Minimization/Maximization of Functions

Numerical Methods for IT

Convex functions

Give $f: \Omega \to \mathbb{R}$; $\emptyset \neq \Omega \subset \mathbb{R}^n$. Ω is a convex open set in $\mathbb{R}^n$ if $\forall x \in \Omega, h \in \mathbb{R}^n$, then $\{t \in \mathbb{R} | x + th \in \mathbb{R}\}$ is a convex open interval in $\mathbb{R}$, with $-\infty \leq a < b \leq \infty$.

$A \in \mathbb{R}^{n \times n}$ is called positive semidefinite (or negative semidefinite), denote $A \geq 0$ $(or\ A \leq 0)$, if $h^T A\, h \geq 0, (or \leq 0), \forall h \in \mathbb{R}^n$.

Theorem. A symmetric matrix $A \geq 0$ $(or \leq 0)$ if $\lambda_i \geq 0, (or \leq 0), \forall i = 1, \dots, n$.

$f: \Omega \to \mathbb{R}^n$, with$\emptyset \neq \Omega \subset \mathbb{R}^n$is a convex open set.
(i) f is called a convex function if $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y), \forall x, y \in \Omega, 0 \leq \theta \leq 1$.
(i) f is called a concave function if $f(\theta x + (1-\theta)y) \geq \theta f(x) + (1-\theta) f(y), \forall x, y \in \Omega, 0 \leq \theta \leq 1$.

$f: \Omega \to \mathbb{R}^n$ is function of class C^2 (is a function continuous up to the second derivative). The gradient vector, $\nabla f(x)$ and Hesse matrix $\nabla^2 f(x)$ of f at x are defined by:

(i) $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right) \in \mathbb{R}^n$. (ii)$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$

Convex optimization

Theorem. $f: \Omega \to \mathbb{R}$ is function of class C^2 that defined on the convex open set $\Omega \in \mathbb{R}^n$. The following statements are equivalent.
(i) f is convex function.
(ii) $f(y) \geq f(x) + \nabla f(x)^T (y - x), \forall x, y \in \Omega.$
(iii)$\nabla^2 f(x) \geq 0, \forall x \in \Omega.$

Theorem. $f: \Omega \to \mathbb{R}$ is function of class C^2 that defined on the convex open set $\Omega \in \mathbb{R}^n$. The following statements are equivalent.
(i) f is concave function.
(ii) $f(y) \leq f(x) + \nabla f(x)^T (y - x), \forall x, y \in \Omega.$
(iii)$\nabla^2 f(x) \leq 0, \forall x \in \Omega.$

Quadratic forms

Given $q = (q_1, \dots, q_n) \in \mathbb{R}^n$ and $A = \left[a_{ij}\right] \in \mathbb{R}^{n\times n}$.
(i) Function $f: \mathbb{R}^n \to \mathbb{R}$ defined by $f_q(x) =< q, x > = \Sigma_{i=1}^{n} q_i x_i = qx^T$ is called linear form.
(ii) Function $f: \mathbb{R}^n \to \mathbb{R}$ defined by $f_A(x) = \Sigma_{i=1}^{n}\Sigma_{j=1}^{n} a_{ij} x_i x_j = xAx^T$, is called quadratic form.

Proposition. Given $f_q(x)$ is linear form, and $f_A(x)$, quadratic form. We have:
(i) $\nabla f_q(x) = q$; $\nabla^2 f_q(x) = 0, \forall x \in \mathbb{R}^n$.
(ii) $\nabla f_A(x) = 2Ax$; $\nabla^2 f_A(x) = 2A, \forall x \in \mathbb{R}^n$.

Given $A \in \mathbb{R}^{N\times n}, N > n$ with $rank(A) = n$, and given $b \in \mathbb{R}^N$.
Consider the function defined by $f(x) = \|Ax - b\|^2$. We have
$f(x) =< Ax - b, Ax - b \geq (Ax - b)(Ax - b)^T$
$= (Ax - b)(x^T A^T - b^T)$
$= Axx^T A^T - Axb^T - bx^T A^T + bb^T$
$= x(AA^T)x^T - 2(b^T A)x^T + bb^T$.
.) $\nabla f(x) = 2AA^T x^T - 2b^T A$
.) $\nabla^2 f(x) = 2AA^T$
Because $h\nabla^2 f(x)h^T = 2[hAA^T h^T] = 2[(hA)(hA)^T] = \|hA\|^2 \geq 0$.

Extreme values of one-dimensional functions

Given $f\colon [a,b] \to \mathbb{R}$ be a continuous function on $[a,b]$, and $\exists a < c < b\colon f(c) < f(a), f(b)$. Find $x^* \in [a,b]\colon f(x^*) \sim \min_{[a,b]} f(x)$.

We know that through 3 points there is a parabola.

Brent's method is approximation of the given function f by a series of parabolas. Therefore, Brent's method is also called parabolic interpolation method.

The formula for the abscissa x that is the minimum of a parabola through 3 points $(\mathrm{a}, f(a)), (b, f(b)), (c, f(c))$:

$$x = b - \frac{1}{2}\frac{(b-a)^2[f(b)-f(c)]-(b-c)^2[f(b)-f(a)]}{(b-a)[f(b)-f(c)]-(b-c)[f(b)-f(a)]}.$$

This formula fails only if three points are collinear, in which case the denominator is zero.

Brent's method is up to the task in all particulars.

Each stage, it is keeping track of 6 function points (not necessarily all distinct): a, b, u, v, w, x defined as:

. The minimum is bracketed between a and b; x is point with the very least function value found so far;

. v is the previous value of w; and u is point at which the function was evaluated most recently.

Line Methods in Multidimensions

If we start at a point **P** in N-dimension space, and proceed from there in some vector direction **n**, then any function of N variables $f(\boldsymbol{P})$ can be minimized along the line **n** by our 1-dimensional methods.
One can dream up various multidimensional minimization method that consist of sequences of such line minimizations.

Method. Given as input the vector **P** and **n**, and the function f, find scalar λ that minimizes $f(\boldsymbol{P} + \lambda\boldsymbol{n})$. Replace **P** by $\boldsymbol{P} + \lambda\boldsymbol{n}$. Replace $\boldsymbol{n}$ by $\lambda\boldsymbol{n}$.