < x_i \le b$ $i = 1, 2 \cdots n$

$\alpha_i \ge 0$, $\sum_{i=1}^n \alpha_i = 1$, then $\left(\sum_{i=1}^n \alpha_i x_i\right)\left(\sum_{i=1}^n \alpha_i x_i^{-1}\right) \le \frac{(a+b)^2}{4ab}$

$\forall x \in \mathbb{R}$, let $0 \le \lambda_1 \le \cdots \le \lambda_n$

$\Rightarrow \frac{(x^T a x)(x^T a^{-1} x)}{(x^T x)^2} \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}$

lemma: $A \in \mathbb{R}^{n \times n}$, SPD, eigenvalues $\lambda_1 \ge \lambda_2 \cdots \lambda_n > 0$

$\Rightarrow \frac{(x^T x)^2}{(x^T A x)(x^T A^{-1} x)} \ge \frac{4\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2}$ $\forall x \neq 0$.

proof: $A = Q^T D Q$, $Q^T Q = I$, $D = diag(\lambda_1, \lambda_2 \cdots \lambda_n)$

$x^T A x = x^T Q^T D Q x = (Qx)^T D (Qx)$

let $y = Qx$ $\quad x^T A x = y^T D y = \sum_{i=1}^n \lambda_i y_i^2$

$x^T A^{-1} x = y^T D^{-1} y = \sum_{i=1}^n \lambda_i^{-1} y_i^2$

$x^T x = (Q^{-1} y)^T (Q^{-1} y) = y^T (Q^{-1})^T Q^{-1} y = y^T y = \sum y_i^2 = \|y\|^2$

$\Rightarrow \frac{(x^T x)^2}{(x^T A x)(x^T A^{-1} x)} = \frac{(y^T y)^2}{\sum \lambda_i y_i^2 \sum \lambda_i^{-1} y_i^2} = \frac{1}{\sum \lambda_i \left(\frac{y_i}{\|y_i\|}\right)^2 \sum \lambda_i^{-1} \left(\frac{y_i}{\|y_i\|}\right)^2}$

let $z = \frac{y_i}{\|y_i\|}$ so $\frac{1}{\sum \lambda_i z_i^2 \sum \lambda_i^{-1} z_i^2} \ge \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}$

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

$$\frac{(x^T x)^2}{(x^T A x)(x^T A^{-1} x)} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_2)^2}$$

Thm: A: spd: eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n > 0$

$$E(x_{k+1}) \leq \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 E(x_k).$$

proof: $$1 - \frac{(g_k^T g_k)^2}{(g_k^T A g_k)(g_k^T A^{-1} g_k)} \leq 1 - \frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2} = \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_2}\right)^2$$

$$\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right) < 1 \Rightarrow E(x_k) \to 0.$$

Convergence order : at least 1

Convergence ratio : $\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2$

let $\delta = \frac{\lambda_1}{\lambda_n}$: condition number of A

$$\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 = \left(\frac{\delta-1}{\delta+1}\right)^2$$

- if δ is close to 1, then converge fast.
- if δ is large, then converge slow.

Ex: $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\lambda_1 = \lambda_2 = 1$ $\quad \delta = 1$

$$E(x_1) \leq 0 \cdot E(x_0)$$

$$E(x_1) = 0.$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}$$

$\lambda_1 = 10 \quad \lambda_2 = 1, \quad \delta = 10.$

$$\left(\frac{\delta-1}{\delta+1}\right)^2 = \left(\frac{9}{11}\right)^2$$

$$f(x_1, x_2) = (x_1 + 2x_2 + 3)^2 + x_1^2 = 2x_1^2 + 4x_2^2 + 4x_1x_2 + 3x_1 + 6x_2 + 9$$

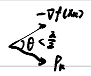
$$= [x_1, x_2] \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [3, 6] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 9.$$

$A = \begin{bmatrix} 4 & 4 \\ 4 & 8 \end{bmatrix} \quad \lambda = 6 \pm \sqrt{10} \quad \delta = \frac{6+\sqrt{10}}{6-\sqrt{10}}$

General Line search methods.

$P_k \in \mathbb{R}^n$: descent direction

$\nabla f(x_k) P < 0$



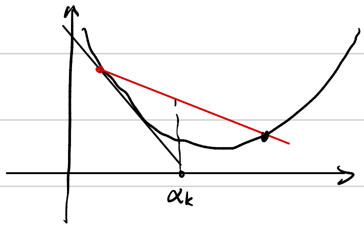
$m > 0 \quad |P_k| \geq m |\nabla f(x_k)|, \forall k.$

$\varphi_k(\alpha) = f(x_k + \alpha P_k).$

Armijo Condition

$f(x_k + \alpha P_k) \approx f(x_k) + \alpha \nabla f(x_k) P_k$

$f(x_k + \alpha_k P_k) \leq f(x_k) + \mu \alpha_k \nabla f(x_k) P_k \quad \mu \in (0,1).$



Sufficient decreasing Condition

backtracking

① $\alpha_k = 1.$

② if Armijio Condition is satisfied, done

if not, set $\alpha_k = \frac{1}{2}$

repeat, $\alpha_k = (\frac{1}{2})^k.$

Assume

$S := \{x : f(x) \leq f(x_0)\}$ is bounded

Assume $\|\nabla f(x) - \nabla f(y)\| - L|x - y|$: Lipschitz condition for ∇f

Thm: under all above assumption $x_{k+1} = x_k + \alpha P_k$ Then $\nabla f(x_k) \to 0$

P_k is called a descent direction at x_k is $\nabla f(x_k) \cdot P_k < 0$

Sufficient descent if $-\frac{\nabla f(x_k) \cdot P_k}{\|\nabla f(x_k)\| \|P_k\|} \geq \varepsilon \quad \varepsilon \in (0,1].$

Steepest descent

$P_k = -\nabla f(x_k).$

$-\frac{\nabla f(x_k) \cdot P_k}{\|\nabla f(x_k)\| \|P_k\|} = 1.$

Newton's method for find root.

find x^* s.t. $g(x^*) = 0.$

$g(x^* + p) \approx g(x_k) + g'(x_k) \cdot p = 0 \quad \Rightarrow p = -\frac{g(x_k)}{g'(x_k)}$

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

$0 = g(x^*) = g(x_k) + g'(x_k)(x^* - x_k) + \frac{1}{2} g''(\xi_k)(x_k - x^*)^2 \quad \xi_k$ between x^* and x_k

let $e_k = x_k - x^*$

$$0 = g(x_k) - g'(x_k)e_k + \frac{1}{2}g''(\xi)e_k^2$$

$$e_{k+1} = x_{k+1} - x^*$$

$$= \left[x_k - \frac{g(x_k)}{g'(x_k)}\right] - x^* = e_k - \frac{g(x_k)}{g'(x_k)}$$

$$e_{k+1} = \underbrace{\frac{g''(\xi_k)}{2g'(x_k)}}_{\frac{g''(x^*)}{2g'(x^*)}} e_k^2 \qquad g'(x^*) \neq 0$$

if $g'(x^*) = 0$, may still converge, but quadratic convergence isn't guaranteed.

Thm: Suppose $g \in C^2(\mathbb{R})$, $g(x^*) = 0$, $g'(x^*) \neq 0$

$|x_0 - x^*|$ sufficiently small

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

Then $x_k \to x^*$ at least quadratically with rate $\frac{g''(x^*)}{2g'(x^*)}$

$$0 = g(x^*) + g'(\xi_k)(x^* - x_k) = g(x_k) - g'(\xi_k)e_k \Rightarrow g(x_k) = g'(\xi_k)e_k \sim g'(x^*)e_k \to 0$$

$g(x) = \tanh x$

n-dim case

$$\begin{cases} g_1(x_1 \cdots x_n) = 0 \\ g_2(x_1 \cdots x_n) = 0 \\ \vdots \\ g_n(x_1 \cdots x_n) = 0 \end{cases} \qquad g(\vec{x}) = \begin{bmatrix} g_1(\vec{x}) \\ \vdots \\ g_n(\vec{x}) \end{bmatrix}$$

$$\nabla g(\vec{x}) = \begin{bmatrix} \nabla g_1(\vec{x}) \\ \vdots \\ \nabla g_n(\vec{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix} \text{ : Jacobian matrix of } \vec{g}$$

$$\vec{g}(\vec{x_k+p}) \approx \vec{g}(\vec{x_k}) + \nabla\vec{g}(x_k)\vec{p} = 0 \Rightarrow \vec{p} = [\nabla\vec{g}(x_k)]^{-1}\vec{g}(\vec{x_k})$$

$$\vec{x_{k+1}} = \vec{x_k} - [\nabla\vec{g}(x_k)]^{-1} g(\vec{x_k}).$$

Thm if $g \in C^2$ $\quad \vec{g}(\vec{x^*}) = 0$

$\nabla\vec{g}(\vec{x^*})$ is nonsingular, $|\vec{x_0} - \vec{x^*}|$ sufficiently small

Then $\vec{x_k} \to \vec{x^*}$ quadratically

Newton's method for optimization

$f' = g \quad g(x^*) \Rightarrow f'(x^*) = 0$

$\min f(x) \Rightarrow \nabla f(x^*) = 0$

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

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

$$g = (\nabla f)^T = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$$\nabla g = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ & & \end{bmatrix} = \nabla^2 f$$

Another point of view

$f(x_k+p) \approx f(x_k) + \nabla f(x_k)p + \frac{1}{2}p^T\nabla^2 f(x_k)p$

$\min_p, \quad p = -[\nabla^2 f(x_k)]^{-1} [\nabla f(x_k)]^T$

Thm: Suppose $f: S \to \mathbb{R}$ $\quad S$: convex

$\nabla^2 f$ is Lipschitz continous

$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*)$: SPD

$|x_0 - x^*|$ is sufficiently small.

Then $x_k \to x^*$ quadratically.

Ex: $f(x_1,x_2) = 4x_1^2+2x_1x_2+9x_2^2$

$$f(x_1,x_2) = [x_1, x_2]\begin{bmatrix}4 & 1\\ 1 & 9\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

$$= \frac{1}{2}x^T A x \qquad x = \begin{bmatrix}8 & 2\\ 2 & 18\end{bmatrix}$$

$$\nabla f = (Ax)^T$$

$$\nabla^2 f = A$$

- Steepest descent

$$g_k = \nabla f(x^{(k)})$$
$$= [Ax^{(k)}].$$

$$\varphi(x^{(k)}) := f(x^{(k)} - \alpha_k g_k).$$

$$\alpha_k = \frac{g_k^T g_k}{g_k^T A g_k}$$

Newton's method

$$x^{(k+1)} = x^{(k)} - [\nabla^2 f(x^{(k+1)})]^{-1}[\nabla f(x^{(k+1)})]^T$$
$$= x^k - A^{-1}Ax^{(k)} = 0$$

Modified Newton's method

$$x^{(k+1)} = x^{(k)} - [\nabla^2 f(x^{(k)}) + \beta I].$$

β_k Large so that $\nabla^2 f(x^{(k)}) + \beta_k I$ is SPD

modified Newton's method with line search

$$p_k = -[\nabla^2 f(x^k) + \beta_k I]^{-1}[\nabla f(x_k)]^T$$

$$x^{(k+1)} = x^{(k)} - \alpha_k p_k$$

find α_k by a line search method.

$$\varphi(\alpha) = f(x_k + \alpha p_k).$$

$\min \varphi(\alpha) \Rightarrow \alpha_k$

Armijo's rule with backtracking

$$f(x_k + \alpha_k p_k) \le f(x_k) + \eta\alpha_k \nabla f(x_k) p_k \qquad \eta \in (0,1).$$

Quasin-Newton method.

classic newton's

- require 2nd derivatives
- $O(n^2)$ Storage
- $O(n^3)$ Computational complexity

Quasi-Newton

approximate $\nabla^2 f(x^{(k)})$ by easier Computation

secant method

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

$$f''(x_{k+1})(x_{k+1} - x_k) \approx f'(x_{k+1}) - f'(x_k)$$

$$\nabla^2 f(x^{(k+1)})(x^{k+1} - x^k) \approx \nabla f(x^{k+1}) - \nabla f(x^k).$$

Find B_{k+1} s.t.

$$B_{k+1}[x^{k+1} - x^k] = [\nabla f(x^{k+1}) - \nabla f(x^k)]^T$$

let $S_k = x^{(k+1)} - x^{(k)}$

$$y_k = [\nabla f(x^{k+1}) - \nabla f(x^k)]^T$$

$$\boxed{B_{k+1} S_k = y_k}$$

secant Condition

$B_{k+1} = B_k +$ update term.

① Symetric rank one update form

$B_{k+1} = B_k + C_k$

C_k: symetric, rank one

General Quasin-newton method

① $k=0$ x_0, B_0

② $P_k = -B_k^{-1} \nabla f(x_k)$

: search direction

③ find the step length α_k by a line search method.

$\varphi(\alpha) = f(x_k + \alpha P_k)$

④ $x_{k+1} = x_k + \alpha_k P_k$

⑤ $B_{k+1} = B_k +$ update term.

⑥ go back to ②

Symmetric rank-one update.

$$B_{k+1} = B_k + C_k$$

C_k: symmetric, rank-one

$C_k = \gamma w w^T$ $\quad$ γ: scalar

$\quad$ w: unit vector.

$$p_k = B_k^{-1} [\nabla f(x_k)]^T$$

$$B_k \cdot p_k = [\nabla f(x_k)]^T$$

$$\underbrace{B_k^{-1} \rightarrow B_{k+1}^{-1}}$$

efficient if C_k is rank-one.

$$B_{k+1} S_k = (B_k + C_k) S_k = (B_k + \gamma w w^T) S_k = B_k S_k + \gamma w (w^T S_k) = y_k$$

$$\boxed{\gamma (w^T S_k) w = y_k - B_k S_k.}$$

if $w^T S_k = 0$, then $y_k = B_k S_k$.

then just take $B_{k+1} = B_k$

Assume $w^T S_k \neq 0$.

$$w = \frac{1}{\gamma (w^T S_k)} (y_k - B_k S_k)$$

$$S_k^T w = \frac{1}{\gamma (S_k^T w)} S_k^T (y_k - B_k S_k).$$

$$\boxed{\gamma (S_k^T w)^2 = S_k^T (y_k - B_k S_k)}$$

$$S_k^T w = \frac{S_k^T (y_k - B_k S_k)}{\| y_k - B_k S_k \|} \Rightarrow \gamma = \frac{\| y_k - B_k S_k \|^2}{S_k^T (y_k - B_k S_k)}$$

$$w = -\frac{y_k - B_k S_k}{\| y_k - B_k S_k \|}.$$

$$C_k = \gamma w w^T = \frac{(y_k - B_k S_k)(y_k - B_k S_k)^T}{S_k^T (y_k - B_k S_k)}. \qquad B_{k+1} = B_k + C_k$$

Ex. $f(x) = \frac{1}{2} x^T Q x - c^T x$

$$Q = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix} \qquad c = \begin{bmatrix} -8 \\ -1 \\ -8 \end{bmatrix}$$

$$x^* = Q^{-1} c = \begin{bmatrix} -4 \\ -\frac{1}{3} \\ -2 \end{bmatrix} \qquad x_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad B_0 = I.$$

$$\nabla f = Qx - c \qquad \nabla f(x_0) = Q x_0 - c = \begin{bmatrix} 8 \\ 1 \\ 8 \end{bmatrix}.$$

$$p_0 = -B_0^{-1} \nabla^T f(x_0)$$

$$= -\begin{bmatrix} 8 \\ 1 \\ 8 \end{bmatrix}.$$

$$\varphi_0(\alpha) = f(x_0 + \alpha p_0) \Rightarrow \alpha = \frac{1}{3} \qquad x_1 = x_0 + \alpha_0 p_0$$

$$S_1 = x_1 - x_0 = -\frac{1}{3} \begin{bmatrix} 8 \\ 1 \\ 8 \end{bmatrix}$$

$$y_1 = \nabla^T f(x_1) - \nabla^T f(x_0) = Q x_1 - c - (Q x_0 - c) = Q (x_1 - x_0).$$