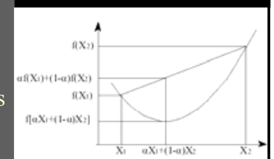
Non-linear Optimization using Calculus

Prof. D. Nagesh Kumar
Dept. of Civil Engg.
IISc, Bangalore – 560 012, India
URL: http://www.civil.iisc.ernet.in/~nagesh

Convexity

- If the objective function or the constraints are not concave or convex, the problem is usually mathematically intractable
- A function is **strictly convex** if a line connecting any two points on the function lies completely above the function
- A function is strictly convex if its slope is continually increasing or $\partial^2 f/\partial x^2 > 0$
- If > is replaced by >= then it is called convex function (but not strictly convex)

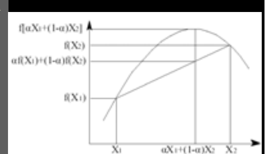


$$f[\alpha x_1 + (1-\alpha)\, x_2\,] < \alpha\, f(x_1) + (1-\alpha)\, f(x_2)$$

where $0 \le \alpha \le 1$

Concavity

- Similarly a function is **strictly concave** if a line connecting any two points on the function lies completely below the function
- A function is strictly concave if its slope is continually decreasing or $\partial^2 f/\partial x^2 < 0$
- If < is replaced by <= then it is called concave function (but not strictly concave)



$$f[\alpha x_1 + (1-\alpha)\, x_2\,] > \alpha\, f(x_1) + (1-\alpha)\, f(x_2)$$

where $0 \le \alpha \le 1$

Functions of two or more variables

- Functions of two or more variables $f(X)$, $X=[x_1, x_2, .., x_n]$ is **strictly convex** if

  $$f[\alpha \mathbf{X}_1 + (1-\alpha)\, \mathbf{X}_2\,] < \alpha\, f(\mathbf{X}_1) + (1-\alpha)\, f(\mathbf{X}_2)$$

- where $\mathbf{X}_1$ and $\mathbf{X}_2$ are points located by the coordinates given in their respective vectors
- To determine convexity or concavity of a function of multiple variables, the eigen values of it Hessian matrix should be examined
  - If all eigen values of the Hessian are negative the function is strictly concave
  - If all eigen values of the Hessian are positive the function is strictly convex
  - If some eigen values are positive and some negative, or some are zero, the function is neither strictly concave nor strictly convex

Hessian Matrix

$$H_{f(x)} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & --- & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & --- & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \begin{matrix} --- \\ --- \\ --- \end{matrix} & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & --- & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Properties of concave and convex functions

- A local minimum of a convex function is also the global minimum, and a local maximum of a concave function is also the global maximum
- A straight line is both concave and convex
- The sum of (strictly) convex functions is (strictly) convex, and the sum of concave functions is concave
- If f(X) is a convex function and k is a constant, then
  - k f(X) is convex if k > 0 and
  - k f(X) is concave if k < 0

Optimization of a Function of One Variable

- To determine the stationary points, the equation $\partial f/\partial x = 0$ should be solved
- To determine convexity or concavity, the second derivatives should be examined
  - If $\partial^2 f/ \partial x^2 > 0$ for all values of x, f(x) is convex and the stationary point is a Global Minimum
  - If $\partial^2 f/ \partial x^2 < 0$ for all values of x, f(x) is concave and the stationary point is a Global Maximum
  - If the function is neither concave nor convex, the following test may be used to classify stationary points
    - Find the first non zero higher-order derivative. Let this be the n^{th} derivative of f(x), i.e., $\partial^n f/ \partial x^n$ at $x_0 \neq 0$, where x_0 is the stationary point
    - If n is odd, x_0 is a Saddle point
    - If n is even, x_0 is a local maximum or minimum
      - If $\partial^n f/ \partial x^n$ at $x_0 < 0$, then x_0 is a local maximum
      - If $\partial^n f/ \partial x^n$ at $x_0 > 0$, then x_0 is a local minimum.

Optimization of a Function of Multiple Variables

- A necessary condition for a stationary point of the function f(X) is that each partial derivative should be equal to zero. In other words, each element of the gradient vector must equal zero where the gradient vector of f(X) is as follows.

Gradient Vector = $[\,\partial f/\partial x_1 \;\; \partial f/\partial x_2 \;\; \partial f/\partial x_3 \; .... \; \partial f/\partial x_n]^T$ where T stands for transpose

- To check the sufficient condition at X_0, Hessian matrix of f(X) at X_0 should be formulated and solved for eigen values. Then stationary point may be classified as per the following rules.
  - If all eigen values of the Hessian are negative at X_0, then X_0 is a local maximum. If all eigen values of the Hessian are negative for all possible values of X, then X_0 is a global maximum
  - If all eigen values of the Hessian are positive at X_0, then X_0 is a local minimum. If all eigen values of the Hessian are positive for all possible values of X, then X_0 is a global minimum
  - If some eigen values of the Hessian are positive and some negative or if some zero, the stationary point, X_0, is neither local minimum nor local maximum.

Example

The yield of a chemical reaction is the actual production as a percent of that which is theoretically possible. In a large commercial operation, production is found to be a function of two catalysts x_1 and x_2, where the objective is to maximize yield (%): $f(X) = 60 + 8x_1 + 2x_2 - x_1^2 - 0.5x_2^2$

$$\frac{\partial f}{\partial x_1} = 8 - 2x_1 = 0 \qquad \frac{\partial f}{\partial x_2} = 2 - x_2 = 0$$

Stationary point X =[4,2]. Compute second derivatives

$$\frac{\partial^2 f}{\partial x_1^2} = -2 \qquad \frac{\partial^2 f}{\partial x_2^2} = -1 \qquad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0$$

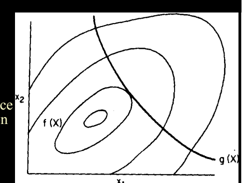
$$H = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix} \qquad |\varepsilon I - H| = \begin{vmatrix} \varepsilon + 2 & 0 \\ 0 & \varepsilon + 1 \end{vmatrix} = (\varepsilon + 2)(\varepsilon + 1) = 0$$

The values of ε do not depend on X and $\varepsilon_1 = -2$, $\varepsilon_2 = -1$.

Since both the eigen values are negative, f(X) is concave and $x_1 = 4\%$, $x_2 = 2\%$ will give the global maximum yield of f(X) = 78.0%

Optimization of a Function of Multiple Variables subject to Equality Constraints

- A function of two variables subject to a single equality constraint
- If there are only two variables, f(X) is analogous to the contour mapping of hill and g(X) to a fence on the hill. The objective is to locate the point on the fence that lies at the highest, or lowest elevation
- Note that the objective function, f(X), is not necessarily linear and the constraint equation, g(X), also may not be linear



A powerful technique for dealing with this type of optimization problem is through Lagrangian multipliers.

The Lagrangian Function is formulated as follows

$$h(X, \lambda) = f(X) - \lambda\, g(X)$$

More than one equality constraints

Maximise (or Minimise) f(X)
Subject to $g_1(X) = 0$
$g_2(X) = 0$
...
...
...
$g_m(X) = 0$

$$h(X, \lambda) = f(X) - \lambda_1 g_1(X) - \lambda_2 g_2(X) - \; .... - \lambda_m g_m(X)$$

Following sets of equations must be solved to obtain a solution.

$$\partial h(X, \lambda)/ \partial x_i = 0 \qquad \text{for } i = 1, 2, ..., n$$
$$\partial h(X, \lambda)/ \partial \lambda_p = 0 \qquad \text{for } p = 1, 2, ..., m$$

There are (n+m) unknowns and (n+m) equations.

Lagangian Approach – contd..

Check for sufficient condition

$$K_{ij} = \partial^2 h(X, \lambda)/ \partial x_i \partial x_j \;\; \text{at } X_o \qquad \text{for } i, j = 1, 2, ..., n$$

$$L_{pi} = \partial g_p(X)/ \partial x_i \qquad p=1,2,.., m, \; i=1,2,\dots,n$$

$$\begin{vmatrix} K_{11} - e & K_{12} & .... & K_{1n} & L_{11} & .... & L_{m1} \\ K_{21} & K_{22} - e & .... & K_{2n} & L_{12} & .... & L_{m2} \\ . & . & .... & . & . & .... & . \\ . & . & .... & . & . & .... & . \\ . & . & .... & . & . & .... & . \\ K_{n1} & K_{n2} & .... & K_{nn} - e & L_{1n} & .... & L_{mn} \\ L_{11} & L_{12} & .... & L_{1n} & 0 & .... & 0 \\ . & . & .... & . & . & .... & . \\ . & . & .... & . & . & .... & . \\ . & . & .... & . & . & .... & . \\ L_{m1} & L_{m2} & .... & L_{mn} & 0 & .... & 0 \end{vmatrix}$$

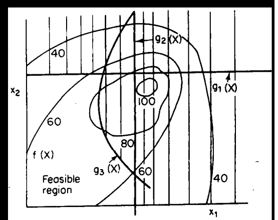
If each root of e of the equation $\Delta e = 0$, is negative, then the point X is a local maximum

If all roots are positive, then the point X is a local minimum

If some are positive and some negative, X is neither local maximum nor a local minimum

Optimization of a Function of Multiple Variables subject to Inequality Constraints

- If there are only two variables, f(X) is analogous to the contour mapping of hill and $g_i(X)$ are like several fences on the hill. The objective is to locate the point on the fences or the field enclosed by the fences that lies at the highest, or lowest elevation
- Convex Region: A region is convex, if a straight line connecting any two points in the region lies entirely in the region.



If all constraints are straight lines, the feasible region will automatically be convex

The procedures explained here are applicable only if the feasible region is convex. Otherwise search procedures are to be used

---

If f(X) is convex for minimising (concave for maximising), the problem is stated in either of the following formats.

	Max f(X)		Min f(X)
subject to	$g_1(X) \le 0$	subject to	$g_1(X) \ge 0$
	$g_2(X) \le 0$		$g_2(X) \ge 0$
	...		...
	...		...
	...		...
	$g_m(X) \le 0$		$g_m(X) \ge 0$
	(all g(X) are convex)		(all g(X) are concave)

A Lagrangian multiplier approach may be used to investigate the function:

$$h(X, \lambda) = f(X) - \lambda_1 g_1(X) - \lambda_2 g_2(X) - \ldots - \lambda_p g_p(X)$$

(The above conditions ensure a convex feasible region).

If any $\lambda_i < 0$, the constraint associated with that λ_i is not active and another solution should be obtained disregarding inactive constraints.

Example

Consider the following optimisation problem

Maximise $f(X) = 60 + 8x_1 + 2x_2 - x_1^2 - 0.5x_2^2$

Subject to $40x_1 + 20x_2 - 140 \le 0$

$50x_1 + 35x_2 - 200 \le 0$

First, establish that f (X) is a concave function

$$h(X, \lambda) = 60 + 8x_1 + 2x_2 - x_1^2 - 0.5x_2^2 - \lambda_1(40x_1 + 20x_2 - 140) - \lambda_2(50x_1 + 35x_2 - 200)$$

$$\partial h/\partial x_1 = 8 - 2x_1 - 40\lambda_1 - 50\lambda_2 = 0$$
$$\partial h/\partial x_2 = 2 - x_2 - 20\lambda_1 - 35\lambda_2 = 0$$
$$\partial h/\partial \lambda_1 = -40x_1 - 20x_2 + 140 = 0$$
$$\partial h/\partial \lambda_2 = -50x_1 - 35x_2 + 200 = 0$$

Solving these equations we find that $x_1=2.25$, $x_2=2.5$, $\lambda_1=0.369$, $\lambda_2=-0.225$.
Since λ_2 is negative, the second constraint is not active.
The second constraint should be deleted and the problem should be solved again as follows.

$$h(X, \lambda) = 60 + 8x_1 + 2x_2 - x_1^2 - 0.5x_2^2 - \lambda_1(40x_1 + 20x_2 - 140)$$

$$\partial h/\partial x_1 = 8 - 2x_1 - 40\lambda_1 = 0$$
$$\partial h/\partial x_2 = 2 - x_2 - 20\lambda_1 = 0$$
$$\partial h/\partial \lambda_1 = -40x_1 - 20x_2 + 140 = 0$$

Solving these equations we get $x_1=3$, $x_2=1$, and $\lambda_1=0.05$
This is the solution for the given problem with two constraints and the Maximised value of f(X) is 76.5.

Kuhn-Tucker Conditions

The Kuhn-Tucker Conditions are the necessary conditions for a point to be a local optimum of a function subject to inequality constraints.

This is a precise mathematical statement of the procedure used in the previous section.

If we wish to maximise f(X) subject to $g_1(X) \le 0, g_1(X) \le 0, \ldots, g_p(X) \le 0$, where $X = [x_1\ x_2\ \ldots\ x_n]$, then Kuhn-Tucker Conditions for $X^* = [x_1^*\ x_2^*\ \ldots\ x_n^*]$ to be a local optimum are

$$\frac{\partial f(X)}{\partial x_i} - \sum_{j=1}^{p} \lambda_j \frac{\partial g_j(X)}{\partial x_i} = 0 \qquad for\ i=1,2,\ldots,n\ at\ X=X^*$$

and $\lambda_j g_j(X) = 0$ $\qquad g_j(X) \le 0$

$\lambda_j \ge 0$ $\qquad for\ j=1,2,\ldots,p\ at\ X=X^*$

These are sufficient conditions for a global maximum if f(X) is concave and the constraints form a convex set