Nonlinear Optimization

Unconstrained Optimization

Problem :- $\min\limits_{x \in \mathbb{R}^n} f(X)$

Solution set : $\{x \in \mathbb{R}^n : \nabla f(X) = 0\}$

The points from the solution set may be a point of local minima, local maxima or saddle. To determine the nature of a point in the solution set, one may evaluate the Hessian matrix (matrix of double (partial) derivatives) say $H_f(x)$

* For a point x^* in the solution set :-

  (a) x^* is a point of local minima iff $H_f(x^*)$ is positive definite (semidefinite)

  (b) x^* is a point of local maxima iff $H_f(x^*)$ is negative definite (semidefinite)

* Recall that a matrix A is positive definite (semidefinite) iff all of its principal minors have positive (non-negative) determinant.

* Equivalently a matrix A is positive definite (semidefinite) iff all the eigenvalues of A are positive (non-negative)

* A matrix A is negative definite (semidefinite) iff $-A$ is positive definite (semidefinite).

* Equivalently, one may qualitatively observe neighboring points (of x^*) to determine the nature of x^*.

Constrained Optimization

Problem :-

$$\left.\begin{aligned} &\text{Min } f(X) \\ &\text{s.t } g_i(X) \leq 0 \quad i = 1,2,\ldots m \\ &\quad h_j(X) = 0 \quad j = 1,2,\ldots k \\ &\quad X \in \mathbb{R}^n \end{aligned}\right\} \quad \text{— (A)}$$

Objective is to find $X \in \mathbb{R}^n$ which minimizes the function $f(X)$ over the region determined by the set of constraints.

If the constraints for the problem are of equality type only, one may use Lagrange's method to solve the given problem.

Lagrange's Method

Problem :-

$$\left.\begin{aligned} &\text{Min } f(X) \\ &\text{s.t. } h_j(x) = 0 \quad j = 1,2,\ldots k \\ &\quad x \in \mathbb{R}^n \end{aligned}\right\} \quad \text{— (B)}$$

To find the solution to the given problem determine

$$L(X,\lambda) = f(X) + \sum_{j=1}^{k} \lambda_j h_j(X)$$

and find the solution set $\{x \in \mathbb{R}^n : \nabla L(X,\lambda) = 0\}$

Note that the points in the solution set are solution to the simultaneous set of equations :-

$$\left.\begin{aligned} &\nabla_x f(X) + \sum_{j=1}^{k} \lambda_j \nabla_x h_j(X) = 0 \\ &h_j(X) = 0 \quad j = 1,2,\ldots k. \end{aligned}\right\} \quad \text{— (I)}$$

If (x^*, λ^*) is a point in the solution set, then x^* is a possible candidate for the optimal solutions of the given problem

Some Results :-

① If f and h_j $(j=1,2,\ldots k)$ are continuously differentiable and x^* is local minima (or maxima) for the problem (B) then there exists constants $\lambda_1^*, \lambda_2^*, \ldots \lambda_k^*$ such that the point $(x^*, \lambda_1^*, \lambda_2^*, \ldots \lambda_k^*)$ is a solution to the set of equations (I).

② If f and h_j $(j=1,2,\ldots k)$ are twice differentiable and (x^*, λ^*) be a solution to the set of equations (I). If for every non-zero vector d satisfying $\nabla h_j(x^*)^T d = 0$ $(j=1,2,\ldots k)$ we have $d^T H_f(x^*, \lambda^*) d > 0$ then x^* is a strict local minima for the problem (B).

③ Let f and h_j $(j=1,2,\ldots k)$ be twice differentiable and let (x^*, λ^*) be a solution to set of equations (I). Let

$$Q = \left[\frac{\partial^2 L(x^*, \lambda^*)}{\partial x_i \partial x_j}\right] ; \quad H = \left[\nabla h_1(x^*) \ \nabla h_2(x^*) \ldots \nabla h_k(x^*)\right]$$

and
$$P(\lambda) = \begin{vmatrix} Q - \lambda I & H \\ H^T & 0 \end{vmatrix} = 0$$

Then, x^* is a strict local minima (or maxima) for f [subject to $h_j(x) = 0 \ \forall j$ if each root of $P(\lambda)$ is positive (or negative).

Example :- Use method of Lagrange multipliers to find the solution to

$$\min\ 2x_1^2 + x_2^2 + 3x_3^2 + 10x_1 + 8x_2 + 6x_3 - 100$$

$$\text{s.t.}\quad x_1 + x_2 + x_3 = 20$$

$$x_i \geq 0 \quad \forall i.$$

$$L(X,\lambda) = 2x_1^2 + x_2^2 + 3x_3^2 + 10x_1 + 8x_2 + 6x_3 - 100 + \lambda(x_1 + x_2 + x_3 - 20)$$

The set of eqns (I) is given by

$$4x_1 + 10 + \lambda = 0$$
$$2x_2 + 8 + \lambda = 0$$
$$6x_3 + 6 + \lambda = 0$$
$$x_1 + x_2 + x_3 - 20 = 0$$

and the solution is $x_1^* = 5,\ x_2^* = 11,\ x_3^* = 4,\ \lambda = -30$

$$Q = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix} \quad ; \quad H = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$P(\lambda) = \begin{vmatrix} Q - \lambda I & H \\ H & 0 \end{vmatrix} = 0$$

or

$$\begin{vmatrix} 4-\lambda & 0 & 0 & 1 \\ 0 & 2-\lambda & 0 & 1 \\ 0 & 0 & 6-\lambda & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix} = 0$$

All roots of $P(\lambda)$ are positive and thus $(5, 11, 4)$ is the point of minima for the given problem.

For a general nonlinear optimization problem of type (A) one may use KKT conditions to find the optimal solution to the given problem.

KKT conditions

For a given NLP problem of type (A), the KKT conditions are given by the set of equations :-

$$\left.\begin{aligned} &(1)\ \nabla f(X) + \sum_{i=1}^{m} \lambda_i \nabla g_i(X) + \sum_{j=1}^{k} \mu_j \nabla h_j(X) = 0 \\ &g_i(X) \leq 0 \quad \forall\, i = 1, 2, \ldots m \\ &h_j(X) = 0 \quad \forall\, j = 1, 2, \ldots k \\ &\lambda_i \geq 0 \quad \forall\, i = 1, 2, \ldots m \\ &\lambda_i g_i(x) = 0 \quad \forall\, i = 1, 2, \ldots m \end{aligned}\right\} - (II)$$

For visualization of the above conditions, one may convert each of the constraints $g_i(x) \leq 0$ into equality type by introducing the slack variable S_i^2 and then use Lagrange's method to solve the reduced nonlinear programming problem (now with only equality constraints).

After converting $g_i(x) \leq 0$ into equality constraints the given problem looks like

$$\min f(X)$$

$$\text{s.t. } g_i(x) + S_i^2 = 0 \quad \forall\, i = 1, 2, \ldots m$$

$$h_j(X) = 0 \quad \forall\, j = 1, 2, \ldots k$$

Constructing the Lagrangian function

$$L(X,\lambda) = f(X) + \sum_{i=1}^{m} \lambda_i \left(g_i(x) + s_i^2\right) + \sum_{j=1}^{K} \mu_j h_j(X)$$

and the Lagrange's conditions are

(1) $$\nabla_x f(X) + \sum_{i=1}^{m} \lambda_i \nabla_x g_i(x) + \sum_{j=1}^{k} \mu_j \nabla_x h_j(X) = 0$$

(2) $g_i(x) \leq 0 \quad \forall i = 1,2,\ldots m$

(3) $h_j(x) = 0 \quad \forall j = 1,2,\ldots k$

(4) $\lambda_i s_i = 0 \quad \forall i = 1,2,\ldots m$

(5) $\lambda_i \geq 0 \quad \forall i = 1,2,\ldots m$

Note that the condition (4) can be replaced by the condition

$$\lambda_i g_i(x) = 0 \quad \forall i = 1,2,\ldots m$$

as $s_i = 0$ if and only if $g_i(x) = 0$

(Think why $\lambda_i \geq 0$ will hold !!!)

The variables $(\lambda_1, \lambda_2, \ldots \lambda_m, \mu_1, \mu_2, \ldots \mu_k)$ are called KKT multiplie

Do KKT multipliers always exist ??

Example :- $\min\ (x_1-4)^2 + (x_2-4)^2$

$$\text{s.t.}\quad x_1+x_2 \le 4$$
$$x_1 - x_2 \le 2$$
$$x_1, x_2 \ge 0$$

Writing NLP in eqns of form (A) we get the problem as

$$\min\ (x_1-4)^2 + (x_2-4)^2$$
$$\text{s.t}\quad x_1+x_2 \le 4 \quad \text{or} \quad x_1+x_2-4 \le 0$$
$$x_1 - x_2 \le 2 \quad \text{or} \quad x_1-x_2-2 \le 0$$
$$-x_1 \le 0$$
$$-x_2 \le 0$$

Writing KKT conditions we obtain

$$\begin{bmatrix} 2(x_1-4) \\ 2(x_2-4) \end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_4 \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left.\begin{matrix} x_1+x_2 \le 4 \\ x_1 - x_2 \le 2 \\ x_1, x_2 \ge 0 \end{matrix}\right\} \rightarrow \text{original constraints}$$

and
$$\lambda_1(x_1+x_2-4) = 0$$
$$\lambda_2(x_1-x_2-2) = 0 \qquad \text{and } \lambda_i \ge 0 \ \forall i = 1,2,3,4$$
$$\lambda_3(-x_1) = 0$$
$$\lambda_4(-x_2) = 0$$

Solving the above set of equations we obtain

$$x_1^* = 2,\ x_2^* = 2,\ \lambda_1^* = 4,\ \lambda_2^* = \lambda_3^* = \lambda_4^* = 0$$

(2,2) is a candidate for optimal solution and is actually the optimal solution to the given problem!!! (why?)

Result :- If $(X^*, \lambda_1^*, \lambda_2^*, \ldots \lambda_k^*)$ ~~are~~ is a solution to the KKT system of equations for the problem

$$\left\{\begin{array}{l} \min f(X) \\ \text{s.t. } g_i(X) \leq 0 \quad i = 1,2,\ldots k \end{array}\right\} \quad - \text{C}$$

and $f, g_1, g_2, \ldots g_k$ are convex functions then X^* is a global minima for the given problem.

Observation :- If KKT system of equations has a solution $(X^*, \lambda_1^*, \ldots \lambda_k^*)$ and problem is a convex programming problem then X^* is optimal solution (global minima) for the given problem.