# Theoretical Models for Computing: Optimization

Presenter: Dr. Ha Viet Uyen Synh.

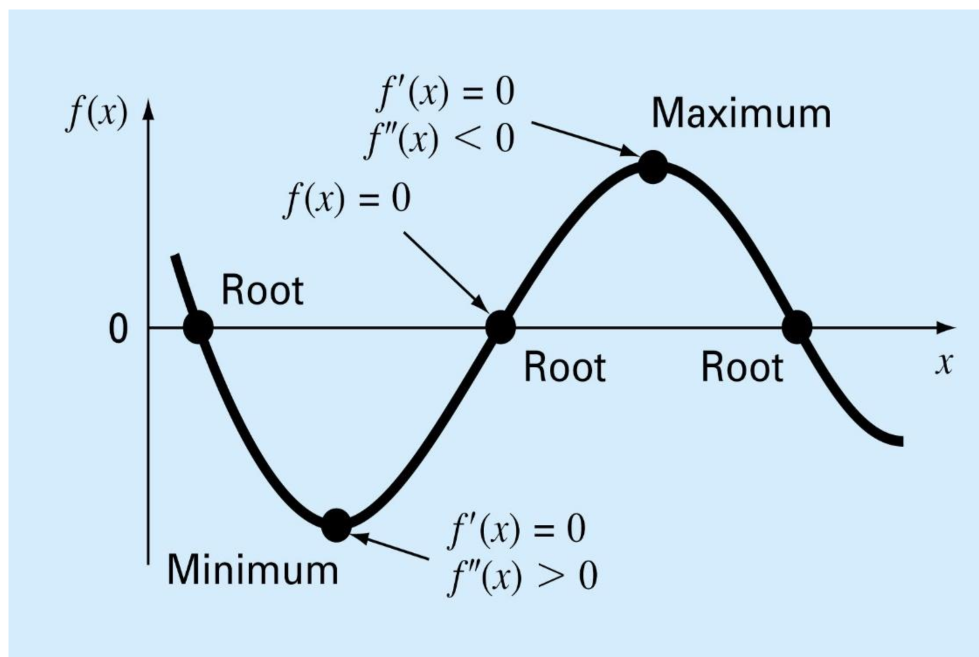
# Optimization

Root finding and optimization are related, both involve guessing and searching for a point on a function.

Fundamental difference is:

Root finding is searching for zeros of a function or functions

Optimization is finding the minimum or the maximum of a function of several variables.



# Mathematical Background

An *optimization* or *mathematical programming* problem generally be stated as:

Find $x$, which minimizes or maximizes $f(x)$ subject to

$$d_i(x) \leq a_i \quad i = 1,2,\ldots,m^*$$
$$e_i(x) = b_i \quad i = 1,2,\ldots,p^*$$

Where $x$ is an $n$-dimensional *design vector*, $f(x)$ is the *objective function*, $d_i(x)$ are *inequality constraints*, $e_i(x)$ are *equality constraints*, and $a_i$ and $b_i$ are constants

Optimization problems can be classified on the basis of the form of *f(x):*

If *f(x)* and the constraints are linear, we have *linear programming*.

If *f(x)* is quadratic and the constraints are linear, we have *quadratic programming*.

If *f(x)* is not linear or quadratic and/or the constraints are nonlinear, we have *nonlinear programming*.

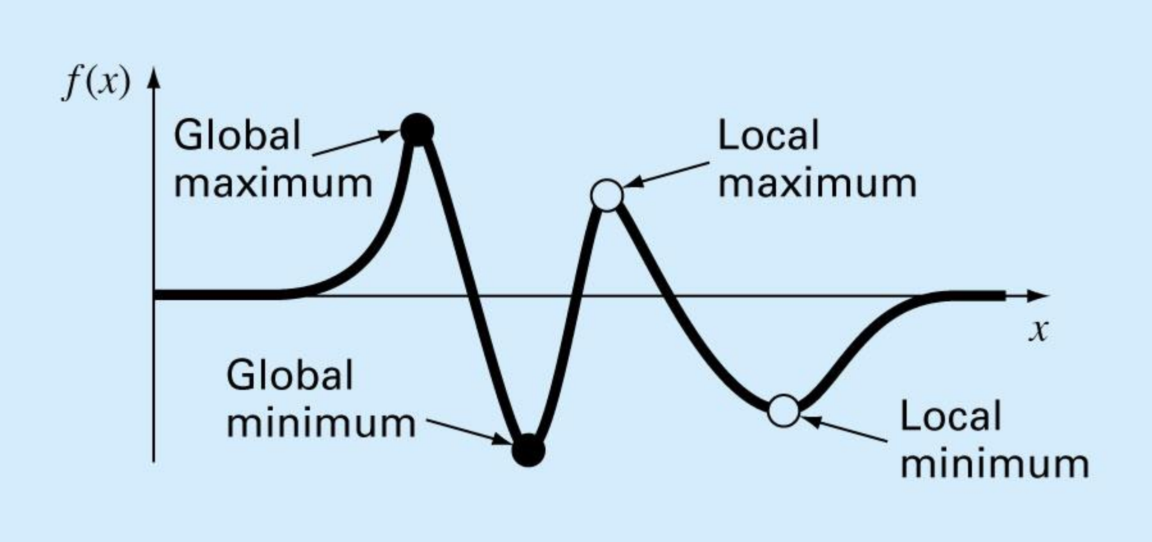
When equations(*) are included, we have a *constrained optimization* problem; otherwise, it is *unconstrained optimization* problem.

PART A

# ONE-DIMENSIONAL UNCONSTRAINED OPTIMIZATION

# One-Dimensional Unconstrained Optimization

In *multimodal* functions, both local and global optima can occur. In almost all cases, we are interested in finding the absolute highest or lowest value of a function.



# How do we distinguish global optimum from local one?

By graphing to gain insight into the behavior of the function.

Using randomly generated starting guesses and picking the largest of the optima as global.

Perturbing the starting point to see if the routine returns a better point or the same local minimum.

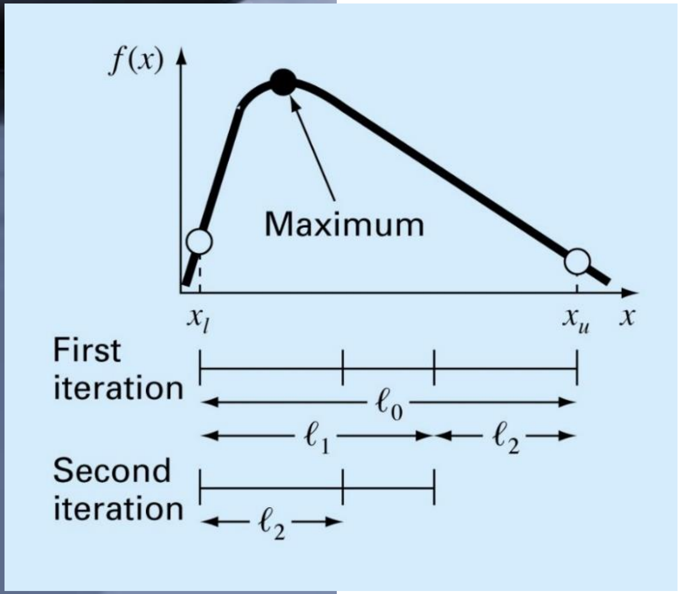
# Golden-Section Search

A *unimodal* function has a *single maximum or a minimum* in the a given interval. For a unimodal function:

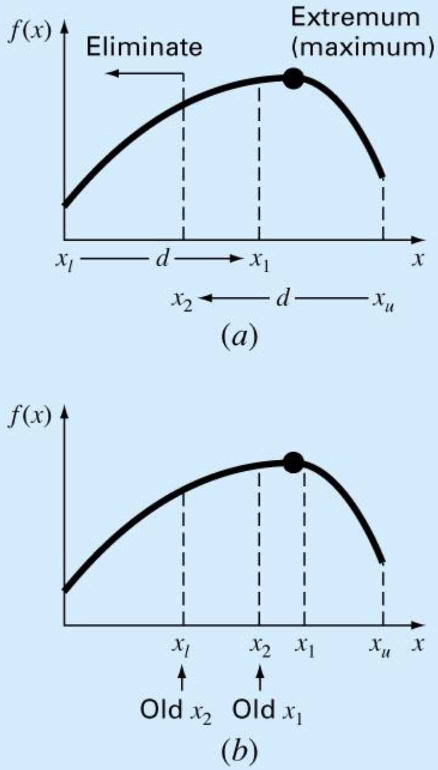
- First pick two points that will bracket your extremum [$x_l$, $x_u$].

- Pick an additional third point within this interval to determine whether a maximum occurred.

- Then pick a fourth point to determine whether the maximum has occurred within the first three or last three points




The key is making this approach efficient by choosing intermediate points wisely thus minimizing the function evaluations by replacing the old values with new values.




$$l_0 = l_1 + l_2$$

$$\frac{l_1}{l_0} = \frac{l_2}{l_1}$$

- The first condition specifies that the sum of the two sub lengths $l_1$ and $l_2$ must equal the original interval length.

- The second say that the ratio of the length must be equal

$$\frac{l_1}{l_1 + l_2} = \frac{l_2}{l_1} \qquad R = \frac{l_2}{l_1}$$

$$1 + R = \frac{1}{R} \qquad R^2 + R - 1 = 0$$

$$R = \frac{-1 + \sqrt{1 - 4(-1)}}{2} = \frac{\sqrt{5} - 1}{2} = 0.61803$$

**Golden Ratio**

The method starts with two initial guesses, $x_l$ and $x_u$, that bracket one local extremum of $f(x)$:

Next two interior points $x_1$ and $x_2$ are chosen according to the golden ratio

$$d = \frac{\sqrt{5}-1}{2}(x_u - x_l)$$

$$x_1 = x_l + d$$

$$x_2 = x_u - d$$

The function is evaluated at these two interior points.

Two results can occur:

If $f(x_1)>f(x_2)$ then the domain of $x$ to the left of $x_2$ from $x_l$ to $x_2$, can be eliminated because it does not contain the maximum. Then, $x_2$ becomes the new $x_l$ for the next round.

If $f(x_2)>f(x_1)$, then the domain of $x$ to the right of $x_1$ from $x_l$ to $x_2$, would have been eliminated. In this case, $x_1$ becomes the new $x_u$ for the next round.

New $x_1$ determined as before

$$x_1 = x_l + \frac{\sqrt{5}-1}{2}(x_u - x_l)$$

$$\varepsilon_a = (1-R)\left|\frac{x_u - x_l}{x_{opt}}\right| 100\%$$

# Example #1

Use the golden-section search to find the maximum of

$$f(x) = 2\sin x - \frac{x^2}{10}$$

within the interval $x_l = 0$ and $x_u = 4$.

First, the golden ratio is used to create the two interior points

$$d = \frac{\sqrt{5}-1}{2}(4-0) = 2.472$$

$$x_1 = 0 + 2.472 = 2.472$$

$$x_2 = 4 - 2.472 = 1.528$$

The function can be evaluated at the interior points

$$f(x_2) = f(1.528) = 2\sin(1.528) - \frac{1.528^2}{10} = 1.765$$

$$f(x_1) = f(2.472) = 0.63$$

Because $f(x_2) > f(x_1)$, the maximum is in the interval defined by $x_l$, $x_2$, and $x_1$. Thus, for the new interval, the lower bound remains $x_l = 0$, and $x_1$ becomes the upper bound, that is, $x_u = 2.472$.

In addition, the former $x_2$ value becomes the new $x_1$, that is, $x_1$ = 1.528. Further, we do not have to recalculate f (x1) because it was determined on the previous iteration as f (1.528) = 1.765.

All that remains is to compute the new values of d and x2,

$$d = \frac{\sqrt{5}-1}{2}(2.472 - 0) = 1.528$$

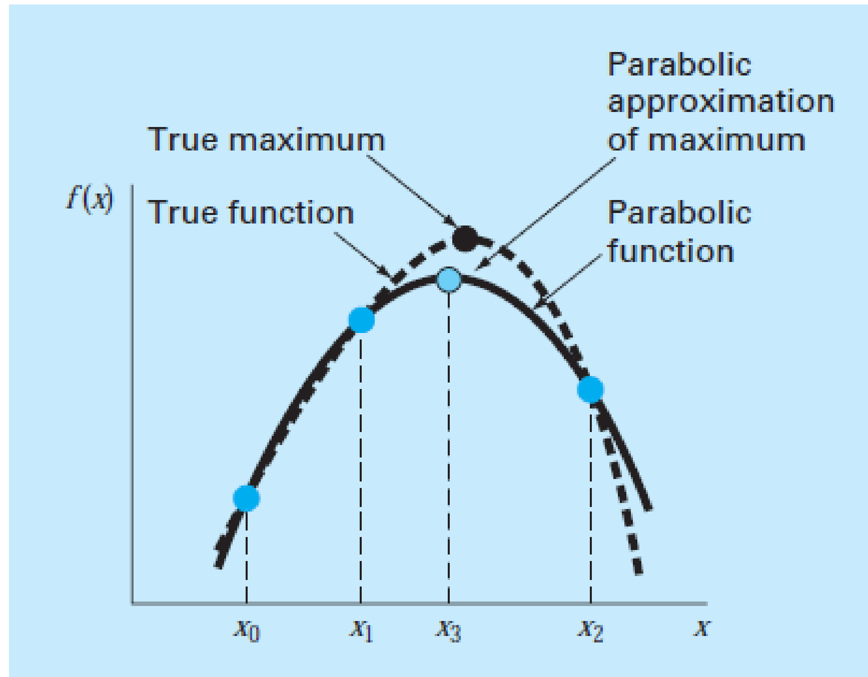
$$x_2 = 2.4721 - 1.528 = 0.944$$

The function evaluation at $x_2$ is f(0.994) = 1.531. Since this value is less than the function value at $x_1$, the maximum is in the interval prescribed by $x_2$, $x_1$, and $x_u$.

| *i* | $x_l$ | $f(x_l)$ | $x_2$ | $f(x_2)$ | $x_1$ | $f(x_1)$ | $x_u$ | $f(x_u)$ | *d* |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 1.5279 | 1.7647 | 2.4721 | 0.6300 | 4.0000 | −3.1136 | 2.4721 |
| 2 | 0 | 0 | 0.9443 | 1.5310 | 1.5279 | 1.7647 | 2.4721 | 0.6300 | 1.5279 |
| 3 | 0.9443 | 1.5310 | 1.5279 | 1.7647 | 1.8885 | 1.5432 | 2.4721 | 0.6300 | 0.9443 |
| 4 | 0.9443 | 1.5310 | 1.3050 | 1.7595 | 1.5279 | 1.7647 | 1.8885 | 1.5432 | 0.5836 |
| 5 | 1.3050 | 1.7595 | 1.5279 | 1.7647 | 1.6656 | 1.7136 | 1.8885 | 1.5432 | 0.3607 |
| 6 | 1.3050 | 1.7595 | 1.4427 | 1.7755 | 1.5279 | 1.7647 | 1.6656 | 1.7136 | 0.2229 |
| 7 | 1.3050 | 1.7595 | 1.3901 | 1.7742 | 1.4427 | 1.7755 | 1.5279 | 1.7647 | 0.1378 |
| 8 | 1.3901 | 1.7742 | 1.4427 | 1.7755 | 1.4752 | 1.7732 | 1.5279 | 1.7647 | 0.0851 |

The result is converging on the true value of 1.7757 at x = 1.4276.

# Parabolic Interpolation




$$x_3 = \frac{f(x_0)\left(x_1^2 - x_2^2\right) + f(x_1)\left(x_2^2 - x_0^2\right) + f(x_2)\left(x_0^2 - x_1^2\right)}{2\, f(x_0)(x_1 - x_2) + 2\, f(x_1)(x_2 - x_0) + 2\, f(x_2)(x_0 - x_1)}$$

where $x_0$, $x_1$, and $x_2$ are the initial guesses, and $x_3$ is the value of x that corresponds to the maximum value of the parabolic fit to the guesses.

# Example #2

Use the parabolic interpolation to find the maximum of

$$f(x) = 2\sin x - \frac{x^2}{10}$$

within the interval x0 = 0, x1=1 and x2 = 4.

The function values at the three guesses can be evaluated,

$$x_0 = 0 \qquad f(x_0) = 0$$
$$x_1 = 1 \qquad f(x_1) = 1.5829$$
$$x_2 = 4 \qquad f(x_2) = -3.1136$$

$$x_3 = \frac{0(1^2 - 4^2) + 1.5829(4^2 - 0^2) + (-3.1136)(0^2 - 1^2)}{2(0)(1 - 4) + 2(1.5829)(4 - 0) + 2(-3.1136)(0 - 1)} = 1.5055$$

which has a function value of f(1.5055) = 1.7691. For the next iteration,

$$x_0 = 1 \qquad f(x_0) = 1.5829$$
$$x_1 = 1.5055 \qquad f(x_1) = 1.7691$$
$$x_2 = 4 \qquad f(x_2) = -3.1136$$

$$x_3 = \frac{1.5829(1.5055^2 - 4^2) + 1.7691(4^2 - 1^2) + (-3.1136)(1^2 - 1.5055^2)}{2(1.5829)(1.5055 - 4) + 2(1.7691)(4 - 1) + 2(-3.1136)(1 - 1.5055)}$$
$$= 1.4903$$

which has a function value of f(1.4903) = 1.7714.

| $i$ | $x_0$ | $f(x_0)$ | $x_1$ | $f(x_1)$ | $x_2$ | $f(x_2)$ | $x_3$ | $f(x_3)$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.0000 | 0.0000 | 1.0000 | 1.5829 | 4.0000 | −3.1136 | 1.5055 | 1.7691 |
| 2 | 1.0000 | 1.5829 | 1.5055 | 1.7691 | 4.0000 | −3.1136 | 1.4903 | 1.7714 |
| 3 | 1.0000 | 1.5829 | 1.4903 | 1.7714 | 1.5055 | 1.7691 | 1.4256 | 1.7757 |
| 4 | 1.0000 | 1.5829 | 1.4256 | 1.7757 | 1.4903 | 1.7714 | 1.4266 | 1.7757 |
| 5 | 1.4256 | 1.7757 | 1.4266 | 1.7757 | 1.4903 | 1.7714 | 1.4275 | 1.7757 |

Thus, within five iterations, the result is converging rapidly on the true value of 1.7757 at x = 1.4276.

# Newton’s Method

A similar approach to Newton- Raphson method can be used to find an optimum of f(x) by defining a new function g(x)=f‘(x). Thus because the same optimal value x* satisfies both

$$f'(x^*)=g(x^*)=0$$

We can use the following as a technique to the extremum of *f(x)*.

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}$$

# Example #3

Use the Newton's Method to find the maximum of

$$f(x) = 2\sin x - \frac{x^2}{10}$$

within an initial guess of $x_0$=2.5.

The first and second derivatives of the function can be evaluated as

$$f'(x) = 2\cos x - \frac{x}{5}$$

$$f''(x) = -2\sin x - \frac{1}{5}$$

$$x_{i+1} = x_i - \frac{2\cos x_i - x_i/5}{-2\sin x_i - 1/5}$$

| *i* | x | f(x) | f′(x) | f″(x) |
|---|---|---|---|---|
| 0 | 2.5 | 0.57194 | −2.10229 | −1.39694 |
| 1 | 0.99508 | 1.57859 | 0.88985 | −1.87761 |
| 2 | 1.46901 | 1.77385 | −0.09058 | −2.18965 |
| 3 | 1.42764 | 1.77573 | −0.00020 | −2.17954 |
| 4 | 1.42755 | 1.77573 | 0.00000 | −2.17952 |

Thus, within four iterations, the result converges rapidly on the true value.

PART B

# MULTIDIMENSIONAL UNCONSTRAINED OPTIMIZATION

# Multidimensional Unconstrained Optimization

Techniques to find minimum and maximum of a function of several variables are described.

These techniques are classified as:

That require derivative evaluation *Gradient* or *descent* (or *ascent*) methods

That do not require derivative evaluation *Non-gradient* or *direct* methods.

# DIRECT METHODS Random Search

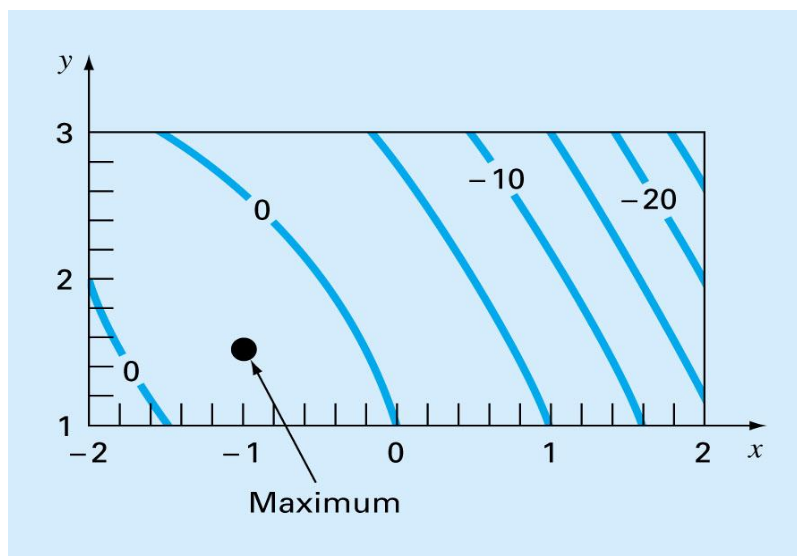
Based on evaluation of the function randomly at selected values of the independent variables.

If a sufficient number of samples are conducted, the optimum will be eventually located.

Example: maximum of a function

$$f(x, y) = y - x - 2x^2 - 2xy - y^2$$

can be found using a random number generator.



**Advantages**

Works even for discontinuous and nondifferentiable functions.

Always finds the global optimum rather than the global minimum.

**Disadvantages**

As the number of independent variables grows, the task can become onerous.

Not efficient, it does not account for the behavior of underlying function.

# Univariate and Pattern Searches

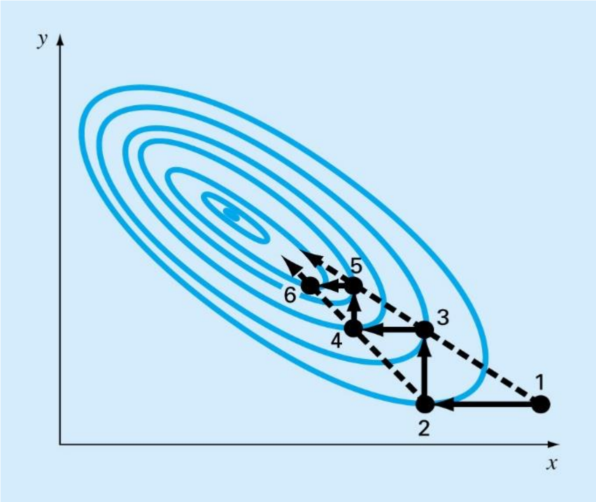
More efficient than random search and still doesn’t require derivative evaluation.

**The basic strategy is:**

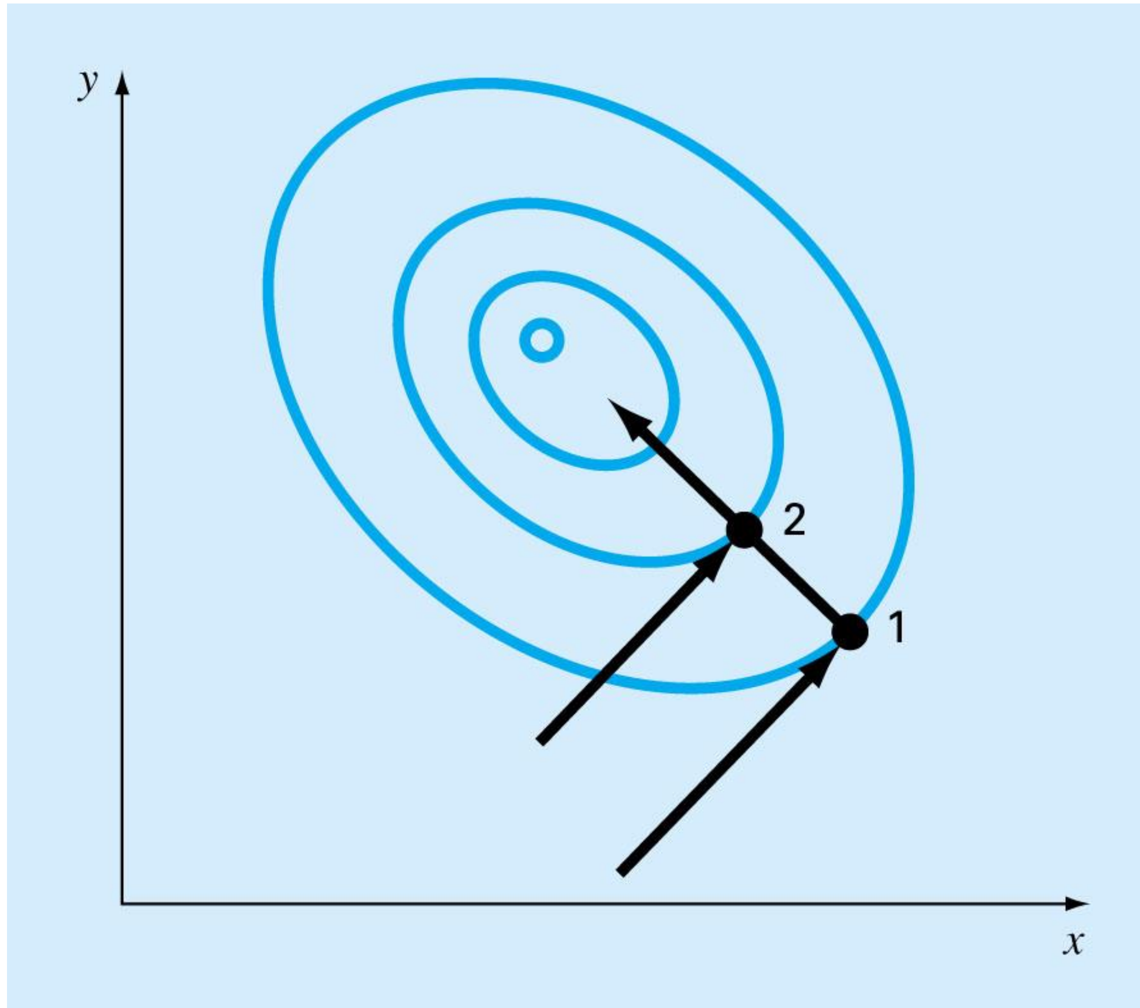
*Change one variable at a time* while the other variables are held constant.

Thus problem is reduced to a sequence of one-dimensional searches that can be solved by variety of methods.

The search becomes less efficient as you approach the maximum.



*Pattern directions* can be used to shoot directly along the ridge towards maximum.



# GRADIENT METHODS
# Gradients and Hessians

## The Gradient

If *f(x,y)* is a two dimensional function, the *gradient* vector tells us

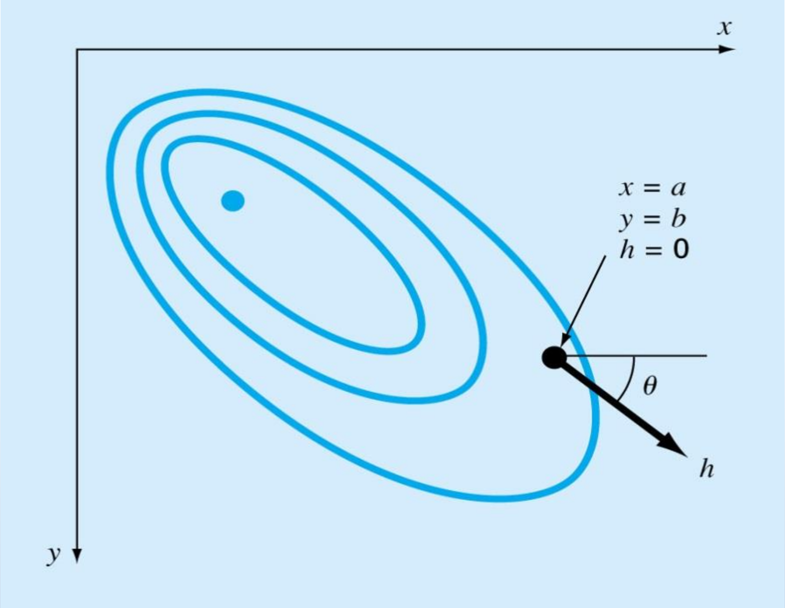
What direction is the steepest ascend?

How much we will gain by taking that step?

$$\nabla f = \frac{\partial f}{\partial x}\mathrm{i} + \frac{\partial f}{\partial y}\mathrm{j} \qquad \text{or } \textit{del f}$$

Directional derivative of *f(x,y)* at point x=a and y=b

•For $n$ dimensions



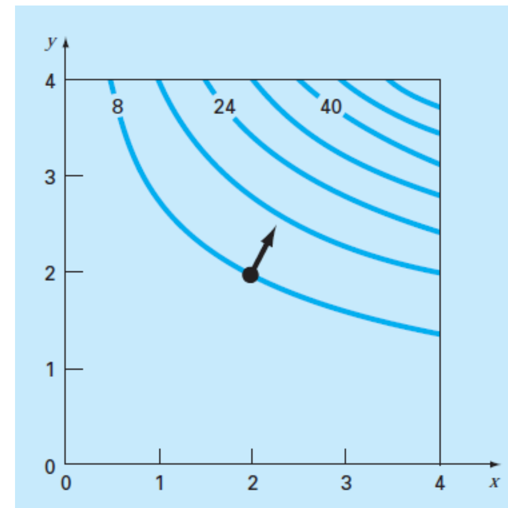

$$\nabla f(x) = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1}(\mathrm{x}) \\ \dfrac{\partial f}{\partial x_2}(\mathrm{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\mathrm{x}) \end{Bmatrix}$$

# Example #4

Employ the gradient to evaluate the steepest ascent direction for the function

$$f(x, y) = xy^2$$

at the point (2, 2). Assume that positive x is pointed east and positive y is pointed north




First, our elevation can be determined as

$$f(2, 2) = 2(2)2 = 8$$

Next, the partial derivatives can be evaluated,

$$\frac{\partial f}{\partial x} = y^2 = 2^2 = 4$$

$$\frac{\partial f}{\partial y} = 2xy = 2(2)(2) = 8$$

which can be used to determine the gradient as

$$\nabla f = 4i + 8j$$

This vector can be sketched on a topographical map of the function. This immediately tells us that the direction we must take is

$$\theta = \tan^{-1}\left(\frac{8}{4}\right) = 1.107 \text{ radians } (= 63.4^\circ)$$

The slope in this direction, which is the magnitude of ∇f , can be calculated as

$$\sqrt{4^2 + 8^2} = 8.944$$

# The Hessian

For one dimensional functions both first and second derivatives  valuable information for searching out optima.

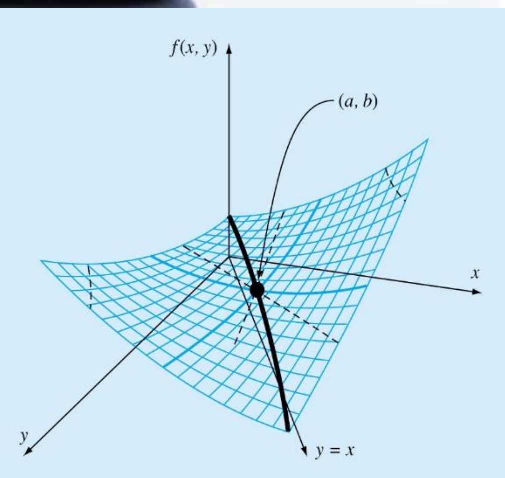
*First derivative provides*

*(a) the steepest trajectory of the function*

*(b) tells us that we have reached the maximum.*

*Second derivative tells us that whether we are a maximum or minimum.*

For *two dimensional functions* whether a maximum or a minimum occurs involves not only the partial derivatives w.r.t. $x$ and $y$ but also the second partials w.r.t. $x$ and $y$.




Assuming that the partial derivatives are continuous at and near the point being evaluated

$$|H| = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$$

$$\text{If } |\text{H}| \succ 0 \text{ and } \frac{\partial^2 f}{\partial x^2} \succ 0\text{, then f(x, y) has a local minimum}$$

$$\text{If } |\text{H}| \succ 0 \text{ and } \frac{\partial^2 f}{\partial x^2} \prec 0\text{, then f(x, y) has a local maximum}$$

$$\text{If } |\text{H}| \prec 0\text{, then f(x, y) has a saddle point}$$

The quantity [H] is equal to the determinant of a matrix made up of second derivatives

# Finite-Difference Approximations

$$\frac{\partial f}{\partial x} = \frac{f(x+\delta x, y) - f(x-\delta x, y)}{2\delta x}$$

$$\frac{\partial f}{\partial y} = \frac{f(x, y+\delta y) - f(x, y-\delta y)}{2\delta y}$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{f(x+\delta x, y) - 2\,f(x, y) + f(x-\delta x, y)}{\delta x^2}$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{f(x, y+\delta y) - 2\,f(x, y) + f(x, y-\delta y)}{\delta y^2}$$

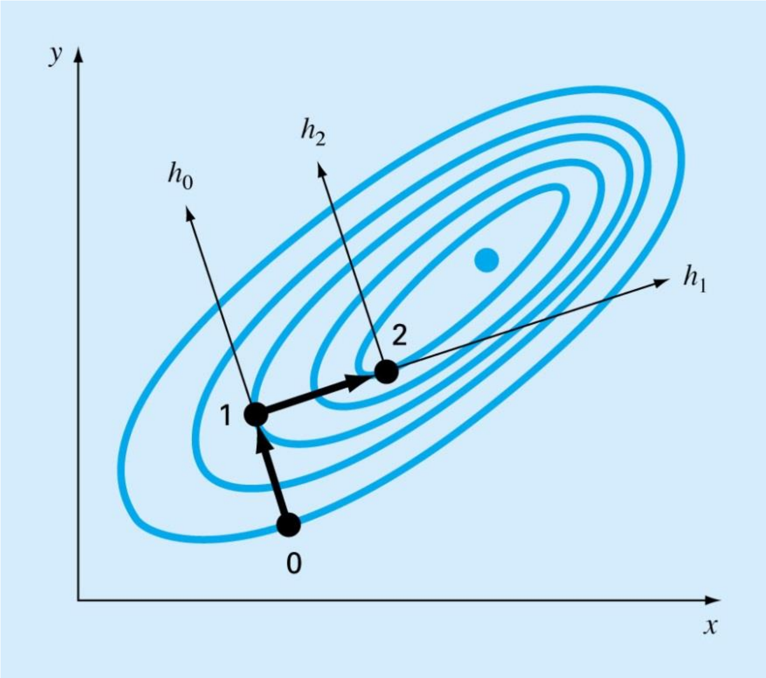
$$\frac{\partial^2 f}{\partial x \partial y} = \frac{f(x+\delta x, y+\delta y) - f(x+\delta x, y-\delta y) - f(x-\delta x, y+\delta y) + f(x-\delta x, y-\delta y)}{4\delta x \delta y}$$

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

# The Steepest Ascend Method

Start at an initial point $(x_o,y_o)$, determine the direction of steepest ascend, that is, the gradient.

Then search along the direction of the gradient, $h_o$, until we find maximum. Process is then repeated.



The problem has two parts

*Determining the "best direction" and*

*Determining the "best value" along that search direction.*

Steepest ascent method uses the gradient approach as its choice for the "best" direction.

To transform a function of *x* and *y* into a function of *h* along the gradient section:

$$x = x_o + \frac{\partial f}{\partial x} h$$

$$y = y_o + \frac{\partial f}{\partial y} h$$

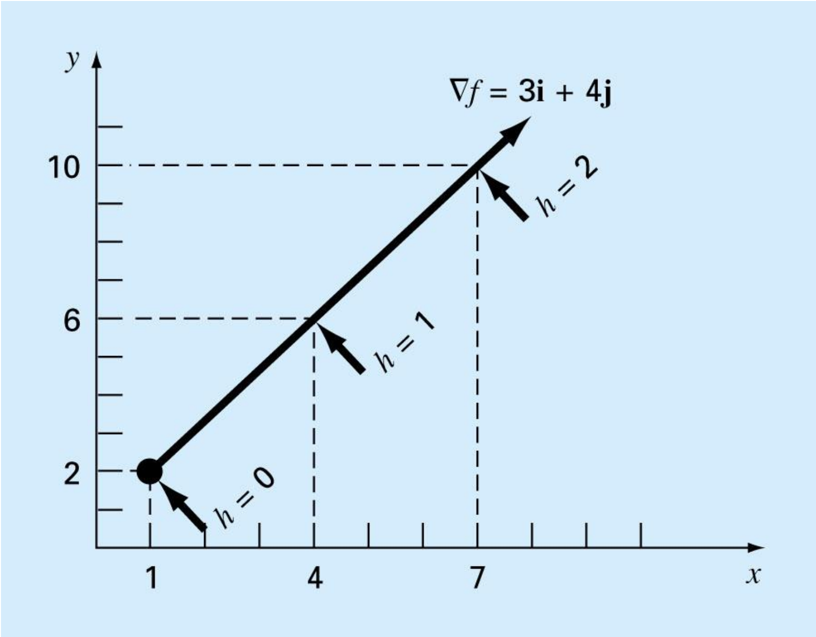
*h* is distance along the h axis

If $x_o$=1 and $y_o$=2

$$\nabla f = 3\mathrm{i} + 4\mathrm{j}$$

$$\mathrm{x} = 1 + 3\mathrm{h}$$

$$\mathrm{y} = 2 + 4\mathrm{h}$$



# Example #5

Suppose we have the following two-dimensional function:

$$f(x, y) = 2xy + 2x - x^2 - 2y^2$$

Develop a one-dimensional version of this equation along the gradient direction at point x=−1 and y = 1.

The partial derivatives can be evaluated at (−1, 1),

$$\frac{\partial f}{\partial x} = 2y + 2 - 2x = 2(1) + 2 - 2(-1) = 6$$

$$\frac{\partial f}{\partial y} = 2x - 4y = 2(-1) - 4(1) = -6$$

Therefore, the gradient vector is

$$\nabla f = 6i - 6j$$

The function can be expressed along this axis as

$$f\left(x_0 + \frac{\partial f}{\partial x}h, y_0 + \frac{\partial f}{\partial y}h\right) = f(-1 + 6h, 1 - 6h)$$

$$= 2(-1 + 6h)(1 - 6h) + 2(-1 + 6h) - (-1 + 6h)^2 - 2(1 - 6h)^2$$

By combining terms, we develop a one-dimensional function g(h) that maps f (x, y) along the h axis,

$$g(h) = -180h^2 + 72h - 7$$

# Example #6

Maximize the following two-dimensional function:

$$f(x, y) = 2xy + 2x - x^2 - 2y^2$$

Using initial guess, x=−1 and y = 1.

<u>**Method #1**</u>: Calculating directly

To do this, the partial derivatives can be evaluated as

$$\frac{\partial f}{\partial x} = 2y + 2 - 2x = 0$$

$$\frac{\partial f}{\partial y} = 2x - 4y = 0$$

This pair of equations can be solved for the optimum, x = 2 and y = 1. The second partial derivatives can also be determined and evaluated at the optimum.

$$\frac{\partial^2 f}{\partial x^2} = -2$$

$$\frac{\partial^2 f}{\partial y^2} = -4$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 2$$

and the determinant of the Hessian is computed

$$|H| = -2(-4) - 22 = 4$$

Therefore, because |H| > 0 and ∂2 f/∂x2 < 0, function value f(2, 1) is a maximum.

**Method #2:** Using Steepest Ascent Method

$$g(h) = -180h^2 + 72h - 7$$

Now, because this is a simple parabola, we can directly locate the maximum (that is, h = h∗) by solving the problem,

$$g'(h^*) = -360h* + 72 = 0$$
$$h^* = 0.2$$

This means that if we travel along the h axis, g(h) reaches a minimum value when h = h∗ = 0.2.

The (x, y) coordinates corresponding to this point,

$$x = -1 + 6(0.2) = 0.2$$
$$y = 1 - 6(0.2) = -0.2$$

The second step is merely implemented by repeating the procedure. First, the partial derivatives can be evaluated at the new starting point (0.2, −0.2) to give

$$\frac{\partial f}{\partial x} = 2(-0.2) + 2 - 2(0.2) = 1.2$$

$$\frac{\partial f}{\partial y} = 2(0.2) - 4(-0.2) = 1.2$$

Therefore, the gradient vector is

$$\nabla f = 1.2\,i + 1.2\,j$$

The coordinates along this new h axis can now be expressed as

$$x = 0.2 + 1.2h$$
$$y = -0.2 + 1.2h$$

Substituting these values into the function yields

$$f(0.2 + 1.2h, -0.2 + 1.2h) = g(h) = -1.44h2 + 2.88h + 0.2$$

The step h∗ to take us to the maximum along the search direction can then be directly computed as

$$g'(h*) = -2.88h* + 2.88 = 0$$
$$h* = 1$$

The (x, y) coordinates corresponding to this new point,

$$x = 0.2 + 1.2(1) = 1.4$$
$$y = -0.2 + 1.2(1) = 1$$

The approach can be repeated with the final result converging on the analytical solution, x = 2 and y = 1.

PART C

# CONSTRAINED OPTIMIZATION

# LINEAR PROGRAMMING

An optimization approach that deals with meeting a desired objective such as maximizing profit or minimizing cost in presence of constraints such as limited resources

Mathematical functions representing both the objective and the constraints are linear.

# Standard Form

Basic linear programming problem consists of two major parts:

The objective function

A set of constraints

For maximization problem, the objective function is generally expressed as

$$\text{Maximize } Z = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

$c_j$= payoff of each unit of the *j*th activity that is undertaken

$x_j$= magnitude of the *j*th activity

$Z$= total payoff due to the total number of activities

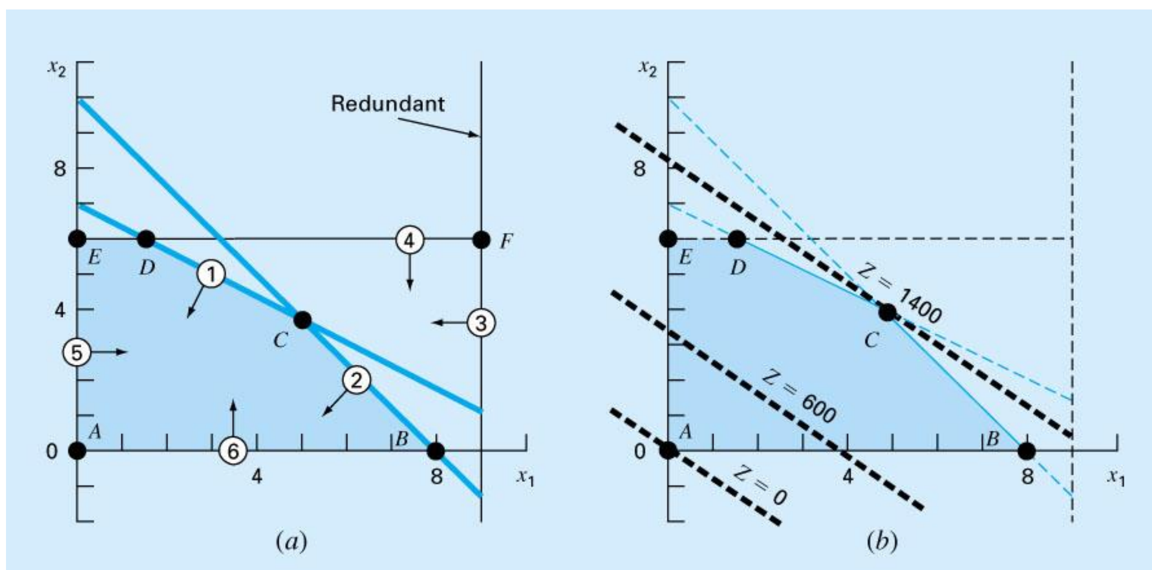
The constraints can be represented generally as

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i$$

Where $a_{ij}$=amount of the ith resource that is consumed for each unit of the *j*th activity and $b_i$=amount of the *i*th resource that is available

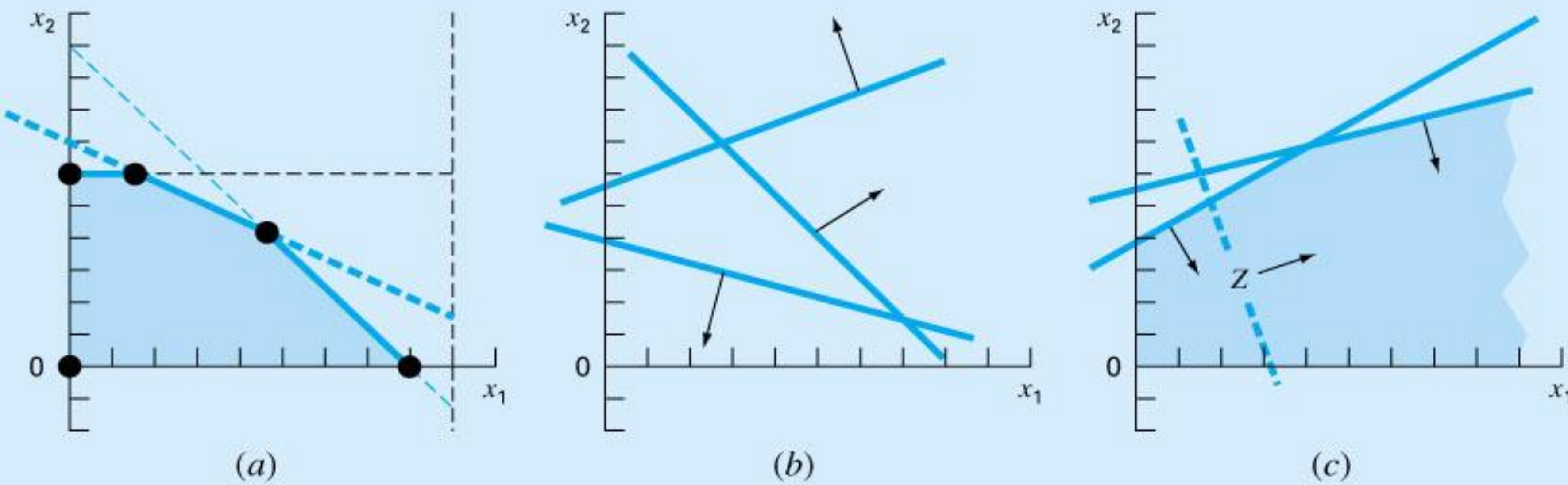
The general second type of constraint specifies that all activities must have a positive value, $x_i > 0$ .

Together, the objective function and the constraints specify the linear programming problem.



Possible outcomes that can be generally obtained in a linear programming problem

1. *Unique solution*. The maximum objective function intersects a single point.
2. *Alternate solutions*. Problem has an infinite number of optima corresponding to a line segment.
3. *No feasible solution*.
4. *Unbounded problems*. Problem is under-constrained and therefore open-ended.



# Any Questions?

✉ hvusynh@hcmiu.edu.vn