مبانی یادگیری ماشین

Intro to Machine Learning

بابک نجار اعرابی

دانشکده مهندسی برق و کامپیوتر دانشگاه تهران

نیم سال اول سال تحصیلی 1403-04

موضوع این جلسه

مروری بر روش های بهینه سازی

جلسه اول

Babak Nadjar Araabi

School of Electrical & Computer Eng

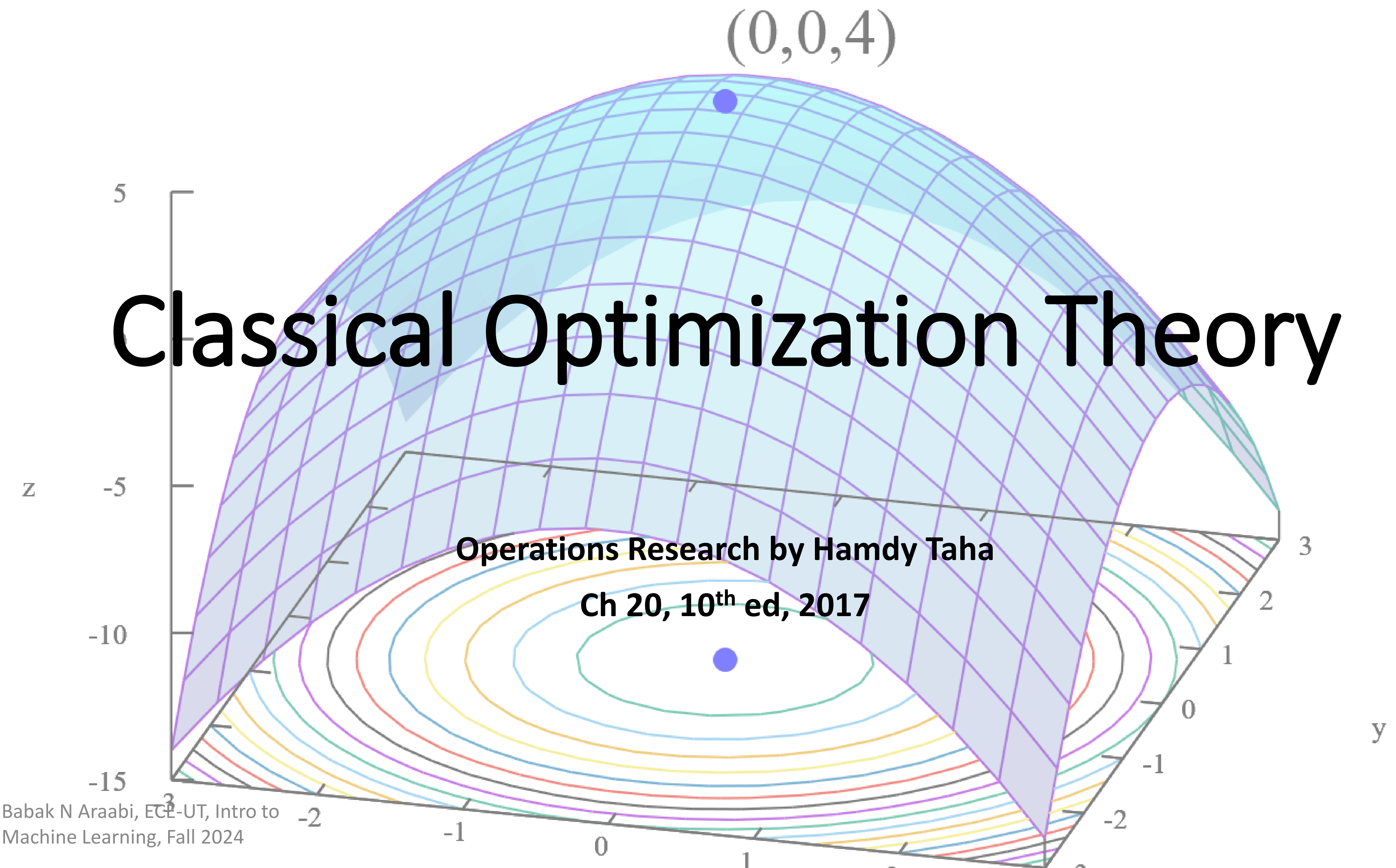
University of Tehran

ECE-UT - Fall 2024

Classical Optimization Theory

Operations Research by Hamdy Taha

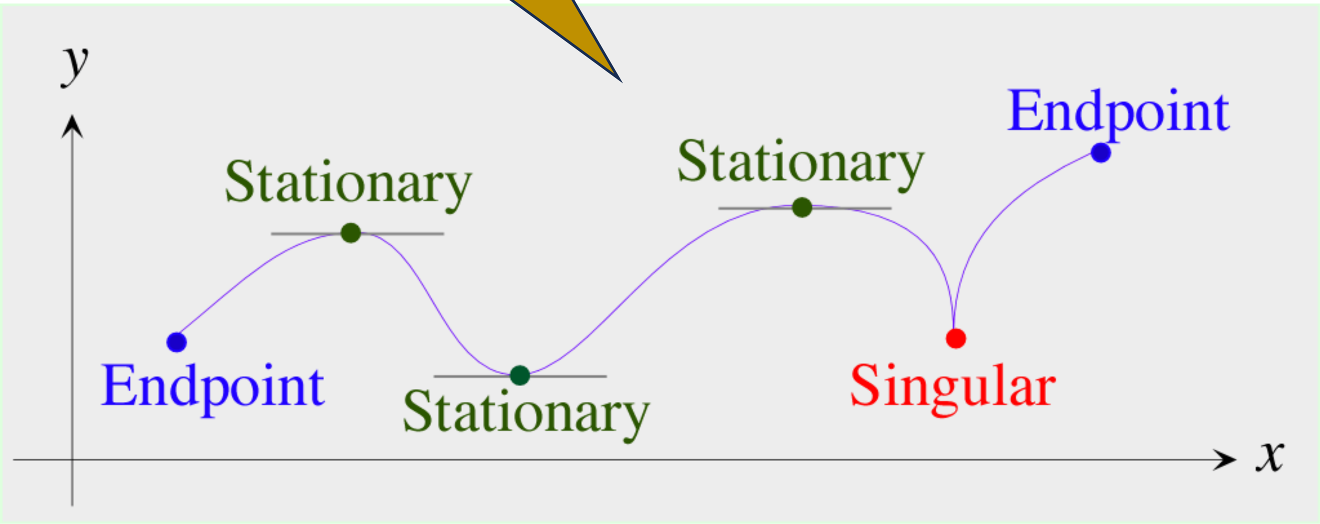
Ch 20, 10th ed, 2017

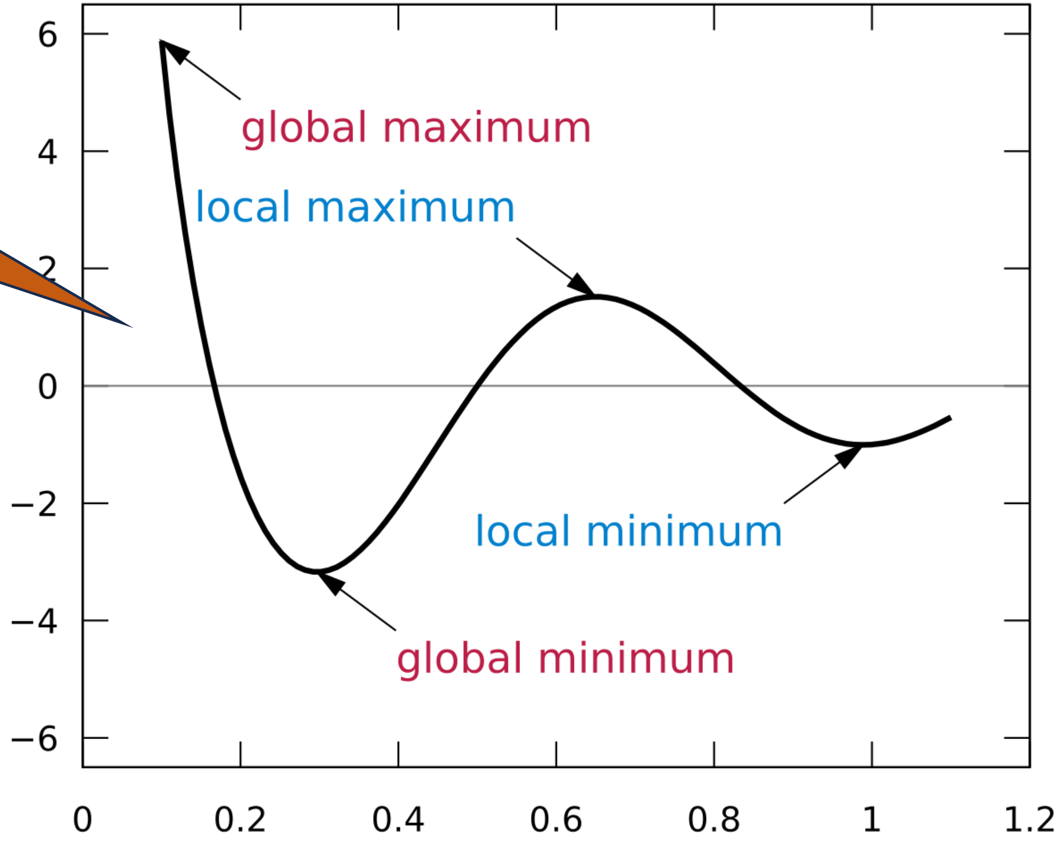


Minima and Maxima

Stationary point
End point
Singular point

Local/Global
Minimum/Maximum

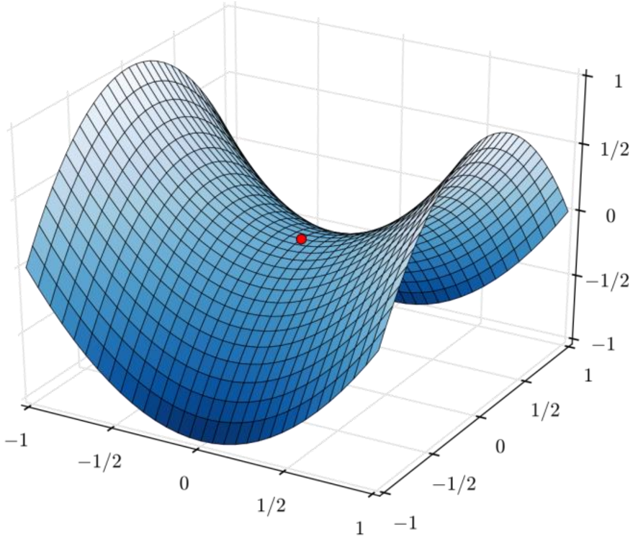


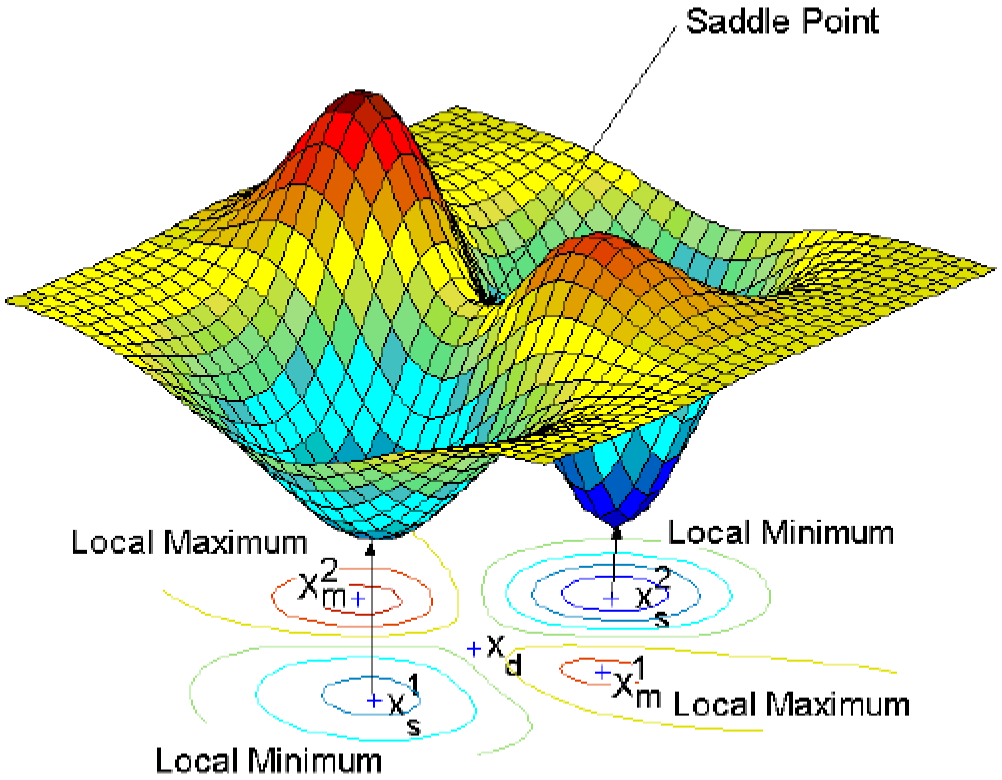


https://www.zweigmedia.com/ & Wikipedia

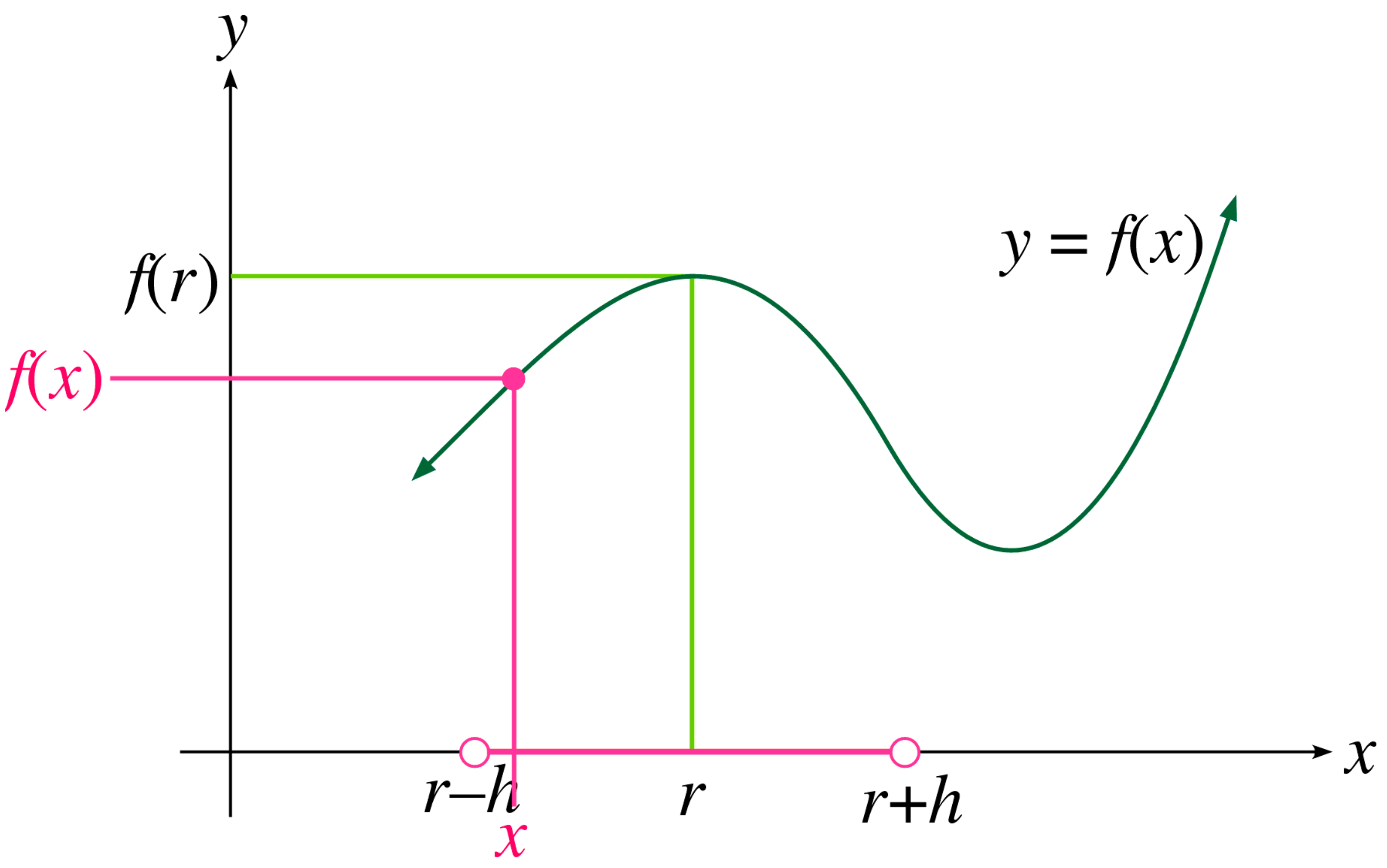
Minima and Maxima, 3D

A saddle point (in red) on the graph of $z = x^2 - y^2$ (hyperbolic paraboloid)





A Local Vicinity (infinite-small)



https://www.zweigmedia.com

Unconstrained Optimization

An extreme point of a function $f(\mathbf{X})$ defines either a maximum or a minimum of the function. Mathematically, a point $\mathbf{X}_0 = (x_1^0, \ldots, x_j^0, \ldots, x_n^0)$ is a maximum if

$$f(\mathbf{X}_0 + \mathbf{h}) \leq f(\mathbf{X}_0)$$

for all $\mathbf{h} = (h_1, \ldots, h_j, \ldots, h_n)$, where $|h_j|$ is sufficiently small for all j. In a similar manner, $\mathbf{X}_0$ is a minimum if

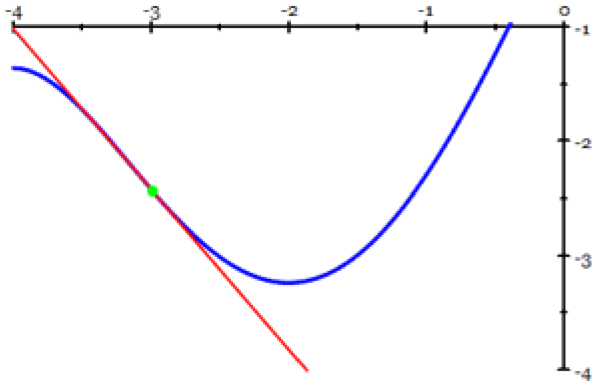
$$f(\mathbf{X}_0 + \mathbf{h}) \geq f(\mathbf{X}_0)$$

Necessary Condition for Extreme Point

Theorem 20.1-1. *A necessary condition for* $\mathbf{X}_0$ *to be an extreme point of* $f(\mathbf{X})$ *is that*

$$\nabla f(\mathbf{X}_0) = 0$$

Because the necessary condition is also satisfied at inflection and saddle points, it is more appropriate to refer to the points obtained from the solution of $\nabla f(\mathbf{X}_0) = \mathbf{0}$ as **stationary** points. The next theorem establishes the sufficiency conditions for $\mathbf{X}_0$ to be an extreme point.



https://www.jonahgreenthal.com/calculus/lessons/3.2/animation.gif

Sufficient Condition for Extremum

Theorem 20.1-2. *A sufficient condition for a stationary point* $\mathbf{X}_0$ *to be an extremum is that the Hessian matrix* $\mathbf{H}$ *evaluated at* $\mathbf{X}_0$ *satisfy the following conditions*:

(i) $\mathbf{H}$ *is positive definite if* $\mathbf{X}_0$ *is a minimum point.*
(ii) $\mathbf{H}$ *is negative definite if* $\mathbf{X}_0$ *is a maximum point.*

In general, if $|\mathbf{H}|_{\mathbf{X}_0}$ is indefinite, $\mathbf{X}_0$ must be a saddle point. For nonconclusive cases, $\mathbf{X}_0$ may or may not be an extremum, and the sufficiency condition becomes rather involved, because higher-order terms in Taylor's expansion must be considered.

1D Function

If $f''(y_0) = 0$, higher-order derivatives must be investigated as the following theorem requires.

Theorem 20.1-3. *Given y_0, a stationary point of $f(y)$, if the first $(n-1)$ derivatives are zero and $f^{(n)}(y_0) \neq 0$, then*

(i) *If n is odd, y_0 is an inflection point.*

(ii) *If n is even, then y_0 is a minimum if $f^{(n)}(y_0) > 0$ and a maximum if $f^{(n)}(y_0) < 0$.*

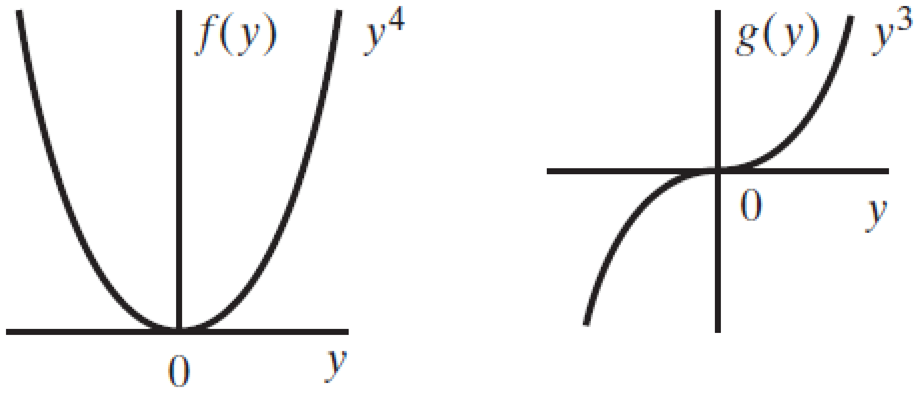


FIGURE 20.2

Extreme points of $f(y) = y^4$ and $g(y) = y^3$

Newton–Raphson Method

In general, the necessary condition $\nabla f(\mathbf{X}) = \mathbf{0}$ may be highly nonlinear and, hence, difficult to solve. The Newton–Raphson method is an iterative algorithm for solving simultaneous nonlinear equations.

Consider the simultaneous equations

$$f_i(\mathbf{X}) = 0, i = 1, 2, \ldots, m$$

Let $\mathbf{X}^k$ be a given point. Then by Taylor's expansion

$$f_i(\mathbf{X}) \approx f_i(\mathbf{X}_k) + \nabla f_i(\mathbf{X}_k)(\mathbf{X} - \mathbf{X}_k), i = 1, 2, \ldots, m$$

Thus, the original equations, $f_i(\mathbf{X}) = 0, i = 1, 2, \ldots, m$, may be approximated as

$$f_i(\mathbf{X}_k) + \nabla f_i(\mathbf{X}_k)(\mathbf{X} - \mathbf{X}_k) = 0, i = 1, 2, \ldots, m$$

These equations may be written in matrix notation as

$$\mathbf{A}_k + \mathbf{B}_k(\mathbf{X} - \mathbf{X}_k) = \mathbf{0}$$

If $\mathbf{B}_k$ is nonsingular, then

$$\mathbf{X} = \mathbf{X}_k - \mathbf{B}_k^{-1}\mathbf{A}_k$$

Iterative Process in Newton–Raphson

The idea of the method is to start from an initial point $\mathbf{X}_0$, and then use the equation above to determine a new point. The process may or may not converge depending on the selection of the starting point. Convergence occurs when two successive points, $\mathbf{X}_k$ and $\mathbf{X}_{k+1}$, are approximately equal (within specified acceptable tolerance).

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$f'(x_k) = \dfrac{f(x_k)}{x_k - x_{k+1}}$—meaning that x_{k+1} is determined from the slope of $f(x)$ at x_k, where tan $\theta = f'(x_k)$