# Constrained Optimization and Lagrange Multipliers

## Tutorial

Prof. Arun Rajkumar
Computer Science and Engineering
Indian Institute of Technology, Madras

Mr. Kumar Ramanand
Tutorial Instructor

IIT Madras
BSc Degree

# Table of Contents



IIT Madras
BSc Degree

# Gradient Descent

## Update Rule

To minimise the function f(x):

- start with arbitrary choice, $x_0 \in \mathbb{R}$
- for $t = 1, 2, 3, .., T$, Update $\boxed{x_{t+1} = x_t + \eta_t d}$

  where,

  - direction of gradient descent, $d = -f'(x_t)$
  - learning rate at t-th step (at point $x_t$), $\eta_t$

## Properties

- Gradient descent algorithm converges when $\eta_t$ is a diverging series (sum of the series tends to infinity)
- Gradient descent algorithm converges to local minimum.
- For convex functions local minimum is also global minimum. Gradient descent algorithm is good fit for convex function to find global minimum.

IIT Madras
BSc Degree

# Gradient Descent

- Taylor series, $f(x+\eta d) = f(x) + \eta d f'(x) + \frac{1}{2}\eta^2 d^2 f''(x) + \ldots$ where $x, d, d^2, f(x), f'(x), f''(x)$ are scalar values
- Local information gives global information. The value at any point can be determined if all the lower and higher order derivatives are known at the point x
- For small positive constant, $\eta$, function evaluation at updated point along the direction $d$,

$$f(x+\eta d) \sim f(x) + \eta d f'(x)$$

- Descent direction is the direction of movement, $d$, such that $df'(x) < 0$, (For $d = -f'(x)$ this condition is satisfied)

IIT Madras
BSc Degree

# Gradient Descent in higher dimension

- Taylor series, $f(x+\eta d) = f(x) + \eta d^T \nabla f(x) + \frac{1}{2}\eta^2 (d^2)^T \nabla^2 f(x) + ...$ where $x, d, d^2, \nabla f(x), \nabla^2 f(x)$ are vectors and $f(x), d^T \nabla f(x), (d^2)^T \nabla^2 f(x)$ are scalar values
- Descent direction is the direction of movement, $d$, such that $d^T \nabla f(x) < 0$, (For $d = -\nabla f(x)$ this condition is satisfied)
- Location vector at (t+1) step along the descent direction, d

$$x_{t+1} = x_t + \eta d^T \nabla f(x_t), d = -\nabla f(x_t)$$

- For small positive constant, $\eta$, function evaluation near a point $x$ along the descent direction, $d$

$$f(x+\eta d) \sim f(x) + \eta d^T \nabla f(x), d = -\nabla f(x)$$

- Note: $d = -\nabla f(x)$ gives steepest descent. Gradient descent is also called steepest descent algorithm.

IIT Madras
BSc Degree

# Constrained optimization

- Minimize $f(x),\ x \in \mathbb{R}^d$ such that $g(x) \leq 0$,
  - (1) Descent direction: A direction that reduces the objective function $f(x)$ value, $d^T \nabla f(x) < 0$
  - (2) Feasible direction: A direction that takes to a point satisfying the constraint function $g(x) \leq 0$
- If $x^*$ is an optimal point, then
  - (1) $g(x^*) \leq 0$
  - (2) No descent direction is a feasible direction

IIT Madras
BSc Degree

- Minimize objective function $f(x),\ x \in \mathbb{R}^d$, under inequality constraint, $g(x) \leq 0$,
- Necessary condition for optimality at a point $x^*$:

$$d^T \nabla g(x^*) = 0$$

- Optimal configuration:

$$\nabla f(x^*) = -\lambda \nabla g(x^*)$$

where $\lambda$ is a positive scalar value

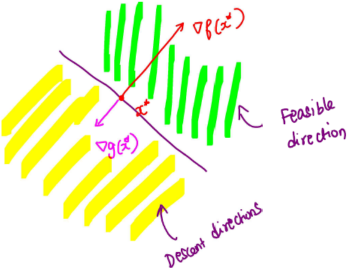



Figure: Feasible and descent direction: anti-parallel to each other

IIT Madras
BSc Degree

# Method of Lagrange multiplier

- Minimize objective function, $f(x),\ x \in \mathbb{R}^d$, under equality constraint, $g(x) = 0$,
- Necessary condition for optimality at a point $x^*$:

$$d^T \nabla g(x^*) = 0$$

- Optimal configuration:

$$\nabla f(x^*) = \lambda \nabla g(x^*)$$

where $\lambda$ is any arbitrary scalar value (positive or negative) known as Lagrange multiplier

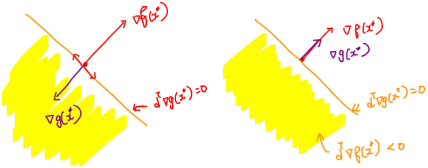



Figure: Feasible and descent direction: parallel or anti-parallel to each other

IIT Madras
BSc Degree

# Convexity

- A set $S \subseteq \mathbb{R}^d$, is a convex set if
  $\forall x_1, x_2 \in S$, the point $\lambda x_1 + (1-\lambda)x_2 \in S, \forall \lambda \in [0,1]$

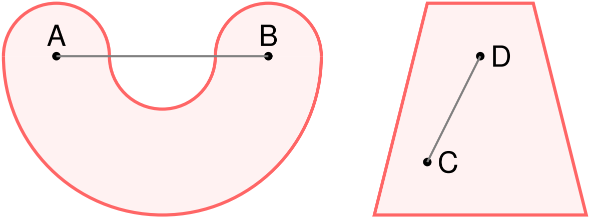



Figure: Left: Non-convex set, Right: Convex set

- Property of a convex set:
  - Intersection of convex sets is convex.
    If $S_1, S_2 \subseteq \mathbb{R}^d$, then $S_{12} = S_1 \cap S_2 = \{x : x \in S_1, x \in S_2\}$

IIT Madras
BSc Degree

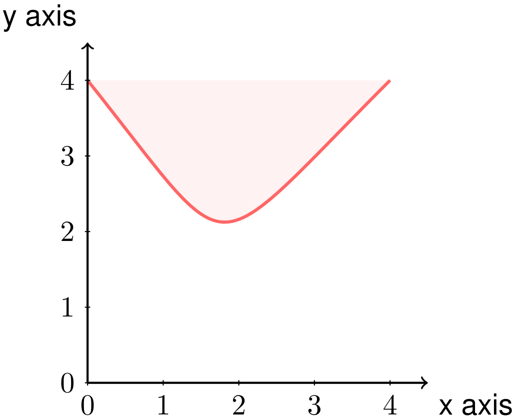
# Convex Hull

- Convex combination: Let $S = \{x_1, x_2, ..., x_n\} \subseteq \mathbb{R}^d$ and $z = \lambda_1 x_1 + \lambda_2 x_2 + ..., \lambda_n x_n$, then z is said to be convex combination in S if there exists $\lambda_1, \lambda_2, ..., \lambda_n$ such that $\lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i = 1$
- Convex hull,
$$CH(S) = \{z : z = \sum_{i=1}^{n} \lambda_i x_i, \ni x_i \in S, \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i = 1\}$$
- A convex hull is a convex set
- Alternate definition of convex hull: A set $\{x_1, x_2, ..., x_n\}$ as the intersection of all convex sets that contains $\{x_1, x_2, ..., x_n\}$

IIT Madras
BSc Degree

# Convex functions

- Given a function, $f : \mathbb{R}^d \to \mathbb{R}$ whose domain is a convex set Epigraph of the function, $\texttt{epi}(f) = \{ \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^d : z \geq f(x) \}$
- A function is a convex function if $epi(f)$ is a convex set



# Convex functions - alternate definition

Alternative definitions:

- A function is said to be convex, if $f(\lambda x_1 + (1-\lambda)x_2 \leq \lambda f(x_1) + (1-\lambda)f(x_2)$ hold for all $\lambda \in [0,1]$ where $x_1, x_2 \in \mathbb{R}^d$
- A differentiable function, $f : \mathbb{R}^d \to \mathbb{R}$ is convex if and only if, for a point y in the neighbourhood of x:
  $f(y) \geq f(x) + (y-x)^T \nabla f(x)$
  In other words, the linear approximation of the function at point $y$ in the neighbourhood of some point $x$ on the function lower bounds the function

IIT Madras
BSc Degree

# Convex functions - alternate definition

Alternative definition:

- A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is is convex if
  - the Hessian matrix $H$ is positive definite or positive semi-definite matrix, $det(H) \geq 0$
  - eigenvalues of the Hessian matrix are non-negative, $\texttt{eigenvalues}(H) \geq 0$
- For a convex function, all local minima are also global minima, if $x^*$ is a local minima and $z$ is a global minima for the convex function $f$:
  $f(z) = f(x^*) \leq f(x), \forall\, x \in [x^* - \delta, x^* + \delta], \delta > 0$

IIT Madras
BSc Degree

# Example 1

A rectangular box without a lid is to be made from $12m^2$ of cardboard. Find the maximum volume of such a box

## Solution:

Let x, y, and z be the length, width, and height respectively, of the box in meters.

We wish to maximize volume, $f(x, y, z) = xyz$ subject to the constraint, $g(x, y, z) = 2xz + 2yz + xy = 12$

- Using the method of Lagrange multipliers, we look for values of x, y, z, and $\lambda$ such that $\nabla f = \lambda \nabla g$

$$\implies \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} = \lambda \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} \implies \begin{bmatrix} yz \\ xz \\ xy \end{bmatrix} = \lambda \begin{bmatrix} (2z + y) \\ (2z + x) \\ (2x + 2y) \end{bmatrix}$$

and $2xz + 2yz + xy = 12$

BSc Degree

# Example 1

## Solution:

- Multiplying 1st equation by x, 2nd equation by y, and 3rd equation by z. The left sides of these equations will be identical: $xyz = \lambda(2xz + xy) = \lambda(2yz + xy) = \lambda(2xz + 2yz)$
- Since, cuboid dimension should greater than 0 for volume to be non-zero and real: $x > 0, y > 0, z > 0 \implies x = y = 2z$
- Solving constraint equation: $2xz + 2yz + xy = 12 \implies (x, y, z) = (2, 2, 1)$
- The volume of the cube, $f(x, y, z) = xyz = 2 * 2 * 1 = 4$

IIT Madras
BSc Degree

# Example 2

Find the maximum value of the function $f(x, y, z) = x + 2y + 3z$ on the curve of intersection of the plane $x–y + z = 1$ and the cylinder $x^2 + y^2 = 1$.

## Solution:

We maximize the function $f(x, y, z) = x + 2y + 3z$ subject to the constraints $g(x, y, z) = x–y + z = 1$ and $h(x, y, z) = x^2 + y^2 = 1$.

- Using the method of Lagrange multipliers, we look for values of x, y, z, $\lambda$ and $\mu$ such that $\nabla f = \lambda \nabla g + \mu \nabla h$

$$\implies \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} = \lambda \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} + \mu \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} \implies \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + \mu \begin{bmatrix} 2x \\ 2y \\ 0 \end{bmatrix}$$

and $x–y + z = 1$, $x^2 + y^2 = 1$

IIT Madras
BSc Degree

# Example 2

## Solution:

- $1 = \lambda + 2\mu x, 2 = -\lambda + 2\mu y, \lambda = 3 \implies x = -\frac{1}{\mu}, y = \frac{5}{2\mu}$
  and $x - y + z = 1, x^2 + y^2 = 1$
- $x^2 + y^2 = \frac{1}{\mu^2} + \frac{25}{4\mu^2} = 1$
  $\implies \mu = \pm\frac{\sqrt{29}}{2} \implies x = \mp\frac{2}{\sqrt{29}}, y = \mp\frac{5}{\sqrt{29}}, z = 1 \pm \frac{7}{\sqrt{29}}$
- Objective function at extreme points,
  $f(x, y, z) = x + 2y + 3z = \mp\frac{2}{\sqrt{29}} + 2 * (\mp\frac{5}{\sqrt{29}}) + 3 * (1 \pm \frac{7}{\sqrt{29}})$
  Maximum value of f is $3 \pm \sqrt{29}$

# Example 3

Check convexity of the function, $f(x,y) = 4x^2 + 2y^2$

First, second order partial derivatives and the Hessian:

- First order derivatives: $f_x = 8x$, $f_y = 4y$
- Second order partial derivatives, $f_{xx} = 8$, $f_{xy} = 0$, $f_{yy} = 4$
- Hessian matrix, $H = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 0 & 4 \end{bmatrix}$

## Hessian determinant test:

- Determinant of the Hessian matrix, $D = f_{xx}f_{yy} - {f_{xy}}^2 = 8 * 4 - 0^2 = 32$
- $f_{xx} > 0$ and $D > 0$,
- Therefore, hessian is positive definite and $f(x,y)$ is a convex function.

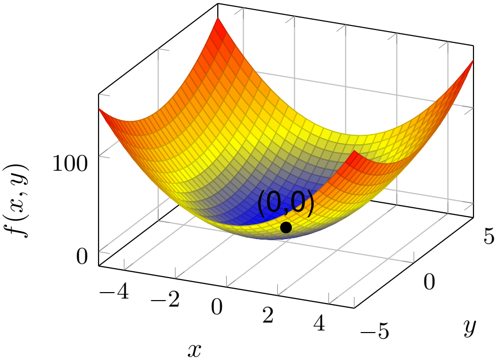
BSc Degree

# Example 3

## Eigenvalue test:

- Hessian matrix is a diagonal matrix. Its diagonal elements are the eigenvalues.
- Both eigenvalues $8, 4$ are positive.
- Therefore, hessian is positive definite and $f(x, y)$ is a convex function.

3D plot: $f(x, y) = 4x^2 + 2y^2$



# Example 4

Check convexity of the function, $f(x, y) = x^2 - y^2$

First, second order partial derivatives and the Hessian:

- First order derivatives: $f_x = 2x$, $f_y = -2y$
- Second order partial derivatives, $f_{xx} = 2$, $f_{xy} = 0$, $f_{yy} = -2$
- Hessian matrix, $H = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$

## Hessian determinant test:

- Determinant of the Hessian matrix, $D = f_{xx} f_{yy} - {f_{xy}}^2 = 2 * -2 - 0^2 = -4$
- $f_{xx} > 0$ and $D < 0$,
- Therefore, hessian matrix is indefinite and $f(x, y)$ is a non-convex function.

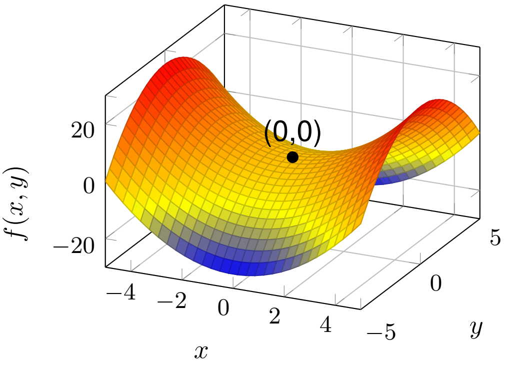
BSc Degree

# Example 4

## Eigenvalue test:

- Hessian matrix is a diagonal matrix. Its diagonal elements are the eigenvalues.
- It has one positive and one negative eigenvalues (2, -2).
- Therefore, hessian matrix is indefinite and $f(x, y)$ is a non-convex function.

3D plot: $f(x, y) = x^2 - y^2$




IIT Madras
BSc Degree

Thank you

IIT Madras
BSc Degree