Optimization :-

An optimization problem is a mathematical problem that involves finding the best solution from a set of feasible solutions. The goal is to either maximize or minimize an objective function, which is a mathematical representation of some performance measure.

Example:

Consider the function

$f(x) = (x-3)^2$

To minimize $f(x)$,

$f'(x) = 0$

$\therefore 2(x-3) = 0$

$\therefore x = 3$.

optimal point $x^* = 3$.

Optimal value $f(x^*) = (3-3)^2 = 0$

In this example, there is no constraints on x, so x can take any real value.

Define unconstrained and constrained optimization :-

Unconstrained optimization refers to the process of finding the maximum or minimum of an objective function without any restrictions on the variables.

Suppose $f(x)$ is a function where x is a vector of variables we have to find x^* such that $f(x^*)$ is either a maximum or a minimum.

Constrained optimization involves finding the maximum or minimum of an objective function subject to a set of constraints.

Suppose $f(x)$ is a function. where x is a vector variables

$g_i(x) = 0$ for $i = 1, 2 \ldots m$

$h_j(x) \leq 0$ for $j = 1, 2 \ldots n$

we have to find x^* such that $f(x^*)$ is optimal and satisfy all constraints $g_i(x^*) = 0$ and $h_j(x^*) \leq 0$

Example:

Consider the function $f(x,y) = x^2 + y^2$ with the constraint $x+y=1$.

To solve this we can use of the method of Langrange multipliers

$$L(x,y,\lambda) = x^2+y^2+\lambda(x+y-1)$$

Taking partial derivatives and setting them to zero.

$$\frac{\partial L}{\partial x} = 2x+\lambda = 0 \quad \text{--- (i)}$$

$$\frac{\partial L}{\partial y} = 2y+\lambda = 0 \quad \text{--- (ii)}$$

$$\frac{\partial L}{\partial \lambda} = x+y-1 = 0$$

From (i), (ii) we get

$x = y$

Now, $x+y=1 \Rightarrow 2x = 1,\ x = 0.5$

$y = 0.5$

optimal point $(x^*, y^*) = (0.5, 0.5)$

optimal value $f(x^*, y^*) = (0.5)^2 + (0.5)^2$

$= 0.5$

In this example, the ~~constra~~ costraint $x+y=1$ restroie the feasible solutions.

Linear programming problem

A linear programming problem is an optimization problem where both the objective function and the constraints are linear.

For a linear function of the decision variables often written as $c^T x$ or $\sum_{i=1}^{n} c_i x_i$

The linear programming written in the form

$$Z = CX \to \max(\min)$$

$$AX \le B \text{ or } AX = B$$

$$X \geq 0$$

let, $x^* = (x_1, \dots, x_n)^T \in \mathbb{R}^n$

to maximize $Z = c_1x_1 + \dots + c_nx_n$

Subject to the constraints

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \le b_1$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \le b_m$$

and bounds $x_1 \geq 0,\ x_2 \geq 0 \dots x_n \geq 0$

Non linear Programming Problems:-

A nonlinear programming problem is a optimization problem where the objective function and one and more of the constraints are nonlinear.

For a function $f(x)$ the non linear programming written as

$g_i(x) = 0$ for equality

$h_i(x) \leq 0$ for inequality.

Convex optimization:

A convex function $f(x)$ where function f is convex if

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

for all x, y in the domain and θ in $[0,1]$.

A set of convext for inequality $h_j(x) \leq 0$ and for equality $g_i(x) = a_i^T x - b_i = 0$.

minimizer of a function

Define local and global minimizer of a function f.

The minimizer of a function is a point in the domain of the function where the function acheives its minimum value.

$f: \mathbb{R}^n \to \mathbb{R}$, a point $x^* \in \mathbb{R}^n$ is called a minimizer of f if $f(x^*) \leq f(x)$ for all $x \in \mathbb{R}^n$.

A point $x^* \in \mathbb{R}^n$ is called a local minimizer of a function $f: \mathbb{R}^n \to \mathbb{R}$ if $f(x^*)$ is less than or equal to $f(x)$ for all x in a neighbourhooh around x^*. This means x^* is the minimum point within some small region around it.

Mathematically:

x^* is a local minimizer if there exists a $\delta > 0$ such that $f(x^*) \le f(x)$ ~~for~~ for all $x \in \mathbb{R}^n$ with $\|x - x^*\| < \delta$.

Example:-

Consider the function $f(x) = (x-1)^2 \sin(x)$ where $x \in \mathbb{R}$;

The function f may have multiple local minimizers at points where it reaches a minimum value within a small interval around those points.

A point $x^* \in \mathbb{R}^n$ is called a global minimizer of a function $f: \mathbb{R}^n \to \mathbb{R}$ if $f(x^*)$ is less than or equal to $f(x)$ for all x in the domain of f. This means x^* is the minimum point over the entire domain.

Mathematically,

x^* is a global minimizer if $f(x^*) \le f(x)$ for all $x \in \mathbb{R}^n$;

example:

Consider the function $f(x) = x^2$ where $x \in \mathbb{R}$

The global minimizer is $x = 0$ because $f(0) = 0$ and $f(x) = x^2 > 0$ for any $x \neq 0$

Argmin:

A function $f: \mathbb{R}^n \to \mathbb{R}$, the argmin of f is defined as

$$\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x) = \{x^* \in \mathbb{R}^n \mid f(x^*) \le f(x) \text{ for all } x \in \mathbb{R}^n\}$$

Let $f: \mathbb{R} \to \mathbb{R}$ be a function defined by $f(x) = (x+1)^2 + 3$, find $\operatorname{argmin} f(x)$.

⇒ Solve:
The given function,

$$f(x) = (x+1)^2 + 3 \text{ and}$$

$$f: \mathbb{R} \to \mathbb{R}.$$

$$f'(x) = 2(x+1)$$

$$2(x+1) = 0$$

$$\therefore x = -1$$

$$f''(x) = 2 > 0$$

Since $f''(x) > 0$, the function confirm that $x = -1$ is a Local (and ~~gol~~ global) minimum

$$f(-1) = (-1+1)^2 + 3 = 3$$

The function value at minimum point is 3.

Since $f(x)$ reaches its minimum value of 3 at $x = -1$, the argmin is.

$$\operatorname{argmin} f(x) = \{-1\}$$

First order necessary condition

Let Ω be a subset of $\mathbb{R}^n$ and $f \in C'$ a real valued function on Ω. If x^* is a local minimizer of f over Ω, then for any feasible direction d at x^* we have $d^T \nabla f(x^*) \geq 0$.

Interior case: Let Ω be a subset of $\mathbb{R}^n$ and $f \in C'$ a real-valued function on Ω. If x^* is a local minimizer of f over Ω and if x^* is an interior point of Ω then

$$\nabla f(x^*) = 0.$$

Consider the problem

minimize $x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$

subject to $x_1, x_2 \geq 0$.

Is the FONC for a local minimizer satisfied at $x = [0, 3]^T$, Solve:

First let $f: \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x) = x_1^4 + 0.5x_2^2 + 3x_2 + 4.5$,

where $x = [x_1, x_2]^T$.

$$\nabla f(x) = [2x_2^{11}, x+3]^T = [0, 6]^T$$

$$d^T \nabla f(x) = [d_1, d_2]\ [0; 6]^T$$

$$= 6d_2 \qquad \text{where, } d = [d_1, d_2]^T$$

But,

$$x + \alpha d \in \mathbb{R}^2$$

$$[0, 3]^T + \alpha [d_1, d_2]^T \in \mathbb{R}^2$$

$$\Rightarrow [0 + \alpha d_1, 3 + \alpha d_2]^T \in \Omega.$$

$$\therefore \alpha d_1 \geq 0, \quad 3 + \alpha d_2 \geq 0.$$

For d to be feasible at x, we need $d_1 \geq 0$ and d_2 can take an arbitary value in $\mathbb{R}$. The point $x = [0, 3]^T$ does not satisfy the FONC for a minimizer because d_2 is allowed to be less than zero. For example.

$d = [1, -1]^T$ is a feasible direction but $d^T \nabla f(x) = -6 < 0$

Q. Define Jacobian, gradient and Hessian matrix of $f: \mathbb{R}^n \to \mathbb{R}$.

For a vector-valued function $f: \mathbb{R}^n \to \mathbb{R}$, where $f = [f_1, f_2, \ldots]^T$, the Jacobian matrix $J(x)$ is an $m \times n$ matrix defined as.

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \text{ then, } \frac{\partial f(x_0)}{\partial x_j} = \begin{bmatrix} \frac{\partial f_1(x_0)}{\partial x_j} \\ \vdots \\ \frac{\partial f_m(x_0)}{\partial x_j} \end{bmatrix}$$

$$\therefore J(x_0) = \begin{bmatrix} \frac{\partial f_1(x_0)}{\partial x_1} & \cdots & \frac{\partial f_1(x_0)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m(x_0)}{\partial x_1} & \cdots & \frac{\partial f_m(x_0)}{\partial x_n} \end{bmatrix}$$

If $f: \mathbb{R}^n \to \mathbb{R}$ is a differentiable, then the function ∇f defined by

$$\nabla f(x_0) = \begin{bmatrix} \frac{\partial f_1(x_0)}{\partial x_1} & \cdots & \frac{\partial f_1(x_0)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m(x_0)}{\partial x_1} & \cdots & \frac{\partial f_m(x_0)}{\partial x_n} \end{bmatrix}^T = Df(x)^T$$

is called the gradient of f.

$f: \mathbb{R}^n \to \mathbb{R}$. If ∇f is differentiable, we say that f is twice differentiable and we write the derivation of ∇f as:

$$D^2 f(x) = \begin{bmatrix} \dfrac{\partial^2 f_1(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f_1(x)}{\partial x_n \partial x_1} \\ \vdots & & \vdots \\ \dfrac{\partial f_m(x)}{\partial x_1 \partial x_n} & \cdots & \dfrac{\partial^2 f_m(x)}{\partial x_n^2} \end{bmatrix}$$

The matrix $D^2 f(x)$ is called Hessian matrix.

A given function

$$f(x) = 5x_1 + 8x_2 + x_1 x_2 - x_1^2 - 2x_2^2, \quad x = (x_1, x_2)^T \in \mathbb{R}^2$$

~~$\nabla f(x)$~~ Jacobian matrix is defined by

$$\nabla f(x) = [5 + x_2 - 2x_1, \ 8 + x_1 - 4x_2]^T.$$

Gradient matrix,

$$Df(x) = (\nabla f(x))^T = [5 + x_2 - 2x_1, \ 8 + x_1 - 4x_2]$$

Hessian matrix.

$$F(x) = D^2 f(x) = \begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} \\ \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f(x)}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}$$

Feasible direction of a vector: A vector $d \in \mathbb{R}^n, d \neq 0$ is a feasible direction at $x \in \Omega$ if there exists $\alpha_0 > 0$ such that $x + \alpha d \in \Omega$ for all $\alpha \in [0, \alpha_0]$

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a real valued function and let d be a feasible direction at $x \in \Omega$. The directional derivative of f in the direction d, denoted $\partial f / \partial d$, is the real valued function defined by

$$\frac{\partial f}{\partial d}(x) = \lim_{\alpha \to 0} \frac{f(x+\alpha d) - f(x)}{\alpha}.$$

$$\frac{\partial f(x)}{\partial d} = \frac{d}{d\alpha} f(x+\alpha d)\Big|_{\alpha=0}$$

$$= \nabla f(x)^T d$$

$$= d^T \nabla f(x)$$

* Define $f: \mathbb{R}^3 \to \mathbb{R}$ by $f(x) = x_1 x_2 x_3$ and let $d = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right]^T$

The directional derivative of f in the direction d is

$$\frac{\partial f(x)}{\partial d} = \nabla f(x)^T d = [x_2 x_3, x_1 x_3, x_1 x_2] \cdot \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \frac{x_2 x_3}{2} + \frac{x_1 x_2}{2} + \frac{x_1 x_2}{\sqrt{2}}.$$

Note that because $\|d\| = \sqrt{\frac{1}{4} + \frac{1}{4} + \frac{1}{2}} = 1$; the rate of increase of f at x in the direction d.

* What is steepest descent algorithm?

In steepest descent algorithm at each step starting from the point x^K we conduct a line search in the direction $-\nabla f(x^{(K)})$ until a minimizer, $x^{(K+1)}$ is found. Specifically, α_K is choosen to minimize $\phi_K(\alpha) \triangleq f(x^{(K)} - \alpha \nabla f(x^{(K)}))$. In other words,

$$\alpha_K = \underset{\alpha \geq 0}{\arg\min}\, f(x^{(K)} - \alpha \nabla f(x^{(K)}))$$

* Use steepest descent to find minimizer of

$$f(x_1, x_2, x_3) = (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$$

The initial point is $x^{(0)} = [4, 2, -1]^T$

Solve:

We find

$$\nabla f(x) = [4(x_1 - 4)^3, 2(x_2 - 3), 16(x_3 + 5)^3]^T$$

Hence $\nabla f(x^{(0)}) = [0, -2, 1024]$

To compute $x^{(1)}$, we need

$$\alpha_0 = \underset{\alpha \geq 0}{\arg\min}\, f(x^{(0)} - \alpha \nabla f(x^{(0)}))$$

$$= \underset{\alpha \geq 0}{\arg\min}\, f([4, 2, -1]^T - [0, -2\alpha, 1024\alpha]^T)$$

$$= \underset{\alpha \geq 0}{\arg\min}\, f([4, 2 + 2\alpha, -1 - 1024\alpha]^T)$$

$\alpha_0 = \operatorname{argmin}_{\alpha \geq 0} \left(0 + (2+2\alpha-3)^2 + 4(-1-1024\alpha+5)^4\right)$

$= \underset{\alpha \geq 0}{\operatorname{argmin}} \; \phi_0(\alpha)$

~~we know, $\alpha_0 = 3.967 \times 10^{-3}$~~

~~$x^{(1)} = x^{(0)} - \alpha_0 \nabla f(x^{(0)})$~~

$\frac{df}{d\alpha} = 0$

$\Rightarrow 2(2+2\alpha-3)\cdot 2 + 16(4-1024\alpha)^3 (-1024) = 0$

$\Rightarrow$ ~~$8 + 8\alpha - 12$~~

solve the equation we assume that

$\alpha_0 = 3.967 \times 10^{-3}$

$x^{(1)} = x^{0} - \alpha_0 \nabla f(x^{(0)})$

$= [4, 2, -1]^T - 3.967 \times 10^{-3} [0, -2, 1024]^T$

$= [4.000, 2.008, -5.062]^T$

To find $x^{(2)}$, we first determine.

$\nabla f(x^{(1)}) = [0, -1.984, -3.81 \times 10^{-3}]$.

Next we find

$$\alpha_1 = \underset{\alpha \geq 0}{\arg\min}\, f\left(x^{(1)} - \alpha \nabla f(x^{(1)})\right)$$

$$= \underset{\alpha \geq 0}{\arg\min}\left(0 + (2.008 + 1.984\alpha - 3)^2 + 4(-5.062 + 3.81\times10^{-3}\alpha + 5)^4\right)$$

$$= \underset{\alpha \geq 0}{\arg\min}\, Q_1(\alpha)$$

Again.

$$\frac{df}{d\alpha} = 0$$

$$2(2.008 + 1.984\alpha - 3)\cdot(1.984) + 16(-5.062 + 3.81\times10^{-3}\alpha + 5)^3 \cdot 3.81\times10^{-3} = 0$$

Solving the equation, $\alpha_1 = 0.5000$

Thus $x^{(2)} = x^{(1)} - \alpha_1 \nabla f(x^{(1)}) = [4, 3, -5.060]^T$

$f:\mathbb{R}^n \to \mathbb{R}$ then for each k the vector $x^{(k+1)} - x^{(k)}$ is orthogonal to the vector $x^{k+2} - x^{(k+1)}$.

From the method of steepest descent it follows that

$$\left(x^{(k+1)} - x^{(k)}, x^{(k+2)} - x^{(k+1)} \right) = \alpha_k \alpha_{k+1} \left(\nabla f(x^{(k)}), \nabla f(x^{(k+1)}) \right)$$

To complete the proof it is enough to show that.

$$\left(\nabla f(x^{(k)}), \nabla f(x^{(k+1)}) \right) = 0$$

we know that

$$\phi_k(\alpha) \triangleq f(x^{(k)} - \alpha \nabla f(x^{(k)}))$$

Hence using the FONC and the chain rule.

$$0 = \phi_k'(\alpha_k)$$

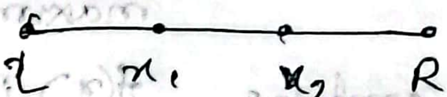
$$= \frac{d\phi_k}{d\alpha}(\alpha_k)$$

$$= \nabla f(x^{(k)} - \alpha_k \nabla f(x^{(k)}))^T (-\nabla f(x^{(k)}))$$

$$= -\left(\nabla f(x^{(k+1)}), \nabla f(x^{(k)}) \right)$$

and the proof is complete

Golden Search:



Consider an initial search be [L, R] then:

$$\frac{R-L}{x_2-L} = \frac{x_2-L}{x_2-x_1} = 1.618$$

$$\Rightarrow \frac{R-L}{x_2-L} = 1.618$$

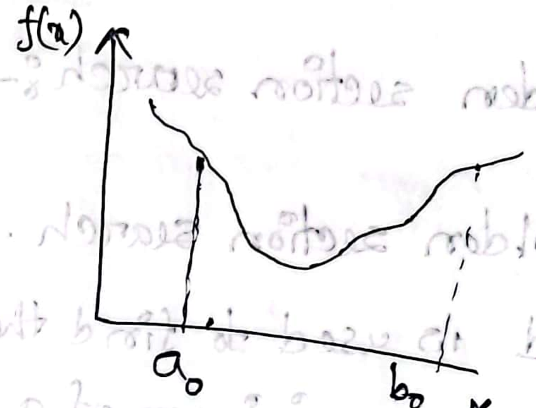
$$\Rightarrow \boxed{x_2 = L + 0.618(R-L)}$$

$$x_1 + x_2 = L + R$$

$$\Rightarrow \boxed{x_1 = L + R - x_2}$$

If $f(x_1) < f(x_2)$ then update R with x_2. Preserve L.

If $f(x_2) \le f(x_1)$ then preserve R and update L with x_1



A: unimodal function

⊛ Define unimodal :—

A unimodal function contains only one minimum or maximum on the interval $[a, b]$.

⊛ Golden section search :—

The Golden section search method is used to find the maximum or minimum of a unimodal function.

$$p(b_1 - a_0) = b_1 - b_2$$

Because $b_1 - a_0 = 1 - p$ and

$$b_1 - b_2 = 1 - 2p$$

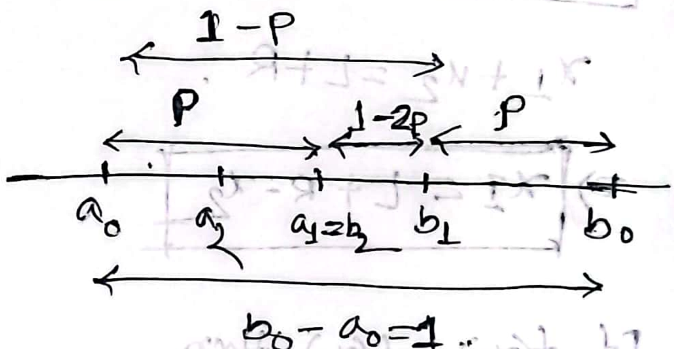
we have

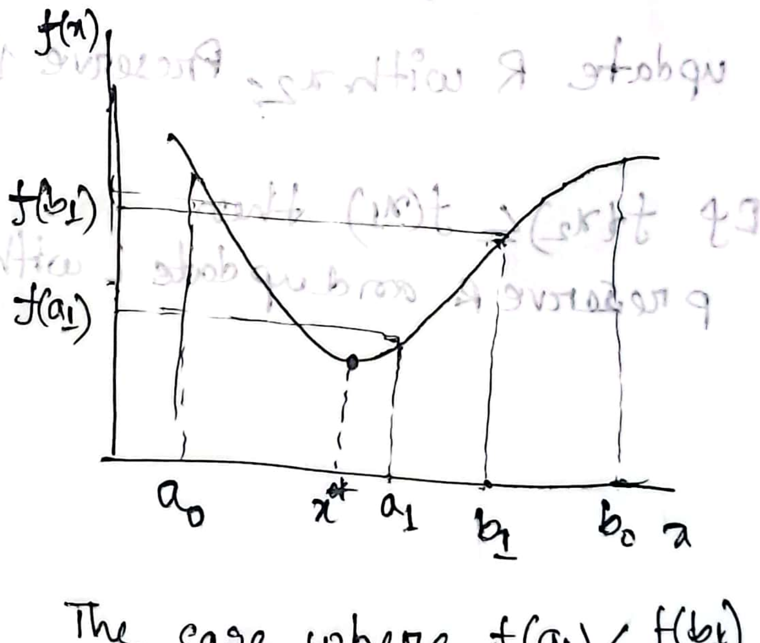
$$p(1-p) = 1 - 2p$$

$$\Rightarrow p^2 - 3p + 1 = 0$$

$$\therefore p_1 = \frac{3+\sqrt{5}}{2}, \quad p_2 = \frac{3-\sqrt{5}}{2}.$$

- If f is unimodal with a maximum at $x^* \in [a, b]$, then $f(x)$ increase for $x \in [a, x^*]$ and decrease for $x \in [x^*, b]$
- If f is unimodal with a minimum at $x^* \in [a, b]$ then $f(x)$ decrease for $x \in [a, x^*]$ and increase for $x \in [x^*, b]$





The case where $f(a_1) < f(b_1)$ the minimizer $x^* \in [a_0, b_1]$

because we require $p < \frac{1}{2}$ we take.

$$p = \frac{3-\sqrt{5}}{2} \approx 0.382.$$

$$\therefore 1-p = \frac{\sqrt{5}-1}{2}.$$

$$\text{and } \frac{p}{1-p} = \frac{3-\sqrt{5}}{\sqrt{5}-1} = \frac{\sqrt{5}-1}{2} = \frac{1-p}{1}$$

$$\therefore \frac{p}{1-p} = \frac{1-p}{1}$$

This method is called the Golden section search method.

Hence N steps of reduction using the Golden Section method reduces the range by the factor

$$(1-p)^N \approx (0.618)^N.$$

(4) Use Golden section search to find the value of x that minimizes.

$$f(x) = x^4 - 14x^3 + 60x^2 - 70x.$$

in the range $[0,2]$. Locate this value of x to within a range of 0.3.

Solve:

After N stages the range reduced by $(0.618)^N$. So we choose N so that.

$$(0.618)^N \leq \frac{0.3}{2-0}$$

$$\Rightarrow N = 4.$$

Four stages of reduction will do that is $N=4$.

Iteration 1:

$$a_1 = a_0 + p(b_0 - a_0)$$

$$= 0 + 0.382 \times 2$$

$$= 0.764.$$

$$b_1 = b_0 + a_0 - a_1$$

$$= 2 - 0.764$$

$$= 1.236.$$

where.

$$p = \frac{3-\sqrt{5}}{2} \approx 0.382.$$

$$f(a_1) = -24.36$$

$$f(b_1) = -18.96.$$

$$f(b_1) > f(a_1)$$

$$\therefore [a_0, b_1] = [0, 1.236]$$

Iteration 2:

$a_2 = a_0 + P(b_1 - a_0)$

$= 0.382 \times 1.236$

$= 0.472.$

$b_2 = a_0 + b_1 - a_2$

$= 1.236 - 0.47$

$= 0.766.$

$f(a_2) = -21.05.$

$f(b_2) = -24.36.$

$f(a_2) > f(b_2).$

$\therefore [a_2, b_1] = [\cancel{-21.05}\ 0.472, 1.236]$

Iteration 3:

$a_3 = a_2 + P(b_1 - a_2)$

$= 0.472 + 0.382 \times 0.764$

$= 0.764.$

$b_3 = a_2 + b_1 - a_3$

$= 0.944.$

$f(a_3) = -24.36$

$f(b_3) = -23.59.$

$\therefore f(b_3) > f(a_3).$

$[a_2, b_3] = [0.472, 0.944]$

Iteration 4:

$a_4 = a_2 + P(b_3 - a_2)$

$= 0.472 + 0.382 \times 0.472$

$= 0.652.$

$b_4 = a_2 + b_3 - a_4.$

$= 1.416 - 0.652$

$= 0.764.$

$f(a_4) = -23.84$

$f(b_4) = -24.36.$

$f(a_4) > f(b_4)$

$\therefore$ the value of x that minimizes f is located in the interval.

$[a_4, b_3] = [0.652, 0.944]$

Newton's Method:

$$x^{k+1} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

Using Newton's Method, find the minimizer of $f(x) = \frac{1}{2}x^2 - \sin x$.

The initial value is $x^{(0)} = 0.5$, the required accuracy is $\epsilon = 10^{-5}$ in the sense that we stop when $|x^{(k+1)} - x^{(k)}| < \epsilon$.

Solve:

$$f(x) = \frac{1}{2}x^2 - \sin x.$$

$$f'(x) = x - \cos x.$$

$$f''(x) = 1 + \sin x.$$

Initially, $x = 0.5$

Hence.

$$x^1 = 0.5 - \frac{0.5 - \cos 0.5}{1 + \sin 0.5}$$

$$= 0.5 - \frac{-0.3775}{1.479}$$

$$= 0.7552.$$

Proceeding a similar manner we obtain.

$$x^{(2)} = x^{(1)} - \frac{0.7552 - \cos 0.7552}{1 + \sin 0.7552}$$

$$= 0.7552 - \frac{0.027}{1.685}$$

$$= 0.7391.$$

$$x^{(3)} = 0.7391 - \frac{0.7391 - \cos 0.7391}{1 + \sin 0.7391}$$

$$= 0.7390.$$

$$x^{(4)} = 0.7390 - \frac{0.7390 - \cos 0.7390}{1 + \sin 0.7390}$$

$$= 0.7390$$

Here

$$|x^{(4)} - x^{(3)}| < \epsilon = 10^{-5}$$

furthermore,

$$f'(x^{(4)}) = -0.00014 \approx 0$$

observed that $f''(x^{(4)}) = 1.673 > 0$

so we can assume that $x^* \approx x^{(4)}$ is a strict minimizer.

Q. Define graph, epigraph and convex function:—

The graph of $f : \Omega \to \mathbb{R}, \Omega \subset \mathbb{R}^n$ is the set of points in $\Omega \times \mathbb{R} \subset \mathbb{R}^{n+1}$ given by.

$$\{ [x, f(x)]^T : x \in \Omega \}.$$

The epigraph of a function. $f : \Omega \to \mathbb{R}, \Omega \subset \mathbb{R}^n$, denoted epi$(f)$ is the set of points in. $\Omega \times \mathbb{R}$ given by.

$$\text{epi}(f) = \{ [x, \beta]^T : x \in \Omega, \beta \in \mathbb{R}, \beta \geq f(x) \}.$$

A set $\Omega \subset \mathbb{R}^n$ is convex if for every $x_1, x_2 \in \Omega$ and $\alpha \in (0,1)$ $\alpha x_1 + (1-\alpha) x_2 \in \Omega$..

A function $f : \Omega \to \mathbb{R}, \Omega \subset \mathbb{R}^n$ is convex on Ω if its epigraph is a convex set.

(4) If a function $f: \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$ is convex on Ω, then Ω is a convex set.

Suppose Ω is not a convex set. Then there exists two points y_1 and y_2 such that for some $\alpha \in (0,1)$

$$z = \alpha y_1 + (1-\alpha) y_2 \notin \Omega .$$

let,

$$\beta_1 = f(y_1), \ \beta_2 = f(y_2) .$$

then the pairs.

$$\begin{bmatrix} y_1 \\ \beta_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ \beta_2 \end{bmatrix}$$

$$w = \alpha \begin{bmatrix} y_1 \\ \beta_1 \end{bmatrix} + (1-\alpha) \begin{bmatrix} y_2 \\ \beta_2 \end{bmatrix}$$

$$\Rightarrow w = \begin{bmatrix} \alpha y_1 \\ \alpha \beta_1 \end{bmatrix} + \begin{bmatrix} (1-\alpha) y_2 \\ (1-\alpha) \beta_2 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha y_1 + (1-\alpha) y_2 \\ \alpha \beta_1 + (1-\alpha) \beta_2 \end{bmatrix}$$

$$\therefore w = \begin{bmatrix} z \\ \alpha \beta_1 + (1-\alpha) \beta_2 \end{bmatrix}$$

$w \notin epi(f)$, because $z \notin \Omega$. Therefore $epi(f)$ is not convex and hence f is not a convex function.

A function $f: \Omega \to \mathbb{R}$ defined on a convex set $\Omega \subset \mathbb{R}^n$ is convex if and only if for all $x, y \in \Omega$ and all $\alpha \in (0,1)$, we have.

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Proof:

Assume that for all $x, y \in \Omega$ and $\alpha \in (0,1)$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y).$$

let $[x^T, a]^T$ and $[y^T, b]^T$ be two points in epi(f) where $a, b \in \mathbb{R}$. From the definition of epi(f) it follows that.

$$f(x) \leq a, \quad f(y) \leq b.$$

Therefore, using the first inequality above, we have:

$$f(\alpha x + (1-\alpha) y) \leq \alpha a + (1-\alpha) b$$

Because Ω is convex, $\alpha x + (1-\alpha) y \in \Omega$ Hence.

$$\begin{bmatrix} \alpha x + (1-\alpha) y \\ \alpha a + (1-\alpha) b \end{bmatrix} \in \text{epi}(f).$$

which implies that epi(f) is a convex set, hence. f is a convex funtion.

~~Assu~~

Assume that $f:\Omega \to \mathbb{R}$ is a convex function. Let $x, y \in \Omega$ and

$$f(x) = a, \quad f(y) = b.$$

Thus $\begin{bmatrix} x \\ a \end{bmatrix}, \begin{bmatrix} y \\ b \end{bmatrix} \in \text{epi}(f)$.

Because f is a convex function, its epigraph is a convex subset of $\mathbb{R}^{n+1}$. Therefore, for all $\alpha \in (0,1)$, we have

$$\alpha \begin{bmatrix} x \\ a \end{bmatrix} + (1-\alpha) \begin{bmatrix} y \\ b \end{bmatrix} = \begin{bmatrix} \alpha x + (1-\alpha) y \\ \alpha a + (1-\alpha) b \end{bmatrix} \in \text{epi}(f).$$

The above implies that for all $\alpha \in (0,1)$

$$f(\alpha x + (1-\alpha) y) \leq \alpha a + (1-\alpha) b = \alpha f(x) + (1-\alpha) f(y).$$

[proved]

Let $f(x) = x_1 x_2$. Is f convex over $\Omega = \{x : x_1 \geq 0, x_2 \geq 0\}$?

⇒ The answer is no. Take for example, $x = [1, 2]^T \in \Omega$ and $y = [2, 1]^T \in \Omega$. Then.

Then,

$$\alpha x + (1-\alpha) y = \alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix} + (1-\alpha) \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha + 2(1-\alpha) \\ 2\alpha + (1-\alpha) \end{bmatrix}$$

$$= \begin{bmatrix} 2-\alpha \\ 1+\alpha \end{bmatrix}$$

Hence,

$$f(\alpha x + (1-\alpha) y) = (2-\alpha)(1+\alpha)$$

$$= 2 + \alpha - \alpha^2$$

$\therefore \alpha f(x) = \alpha \cdot 2 \cdot 1 = 2\alpha$.

$(1-\alpha) f(y) = (1-\alpha) \cdot 2 \cdot 1 = 2(1-\alpha)$

$\therefore \alpha f(x) + (1-\alpha) f(y) = 2$.

if, for example $\alpha = \frac{1}{2} \in (0, 1)$ then

$$f\left(\tfrac{1}{2}x + \tfrac{1}{2}y\right) = \frac{9}{4} > \frac{1}{2} f(x) + \frac{1}{2} f(y)$$

which shows that f is not convex over Ω.

(b) A quadratic form $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, given by $f(x) = x^T Q x$, $Q \in \mathbb{R}^{n\times n}$, $Q = Q^T$, is convex on Ω if and only if for all $x, y \in \Omega$, $(x-y)^T Q (x-y) \geq 0$.

Solve:

The function $f(x) = x^T Q x$ is convex if and only if for every $\alpha \in (0,1)$ and every $x, y \in \mathbb{R}^n$ we have.

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y).$$

or $f(\alpha x + (1-\alpha)y) - \alpha f(x) - (1-\alpha) f(y) \geq 0$

$\Rightarrow \alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y) \geq 0$

$\Rightarrow$

$$\alpha x^T Q x + (1-\alpha) y^T Q y - (\alpha x + (1-\alpha) y)^T Q (\alpha x + (1-\alpha) y)$$

$$= \alpha x^T Q x + y^T Q y - \alpha y^T Q y - \alpha^2 x^T Q x - (2\alpha - 2\alpha^2) x^T Q y - (1 - 2\alpha + \alpha^2) y^T Q y$$

$$= \alpha(1-\alpha) x^T Q x - 2\alpha(1-\alpha) x^T Q y + \alpha(1-\alpha) y^T Q y$$

$$= \alpha(1-\alpha)(x-y)^T Q (x-y) \geq 0.$$

Therefore f is convex if and only if

$$\alpha(1-\alpha)(x-y)^T Q (x-y) \geq 0$$

⑨ Define Lagrangian function:—

Let x^* be a local minimizer of $f: \mathbb{R}^n \to \mathbb{R}$ subject to $h(x) = 0$, $h: \mathbb{R}^n \to \mathbb{R}^m$, $m \le n$. Assume that x^* is a regular point. Then, there exists $\lambda^* \in \mathbb{R}^m$ such that.

$$Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T.$$

Lagrangian conditions

The Lagrangian condition refers to the set of necessary condition that an optimal solution must satisfy in a constrained optimization problem. These condition are derived from the Lagrangian function and are commonly known as the Karush-Kuhn-Tucker (KKT).

Karush-Kuhn-Tucker conditions

Let $f, h, g \in C'$, let x^* be a regular point and a local minimizer for the problem of minimizing f subject to $h(x) = 0$, $g(x) \le 0$ then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that.

1. $\mu^* \geq 0$
2. $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$
3. $\mu^{*T} g(x^*) = 0$.

Q. Consider the problem of extremizing the objective function $f(x) = x_1^2 + x_2^2$ on the ellipse $\{[x_1, x_2]^T : h(x) = x_1^2 + 2x_2^2 - 1 = 0,\}$ where h is equality constraints. Find the extremizers of $f(x)$ using the idea of Lagrange condition.

$$\nabla f(x) = [2x_1, 2x_2]^T$$

$$\nabla h(x) = [2x_1, 4x_2]^T.$$

Thus,

$$D_x l(x,\lambda) = D_x[f(x) + \lambda h(x)] = [2x_1 + 2\lambda x_1, 2x_2 + 4\lambda x_2]$$

and

$$D_\lambda l(x,y) = h(x) = x_1^2 + 2x_2^2 - 1.$$

Setting $D_x l(x,\lambda) = 0^T$ and $D_\lambda l(x,\lambda) = 0$ we obtain three equations in three unknowns.

$$2x_1 + 2\lambda x_1 = 0 \quad \text{— (i)}$$

$$2x_2 + 4\lambda x_2 = 0 \quad \text{— (ii)}$$

$$x_1^2 + 2x_2^2 = 1 \quad \text{— (iii)}$$

From (i) we get either $x_1 = 0$ or $\lambda = -1$.

if $x_1 = 0$ (ii) and (iii) equation emplies that $\lambda = -\frac{1}{2}$ and $x_2 = \pm\frac{1}{\sqrt{2}}$.

if $\lambda = -1$ then from (ii) and (iii) we get $x_2 = 0$, $x_1 = \pm 1$

Thus, the points that satisfy the Lagrange condition for extreme are:

$$x^{(1)} = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad x^{(2)} = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad x^{(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad x^{4} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Because $f(x^1) = f(x^2) = \frac{1}{2}$

and $f(x^3) = f(x^4) = 1$.

we conclude that if there are minimizers, then they are located at $x^{(1)}$ and $x^{(2)}$ and if there are maximizers, then they are located at $x^{(3)}$ and $x^{(4)}$.

④ Prove, Lagrange condition for a local minimizer x^* can be represented using the Lagrangian function as $Dl(x^*, \lambda^*) = 0^T$ for some λ^*, where derivative operator D is with respect to the entire argument $(x^T, \lambda^T)^T$.

<u>Solve:</u>

A function $L: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be defined by.

$L(x, \lambda) = f(x) + \lambda^T h(x)$ is called a Lagrange function

$$D_x l(x, \lambda) = Df(x) + \lambda^T Dh(x).$$

$$D_\lambda L(x, \lambda) = h(x)^T$$

Denote the derivative of l with respect to x as $D_x L$ and the derivative of l with respect to λ as $D_\lambda l$ the.

$$Dl(x, \lambda) = [D_x l(x, \lambda), D_\lambda l(x, \lambda)]$$

Therefore the Lagrange's theorem for a local minimizer x^* can be stated as

$$D_x l(x^*, \lambda^*) = 0^T.$$

$$D_\lambda L(x^*, \lambda^*) = 0^T$$ for some λ^*, which is

equivalent to.

$$Dl(x^*, \lambda^*) = 0^T.$$