Week (4) Notes

Dr. Amr Amin

Ahmed Al-Qassas

Optimization

Optimization focuses on **finding the best solution** among a list of feasible solutions by minimizing or maximizing an *objective function* based on the values of decision variables.

1 Optimization using the 2nd order derivatives

Let's consider a simple optimization problem where we aim to minimize a scalar-valued function $f(x) = x^3 - 6x^2 + 9x + 1$ with respect to x.

STEP 1: We solve $f'(x) = 0$ to obtain the critical points (Extreme points)

$$f'(x) = 3x^2 - 12x + 9 = 0$$

$$x = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

STEP 2: We then find $f''(x)$ for all critical points and compare it to 0.

$$f''(x) = 6x - 12$$

$$f''\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} -6 \\ 6 \end{bmatrix}$$

Since $f''(1) < 0$, then $x = 1$ is a local maximum.

Since $f''(3) > 0$, then $x = 3$ is a local minimum.

```
syms x
f(x) = x^3 - 6*x^2 + 9*x + 1
```

f(x) = $x^3 - 6\,x^2 + 9\,x + 1$

```
% Find f'(x)
fderiv = diff(f)
```

fderiv(x) = $3\,x^2 - 12\,x + 9$

```
% Solve f'(x) = 0
sol = solve(fderiv)
```

sol =

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

```
% Find f''(x)
sderiv = diff(f, 2)
```

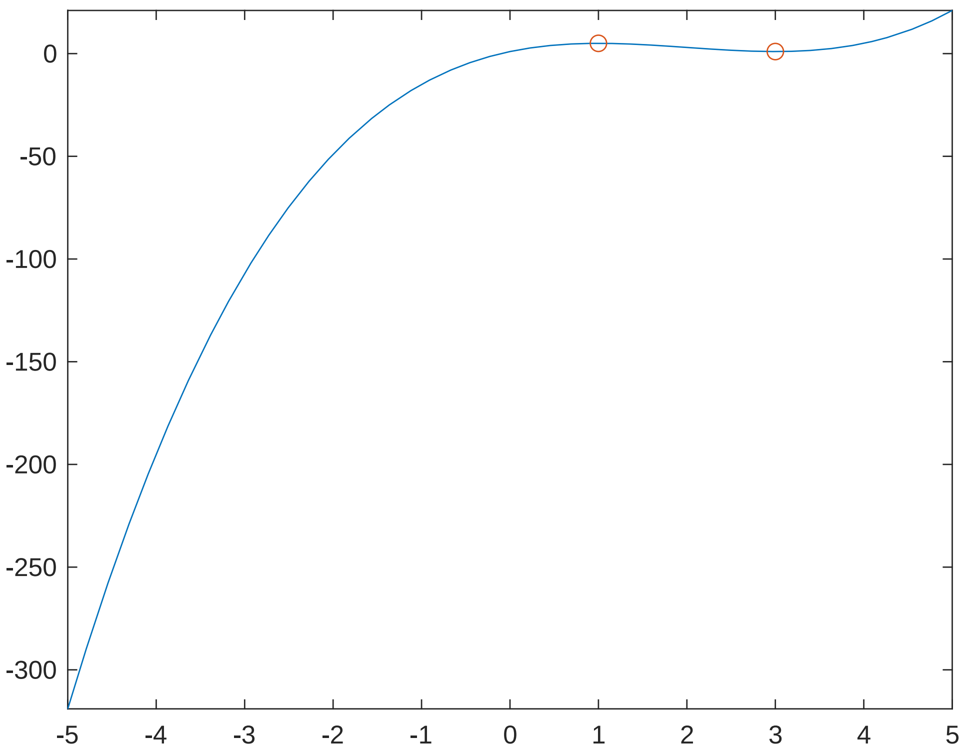
sderiv(x) = $6\,x - 12$

```
% Substitute with Critical Points
cp = sderiv(sol)
```

cp =

$$\begin{pmatrix} -6 \\ 6 \end{pmatrix}$$

```
% Plotf(x) with local maximum/minimum
fplot(f)
hold on
scatter(sol, f(sol))
hold off
```



Limitations with Direct Use of 2nd order derivative

When optimizing using the second derivative, you compute and directly utilize **the second-order partial derivatives of the objective function with respect to each variable**. This approach can be **computationally expensive and numerically unstable**, especially for high-dimensional problems, as it requires computing and storing multiple second-order derivatives.

The Hessian matrix, on the other hand, **aggregates all the second-order partial derivatives into a symmetric matrix**. By computing the Hessian matrix, you obtain a compact representation of the *curvature* of the objective function. This matrix provides valuable information about the local curvature around a given point in the optimization space.

2 Hessian Matrix

The Hessian matrix of $f(x)$ is the square matrix of the second partial derivatives of $f(x)$.

Given a function $f(X)$ where $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ the Hessian matrix $H(f(X)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$

Classify the extreme points for the function $f(X) = 3{x_1}^2 + 2{x_2}^3 - 6x_1x_2$ where $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

STEP 1: Find the critical points using the gradient to the function:

$$\nabla(f) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 6x_1 - 6x_2 \\ 6x_2^2 - 6x_1 \end{bmatrix}$$

Set $\nabla(f) = 0$ to find critical points

For $x_1 = x_2$ and $x_2^2 = x_1$, We got 2 critical points as follows:

$p1 = (0, 0)$
$p2 = (1, 1)$

STEP 2: Find the Hessian Matrix $H(f(X))$

$$H(f(X)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 6 & -6 \\ -6 & 12x_2 \end{bmatrix}$$

STEP 3: Classify the critical points as follows:

- Find the determinant of the Hessian matrix at each critical point.
- If the determinatn is **negative** then the point is **Saddle.**
- Else check the value of $\frac{\partial^2 f}{\partial x_1^2}$. **If positive** the point is **local minimum. If negative** the point is **local maximum.**

$|H(f(0, 0))| = \begin{bmatrix} 6 & -6 \\ -6 & 0 \end{bmatrix} = 0 - 36 = -36$. Hence, $(0, 0)$ is *saddle* point.

$|H(f(1, 1))| = \begin{bmatrix} 6 & -6 \\ -6 & 12 \end{bmatrix} = 72 - 36 = 36$. Hence, (1, 1) is *local minumum* because $\frac{\partial^2 f}{\partial x_1^2} > 0$.

```
syms x1 x2
f(x1, x2) = 3*x1^2 + 2*x2^3 - 6*x1*x2
```

f(x1, x2) = $3x_1^2 - 6x_1x_2 + 2x_2^3$

```
g = gradient(f)
```

g(x1, x2) =

$$\begin{pmatrix} 6x_1 - 6x_2 \\ 6x_2^2 - 6x_1 \end{pmatrix}$$

```
s = solve(g)
```

```
s = struct with fields:
    x1: [2x1 sym]
    x2: [2x1 sym]
```

```
p1 = [s.x1(1) s.x2(1)]
```

p1 = $\begin{pmatrix} 0 & 0 \end{pmatrix}$

```
p2 = [s.x1(2) s.x2(2)]
```

p2 = $\begin{pmatrix} 1 & 1 \end{pmatrix}$

```
% we can directly find the hessian matrix
hmat = hessian(f)
```

hmat(x1, x2) =

$$\begin{pmatrix} 6 & -6 \\ -6 & 12x_2 \end{pmatrix}$$

```
% find the determinant
detp1 = det(hmat(s.x1(1), s.x2(1)))
```

detp1 = -36

```
detp2 = det(hmat(s.x1(2), s.x2(2)))
```

detp2 = 36

```
% find 2nd order partial derivative of x1
p2hmat = hmat(1, 1);
if p2hmat(1) > 0
    disp("p2 is local min")
else
    disp("p2 is local max")
end
```

```
p2 is local min
```

3 Example

Classify the critical points to this function $f(V) = x^3 + y^3 + 2x^2 + 4y^2 + 6$ where $V = \begin{bmatrix} x \\ y \end{bmatrix}$.

STEP1: Find Critical Points

$$\nabla f = \begin{bmatrix} 3x^2 + 0 \ + 4x + 0 + 0 \\ 0 \ + \ 3y^2 + 0 + 8y + 0 \end{bmatrix} = \begin{bmatrix} 3x^2 \ + 4(x) \\ 3y^2 + 8y \end{bmatrix} = \begin{bmatrix} x(3x+4) \\ y(3y+8) \end{bmatrix}$$

For $\nabla f = 0$.

- $x = 0$ or $x = \frac{-4}{3}$
- $y = 0$ or $y = \frac{-8}{3}$

So we got 4 possible critical points as follows:

- $(0, 0)$
- $\left(0, \frac{-8}{3}\right)$
- $\left(\frac{-4}{3}, 0\right)$
- $\left(\frac{-4}{3}, \frac{-8}{3}\right)$

STEP 2: Find the Hessian matrix

$$H(f) = \begin{bmatrix} 6x+4 & 0 \\ 0 & 6y+8 \end{bmatrix}$$

STEP 3: Classify critical points

$|H(f(0, 0))| = 4 \times 8 = 32$. Hence, (0, 0) is ***local minumum*** because $\frac{\partial^2 f}{\partial x_1^2} = 6 \ \times 0 + 4 = 4 > 0.$

$\left|H\left(f\left(0, \frac{-8}{3}\right)\right)\right| = 4 \times -8 = -32$. Hence, $\left(0, \frac{-8}{3}\right)$ is ***saddle point.***

$\left|H\left(f\left(\frac{-4}{3}, 0\right)\right)\right| = -4 \times 8 = -32$. Hence, $\left(\frac{-4}{3}, 0\right)$ is ***saddle point.***

$\left|H\left(f\left(\frac{-4}{3}, \frac{-8}{3}\right)\right)\right| = -4 \times -8 = 32$. Hence, $\left(\frac{-4}{3}, \frac{-8}{3}\right)$ is ***local maximum*** because $\frac{\partial^2 f}{\partial x_1^2} = 6 \ \times \frac{-4}{3} + 4 = -4 < 0.$

```
syms x y
f2(x, y) = x^3 + y^3 + 2*x^2 + 4*y^2 + 6
```

f2(x, y) = $x^3 + 2\,x^2 + y^3 + 4\,y^2 + 6$

```
g = gradient(f2)
```

g(x, y) =

$$\begin{pmatrix} 3x^2+4x \\ 3y^2+8y \end{pmatrix}$$

```
h = hessian(f2)
```

h(x, y) =

$$\begin{pmatrix} 6x+4 & 0 \\ 0 & 6y+8 \end{pmatrix}$$

```
det1 = det(h(0, 0))
```

det1 = 32

```
h1 = h(0, 0)
```

h1 =

$$\begin{pmatrix} 4 & 0 \\ 0 & 8 \end{pmatrix}$$

```
d2fdx2_1 = h1(1)
```

d2fdx2_1 = 4

```
det2 = det(h(0, -8/3))
```

det2 = -32

```
det3 = det(h(-4/3, 0))
```

det3 = -32

```

det4 = det(h(-4/3, -8/3))
```

det4 = 32

```
h4 = h(-4/3, -8/3)
```

h4 =

$$\begin{pmatrix} -4 & 0 \\ 0 & -8 \end{pmatrix}$$

```
d2fdx2_4 = h4(1)
```

d2fdx2_4 = -4