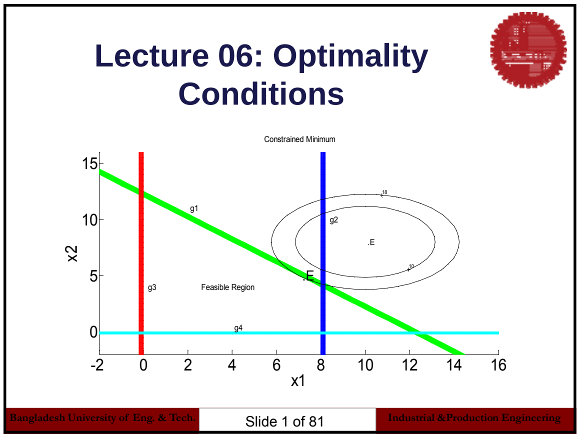
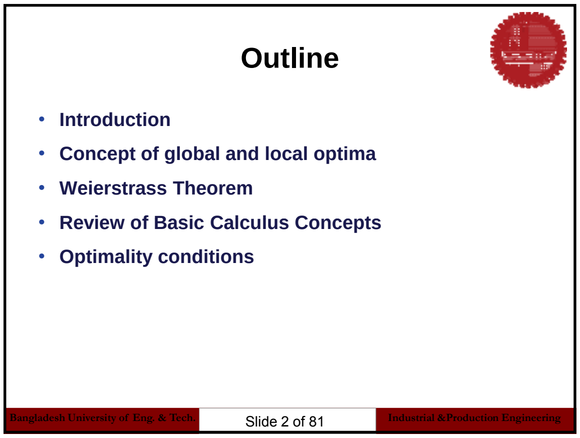
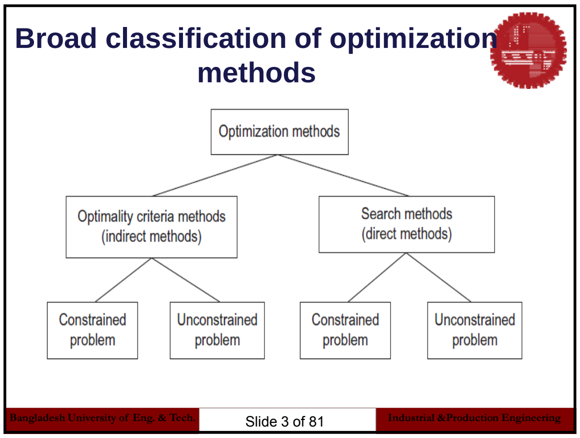


Lecture 06: Optimality Conditions



Outline

- **Introduction**
- **Concept of global and local optima**
- **Weierstrass Theorem**
- **Review of Basic Calculus Concepts**
- **Optimality conditions**



Broad classification of optimization methods

Optimization methods

Optimality criteria methods (indirect methods)

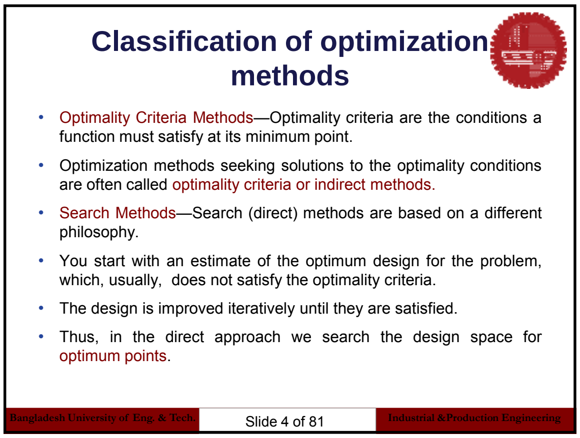
Search methods (direct methods)

Constrained problem

Unconstrained problem

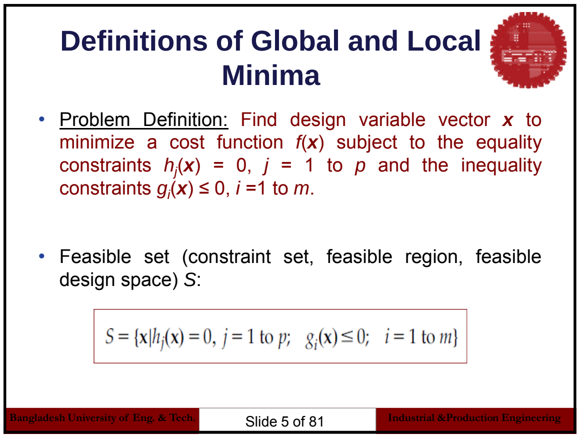
Constrained problem

Unconstrained problem



Classification of optimization methods

- Optimality Criteria Methods—Optimality criteria are the conditions a function must satisfy at its minimum point.
- Optimization methods seeking solutions to the optimality conditions are often called optimality criteria or indirect methods.
- Search Methods—Search (direct) methods are based on a different philosophy.
- You start with an estimate of the optimum design for the problem, which, usually, does not satisfy the optimality criteria.
- The design is improved iteratively until they are satisfied.
- Thus, in the direct approach we search the design space for optimum points.

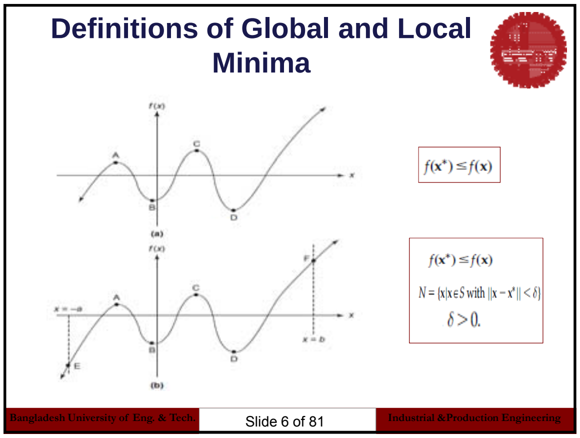


Definitions of Global and Local Minima

- Problem Definition: Find design variable vector $\mathbf{x}$ to minimize a cost function $f(\mathbf{x})$ subject to the equality constraints $h_j(\mathbf{x}) = 0$, $j = 1$ to p and the inequality constraints $g_i(\mathbf{x}) \leq 0$, $i = 1$ to m.

- Feasible set (constraint set, feasible region, feasible design space) S:

$$S = \{\mathrm{x} \mid h_j(\mathrm{x}) = 0,\ j = 1 \text{ to } p;\quad g_i(\mathrm{x}) \leq 0;\quad i = 1 \text{ to } m\}$$



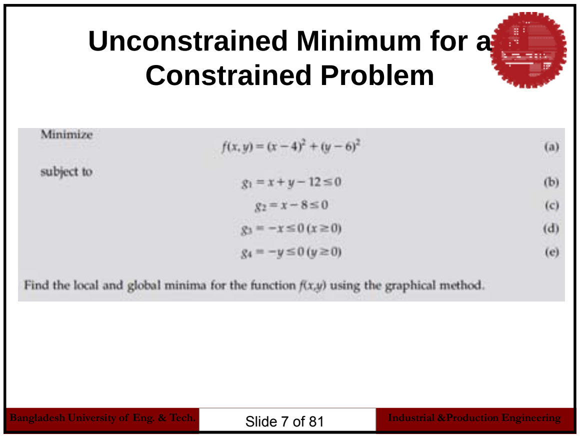
Definitions of Global and Local Minima

$$f(\mathbf{x}^*) \leq f(\mathbf{x})$$

$$f(\mathbf{x}^*) \leq f(\mathbf{x})$$

$$N = \{\mathrm{x} \mid \mathrm{x} \in S \text{ with } \|\mathrm{x} - \mathrm{x}^*\| < \delta\}$$

$$\delta > 0.$$



Unconstrained Minimum for a Constrained Problem

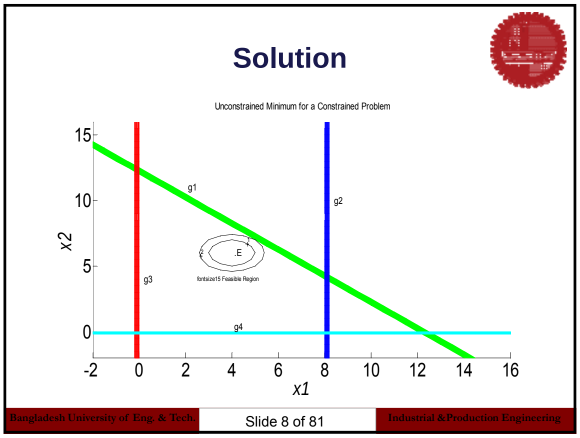
Minimize

$$f(x,y) = (x-4)^2 + (y-6)^2 \quad \text{(a)}$$

subject to

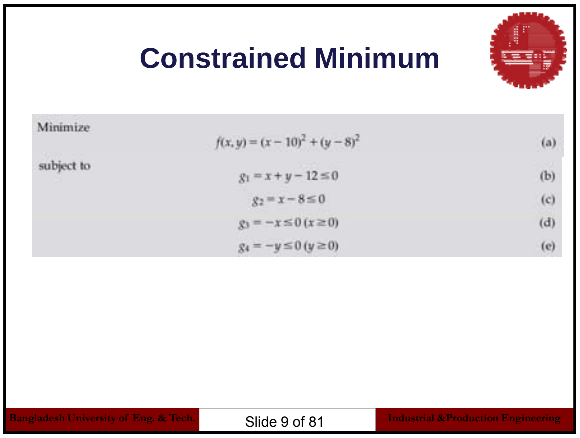
$$g_1 = x + y - 12 \le 0 \quad \text{(b)}$$
$$g_2 = x - 8 \le 0 \quad \text{(c)}$$
$$g_3 = -x \le 0 \, (x \ge 0) \quad \text{(d)}$$
$$g_4 = -y \le 0 \, (y \ge 0) \quad \text{(e)}$$

Find the local and global minima for the function $f(x,y)$ using the graphical method.



Solution

Unconstrained Minimum for a Constrained Problem



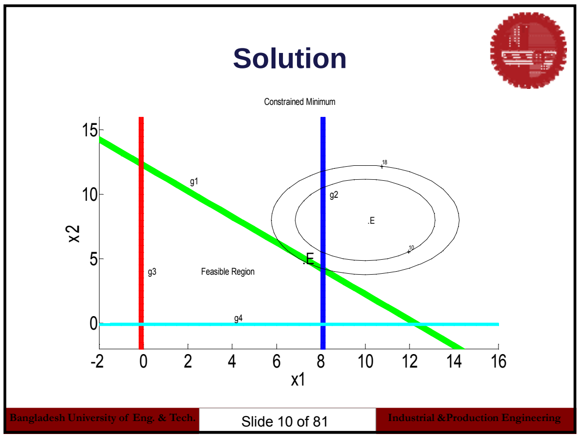
Constrained Minimum

Minimize

$$f(x,y) = (x-10)^2 + (y-8)^2 \quad \text{(a)}$$

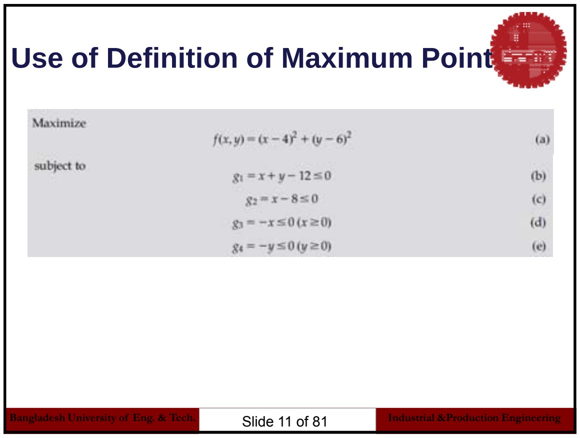
subject to

$$g_1 = x + y - 12 \le 0 \quad \text{(b)}$$
$$g_2 = x - 8 \le 0 \quad \text{(c)}$$
$$g_3 = -x \le 0 \, (x \ge 0) \quad \text{(d)}$$
$$g_4 = -y \le 0 \, (y \ge 0) \quad \text{(e)}$$



Solution

Constrained Minimum



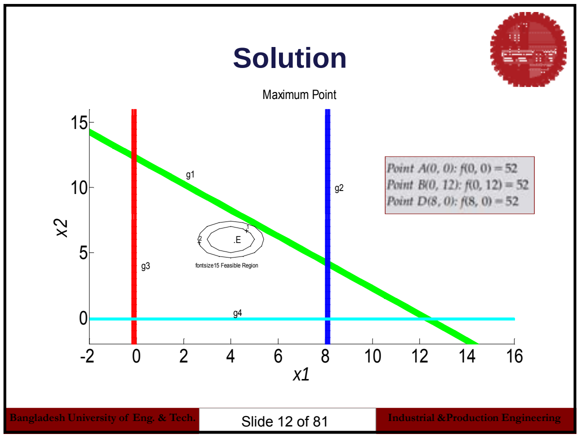
Use of Definition of Maximum Point

Maximize

$$f(x,y) = (x-4)^2 + (y-6)^2 \quad \text{(a)}$$

subject to

$$g_1 = x + y - 12 \le 0 \quad \text{(b)}$$
$$g_2 = x - 8 \le 0 \quad \text{(c)}$$
$$g_3 = -x \le 0 \, (x \ge 0) \quad \text{(d)}$$
$$g_4 = -y \le 0 \, (y \ge 0) \quad \text{(e)}$$



Solution

Maximum Point

Point $A(0, 0)$: $f(0, 0) = 52$
Point $B(0, 12)$: $f(0, 12) = 52$
Point $D(8, 0)$: $f(8, 0) = 52$

Existence of a Minimum

- Weierstrass Theorem—Existence of a Global Minimum

  If $f(\mathbf{x})$ is continuous on a nonempty feasible set S that is closed and bounded, then $f(\mathbf{x})$ has a global minimum in S.

- Closed and bounded set:
  - A set S is closed if it includes all of its boundary points and every sequence of points has a subsequence that converges to a point in the set.
  - A set is bounded if for any point, $\mathbf{x} \in S$, $\mathbf{x}^T\mathbf{x} < c$, where c is a finite number.

Existence of a Global Minimum using Weierstrass Theorem

Consider a function $f(x) = -1/x$ defined on the set $S = \{x \mid 0 < x \le 1\}$. Check the existence of a global minimum for the function.

- If Weierstrass theorem is satisfied, the existence of a global optimum is guaranteed.
- However, when it is not satisfied, a global solution may still exist; i.e., it is not an “if-and-only-if” theorem.
- Minimize $f(x) = x^2$

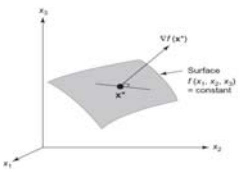
  subject to $-1 < x < 1$.

Review of Basic Calculus Concepts

- Gradient Vector: Partial Derivatives of a Function

$$\mathbf{c} = \nabla f(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial f(\mathbf{x}^*)}{\partial x_1} \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_n} \end{bmatrix} = \left[\frac{\partial f(\mathbf{x}^*)}{\partial x_1} \frac{\partial f(\mathbf{x}^*)}{\partial x_2} \cdots \frac{\partial f(\mathbf{x}^*)}{\partial x_n}\right]^T$$

  - Geometrically, the gradient vector is normal to the tangent plane at the point $\mathbf{x}^*$.
  - Also, it points in the direction of maximum increase in the function.



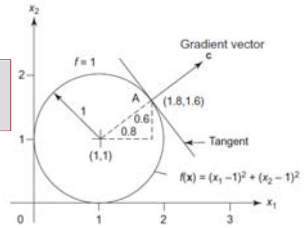
Review of Basic Calculus Concepts

EXAMPLE 4.5 CALCULATION OF A GRADIENT VECTOR

Calculate the gradient vector for the function $f(\mathbf{x}) = (x_1 - 1)^2 + (x_2 - 1)^2$ at the point $\mathbf{x}^* = (1.8, 1.6)$.

$$\frac{\partial f}{\partial x_1}(1.8, 1.6) = 2(x_1 - 1) = 2(1.8 - 1) = 1.6$$

$$\frac{\partial f}{\partial x_2}(1.8, 1.6) = 2(x_2 - 1) = 2(1.6 - 1) = 1.2$$



Review of Basic Calculus Concepts

- Hessian Matrix: Second-Order Partial Derivatives

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- The Hessian is always a symmetric matrix.

Review of Basic Calculus Concepts

For the following function, calculate the gradient vector and the Hessian matrix at the point (1, 2):

$$f(\mathbf{x}) = x_1^3 + x_2^3 + 2x_1^2 + 3x_2^2 - x_1x_2 + 2x_1 + 4x_2 \qquad \text{(a)}$$

Solution

The first partial derivatives of the function are given as

$$\frac{\partial f}{\partial x_1} = 3x_1^2 + 4x_1 - x_2 + 2; \quad \frac{\partial f}{\partial x_2} = 3x_2^2 + 6x_2 - x_1 + 4$$

Substituting the point $x_1 = 1$, $x_2 = 2$, the gradient vector is given as $\mathbf{c} = (7, 27)$.

The second partial derivatives of the function are calculated as

$$\frac{\partial^2 f}{\partial x_1^2} = 6x_1 + 4; \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -1; \quad \frac{\partial^2 f}{\partial x_2 \partial x_1} = -1; \quad \frac{\partial^2 f}{\partial x_2^2} = 6x_2 + 6.$$

Therefore, the Hessian matrix is given as

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} 6x_1 + 4 & -1 \\ -1 & 6x_2 + 6 \end{bmatrix}$$

The Hessian matrix at the point (1, 2) is given as

$$\mathbf{H}(1, 2) = \begin{bmatrix} 10 & -1 \\ -1 & 18 \end{bmatrix}$$

Review of Basic Calculus Concepts

- <u>Taylor's Expansion</u>
  - A function can be approximated by polynomials in a neighborhood of any point in terms of its value and derivatives using Taylor's expansion.
  - <u>For a function of one variable</u>:

$$f(x) = f(x^*) + \frac{df(x^*)}{dx}(x - x^*) + \frac{1}{2}\frac{d^2f(x^*)}{dx^2}(x - x^*)^2 + R$$

$$f(x^* + d) = f(x^*) + \frac{df(x^*)}{dx}d + \frac{1}{2}\frac{d^2f(x^*)}{dx^2}d^2 + R$$

  - <u>For a function of two variables</u>:

$$f(x_1, x_2) = f(x_1^*, x_2^*) + \frac{\partial f}{\partial x_1}d_1 + \frac{\partial f}{\partial x_2}d_2 + \frac{1}{2}\left[\frac{\partial^2 f}{\partial x_1^2}d_1^2 + 2\frac{\partial^2 f}{\partial x_1 \partial x_2}d_1 d_2 + \frac{\partial^2 f}{\partial x_2^2}d_2^2\right]$$

$$f(x_1, x_2) = f(x_1^*, x_2^*) + \sum_{i=1}^{2}\frac{\partial f}{\partial x_i}d_i + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial^2 f}{\partial x_i \partial x_j}d_i d_j$$

Review of Basic Calculus Concepts

- Taylor's expansion in matrix notation:

$$f(x_1, x_2) = f(x_1^*, x_2^*) + \sum_{i=1}^{2}\frac{\partial f}{\partial x_i}d_i + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial^2 f}{\partial x_i \partial x_j}d_i d_j$$

$$\boxed{f(\mathbf{x}^* + \mathbf{d}) = f(\mathbf{x}^*) + \nabla f^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{H}\mathbf{d} + R}$$

where $\mathbf{x} = (x_1, x_2)$, $\mathbf{x}^* = (x_1^*, x_2^*)$, $\mathbf{x} - \mathbf{x}^* = \mathbf{d}$, and $\mathbf{H}$ is the 2×2 Hessian matrix.

$$\Delta f = \nabla f^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{H}\mathbf{d} + R$$

Review of Basic Calculus Concepts

- Taylor's expansion in matrix notation:

$$\Delta f = \nabla f^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{H}\mathbf{d} + R$$

$$\boxed{\delta f = \nabla f^T\delta\mathbf{x} = \nabla f \cdot \delta\mathbf{x}}$$

Example: Taylor's Expansion of a Function of One Variable

Approximate $f(x) = \cos x$ around the point $x^* = 0$.

Solution

Derivatives of the function $f(x)$ are given as

$$\frac{df}{dx} = -\sin x, \quad \frac{d^2f}{dx^2} = -\cos x \qquad \text{(a)}$$

Therefore, using Eq. (4.9), the second-order Taylor's expansion for $\cos x$ at the point $x^* = 0$ is given as

$$\cos x \approx \cos 0 - \sin 0(x - 0) + \frac{1}{2}(-\cos 0)(x - 0)^2 = 1 - \frac{1}{2}x^2 \qquad \text{(b)}$$

Example: Taylor's Expansion of a Function of Two Variables

Obtain a second-order Taylor's expansion for the function $f(\mathrm{x}) = 3x_1^3x_2$ at the point $\mathrm{x}^* = (1, 1)$.

Solution

The gradient and Hessian of the function $f(\mathbf{x})$ at the point $\mathbf{x}^* = (1,1)$ using Eqs. (4.5) and (4.8) are

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 9x_1^2x_2 \\ 3x_1^3 \end{bmatrix} = \begin{bmatrix} 9 \\ 3 \end{bmatrix}; \quad \mathbf{H} = \begin{bmatrix} 18x_1x_2 & 9x_1^2 \\ 9x_1^2 & 0 \end{bmatrix} = \begin{bmatrix} 18 & 9 \\ 9 & 0 \end{bmatrix} \qquad \text{(a)}$$

Substituting these in the matrix form of Taylor's expression given in Eq. (4.12), and using $\mathbf{d} = \mathbf{x} - \mathbf{x}^*$, we obtain an approximation $\bar{f}(\mathbf{x})$ for $f(\mathbf{x})$ as

$$\bar{f}(\mathbf{x}) = 3 + \begin{bmatrix} 9 \\ 3 \end{bmatrix}^T \begin{bmatrix} (x_1 - 1) \\ (x_2 - 1) \end{bmatrix} + \frac{1}{2}\begin{bmatrix} (x_1 - 1) \\ (x_2 - 1) \end{bmatrix}^T \begin{bmatrix} 18 & 9 \\ 9 & 0 \end{bmatrix} \begin{bmatrix} (x_1 - 1) \\ (x_2 - 1) \end{bmatrix} \qquad \text{(b)}$$

where $f(\mathbf{x}^*) = 3$ has been used. Simplifying the expression by expanding vector and matrix products, we obtain Taylor's expansion for $f(\mathbf{x})$ about the point (1, 1) as

$$\bar{f}(\mathbf{x}) = 9x_1^2 + 9x_1x_2 - 18x_1 - 6x_2 + 9 \qquad \text{(c)}$$

Example : A Linear Taylor's Expansion of a Function

Obtain a linear Taylor's expansion for the function

$$f(\mathbf{x}) = x_1^2 + x_2^2 - 4x_1 - 2x_2 + 4 \qquad \text{(a)}$$

at the point $\mathbf{x}^* = (1, 2)$. Compare the approximate function with the original function in a neighborhood of the point (1, 2).

Solution

The gradient of the function at the point (1, 2) is given as

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} (2x_1 - 4) \\ (2x_2 - 2) \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \qquad \text{(b)}$$

Since $f(1, 2) = 1$, Eq. (4.13) gives a linear Taylor's approximation for $f(\mathbf{x})$ as

$$\bar{f}(\mathbf{x}) = 1 + [-2\ 2]\begin{bmatrix} (x_1 - 1) \\ (x_2 - 2) \end{bmatrix} = -2x_1 + 2x_2 - 1 \qquad \text{(c)}$$

To see how accurately $\bar{f}(\mathbf{x})$ approximates the original $f(\mathbf{x})$ in the neighborhood of (1, 2), we calculate the functions at the point (1.1, 2.2), a 10 percent change in the point as $\bar{f}(\mathbf{x}) = 1.20$ and $f(\mathbf{x}) = 1.25$. We see that the approximate function underestimates the real function by 4 percent. An error of this magnitude is quite acceptable in many applications. Note, however, that the errors will be different for different functions and can be larger for highly nonlinear functions.

Review of Basic Calculus Concepts

- Quadratic Form

$$F(\mathbf{x}) = x_1^2 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 2x_2x_3 + 2x_3x_1$$

$$\boxed{F(\mathbf{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}x_ix_j}$$

$$f(x_1, x_2) = f(x_1^*, x_2^*) + \sum_{i=1}^{2}\frac{\partial f}{\partial x_i}d_i + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial^2 f}{\partial x_i \partial x_j}d_id_j$$

Review of Basic Calculus Concepts

- The Matrix of the Quadratic Form

$$F(\mathrm{x}) = \mathrm{x}^{\mathrm{T}}\mathrm{Px}$$

- ***P*** (an *n*×*n* matrix*)* is called the matrix of the quadratic form *F*(***x***).
- There are many matrices associated with a quadratic form. However, there is only one symmetric matrix associated with it.

$$\mathbf{A} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^{\mathrm{T}}) \text{ or } a_{ij} = \frac{1}{2}(p_{ij} + p_{ji}),\ i, j = 1 \text{ to } n \qquad \boxed{F(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}}$$

Example: A Matrix of Quadratic Form

Identify a matrix associated with the quadratic form:

$$F(x_1, x_2, x_3) = 2x_1^2 + 2x_1x_2 + 4x_1x_3 - 6x_2^2 - 4x_2x_3 + 5x_3^2$$

Solution

Writing *F* in the matrix form ($F(\mathbf{x}) = \mathbf{x}^T\mathbf{P}\mathbf{x}$), we obtain

$$F(\mathbf{x}) = [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 2 & 2 & 4 \\ 0 & -6 & -4 \\ 0 & 0 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \text{(b)}$$

$$\mathbf{P} = \begin{bmatrix} 2 & 0.5 & 1 \\ 1.5 & -6 & -6 \\ 3 & 2 & 5 \end{bmatrix};\ \mathbf{P} = \begin{bmatrix} 2 & 4 & 5 \\ -2 & -6 & 4 \\ -1 & -8 & 5 \end{bmatrix} \qquad \text{(c)}$$

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & -6 & -2 \\ 2 & -2 & 5 \end{bmatrix} \qquad \text{(d)}$$

Review of Basic Calculus Concepts

- Form of a Matrix
- Quadratic form *F*(***x***) = ***x***^{T}A***x*** may be either positive, negative, or zero for any ***x***.

1. Positive Definite. *F*(***x***) > 0 for all x ≠ 0.
2. Positive Semidefinite. F(x) ≥ 0 for all x ≠ 0.
3. Negative Definite. F(x) < 0 for all x ≠ 0.
4. Negative Semidefinite. F(x) ≤ 0 for all x ≠ 0.
5. Indefinite. The quadratic form is called indefinite if it is positive for some values of ***x*** and negative for some others.

Example: Determination of Form of a Matrix

Determine the form of the following matrices:

$$(i)\ \mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad (ii)\ \mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Solution

The quadratic form associated with the matrix (*i*) is always positive because

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = (2x_1^2 + 4x_2^2 + 3x_3^2) > 0 \qquad \text{(b)}$$

unless $x_1 = x_2 = x_3 = 0$ ($\mathbf{x} = 0$). Thus, the matrix is positive definite.

The quadratic form associated with the matrix (*ii*) is negative semidefinite, since

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = (-x_1^2 - x_2^2 + 2x_1x_2 - x_3^2) = \{-x_3^2 - (x_1 - x_2)^2\} \le 0 \qquad \text{(c)}$$

for all x, and $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ when $x_3 = 0$, and $x_1 = x_2$ (e.g., $\mathbf{x} = (1, 1, 0)$). The quadratic form is not negative definite but is negative semidefinite since it can have a zero value for nonzero x. Therefore, the matrix associated with it is also negative semidefinite.

Methods for checking positive definiteness or semidefiniteness

Theorem

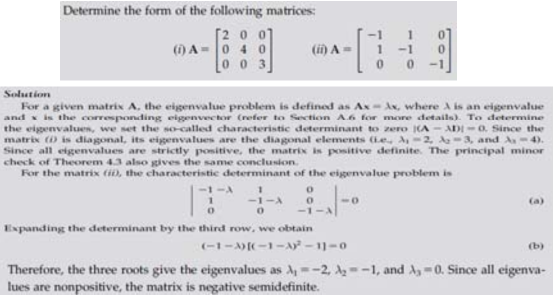
Eigenvalue Check for the Form of a Matrix Let λ_i, $i = 1$ to n be the eigenvalues of a symmetric $n \times n$ matrix **A** associated with the quadratic form $F(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ (since **A** is symmetric, all eigenvalues are real). The following results can be stated regarding the quadratic form *F*(x) or the matrix **A**:

1. *F*(x) is *positive definite if and only if* all eigenvalues of **A** are strictly positive; i.e., $\lambda_i > 0$, $i = 1$ to n.
2. *F*(x) is *positive semidefinite if and only if* all eigenvalues of **A** are non-negative; i.e., $\lambda_i \ge 0$, $i = 1$ to n (note that at least one eigenvalue must be zero for it to be called positive semidefinite).
3. *F*(x) is *negative definite if and only if* all eigenvalues of **A** are strictly negative; i.e., $\lambda_i < 0$, $i = 1$ to n.
4. *F*(x) is *negative semidefinite if and only if* all eigenvalues of **A** are nonpositive; i.e., $\lambda_i \le 0$, $i = 1$ to n (note that at least one eigenvalue must be zero for it to be called negative semidefinite).
5. *F*(x) is *indefinite if* some $\lambda_i < 0$ and some other $\lambda_j > 0$.

Use of Quadratic Forms

- The theory of quadratic forms is used in the second-order conditions for a local optimum point.
- Also, it is used to determine the convexity of functions of the optimization problem.

Example: Determination of Form of a Matrix



Determine the form of the following matrices:

$$(i)\ \mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad (ii)\ \mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Solution

For a given matrix $\mathbf{A}$, the eigenvalue problem is defined as $\mathbf{Ax} = \lambda\mathbf{x}$, where λ is an eigenvalue and $\mathbf{x}$ is the corresponding eigenvector (refer to Section A.6 for more details). To determine the eigenvalues, we set the so-called characteristic determinant to zero $|(\mathbf{A} - \lambda\mathbf{I})| = 0$. Since the matrix *(i)* is diagonal, its eigenvalues are the diagonal elements (i.e., $\lambda_1 = 2$, $\lambda_2 = 3$, and $\lambda_3 = 4$). Since all eigenvalues are strictly positive, the matrix is positive definite. The principal minor check of Theorem 4.3 also gives the same conclusion.

For the matrix *(ii)*, the characteristic determinant of the eigenvalue problem is

$$\begin{vmatrix} -1-\lambda & 1 & 0 \\ 1 & -1-\lambda & 0 \\ 0 & 0 & -1-\lambda \end{vmatrix} = 0 \qquad (a)$$

Expanding the determinant by the third row, we obtain

$$(-1-\lambda)[(-1-\lambda)^2 - 1] = 0 \qquad (b)$$

Therefore, the three roots give the eigenvalues as $\lambda_1 = -2$, $\lambda_2 = -1$, and $\lambda_3 = 0$. Since all eigenvalues are nonpositive, the matrix is negative semidefinite.

Differentiation of a Quadratic Form

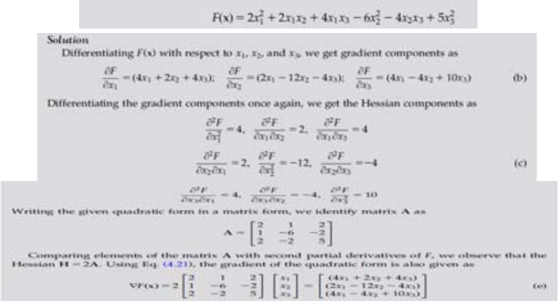
- Often we need to find the gradient and Hessian matrix for the quadratic form:

$$F(\mathbf{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}x_i x_j \qquad F(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$$

$$\frac{\partial F(\mathbf{x})}{\partial x_i} = 2\sum_{j=1}^{n} a_{ij}x_j; \quad \text{or} \quad \nabla F(\mathbf{x}) = 2\mathbf{Ax}$$

$$\frac{\partial^2 F(\mathbf{x})}{\partial x_j \partial x_i} = 2a_{ij}; \quad \text{or} \quad \mathbf{H} = 2\mathbf{A}$$

Example: Calculations for Gradient and Hessian of Quadratic Form



Calculate the gradient and Hessian of the following quadratic form:

$$F(\mathbf{x}) = 2x_1^2 + 2x_1x_2 + 4x_1x_3 - 6x_2^2 - 4x_2x_3 + 5x_3^2$$

Solution

Differentiating $F(\mathbf{x})$ with respect to x_1, x_2, and x_3, we get gradient components as

$$\frac{\partial F}{\partial x_1} = (4x_1 + 2x_2 + 4x_3); \quad \frac{\partial F}{\partial x_2} = (2x_1 - 12x_2 - 4x_3); \quad \frac{\partial F}{\partial x_3} = (4x_1 - 4x_2 + 10x_3) \qquad (b)$$

Differentiating the gradient components once again, we get the Hessian components as

$$\frac{\partial^2 F}{\partial x_1^2} = 4, \quad \frac{\partial^2 F}{\partial x_1 \partial x_2} = 2, \quad \frac{\partial^2 F}{\partial x_1 \partial x_3} = 4$$

$$\frac{\partial^2 F}{\partial x_2 \partial x_1} = 2, \quad \frac{\partial^2 F}{\partial x_2^2} = -12, \quad \frac{\partial^2 F}{\partial x_2 \partial x_3} = -4 \qquad (c)$$

$$\frac{\partial^2 F}{\partial x_3 \partial x_1} = 4, \quad \frac{\partial^2 F}{\partial x_3 \partial x_2} = -4, \quad \frac{\partial^2 F}{\partial x_3^2} = 10$$

Writing the given quadratic form in a matrix form, we identify matrix $\mathbf{A}$ as

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & -6 & -2 \\ 2 & -2 & 5 \end{bmatrix}$$

Comparing elements of the matrix $\mathbf{A}$ with second partial derivatives of F, we observe that the Hessian $\mathbf{H} = 2\mathbf{A}$. Using Eq. (4.21), the gradient of the quadratic form is also given as

$$\nabla F(\mathbf{x}) = 2\begin{bmatrix} 2 & 1 & 2 \\ 1 & -6 & -2 \\ 2 & -2 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (4x_1 + 2x_2 + 4x_3) \\ (2x_1 - 12x_2 - 4x_3) \\ (4x_1 - 4x_2 + 10x_3) \end{bmatrix} \qquad (e)$$

Concept of Necessary and Sufficient Conditions

- Necessary Conditions
  - The conditions that must be satisfied at the optimum point are called necessary.
  - Points satisfying the necessary conditions are called candidate optimum points.
- Sufficient Condition
  - If a candidate optimum point satisfies the sufficient condition, then it is indeed an optimum point.

Summary of Necessary and Sufficient Conditions

1. Optimum points must satisfy the necessary conditions. Points that do not satisfy them cannot be optimum.
2. A point satisfying the necessary conditions need not be optimum; that is, nonoptimum points may also satisfy the necessary conditions.
3. A candidate point satisfying a sufficient condition is indeed optimum.
4. If the sufficiency condition cannot be used or it is not satisfied, we may not be able to draw any conclusions about the optimality of the candidate point.

Optimality Conditions: Unconstrained Problem

"Minimize *f*(x) without any constraints on x."

The optimality conditions for unconstrained or constrained problems can be used in two ways:

1. They can be used to check whether a given point is a local optimum for the problem.
2. They can be solved for local optimum points.

Concepts Related to Optimality Conditions

- You are at a minimum point ***x**** and then examine its neighborhood to study properties of the function and its derivatives.

$$\Delta f = f(\mathbf{x}) - f(\mathbf{x}^*) \geq 0$$

- ***x*** is a new point in the neighborhood of ***x****.

Optimality Conditions for Functions of a Single Variable

- First-Order Necessary Condition

$$f(x) = f(x^*) + f'(x^*)d + \frac{1}{2}f''(x^*)d^2 + R$$

$$\Delta f(x) = f'(x^*)d + \frac{1}{2}f''(x^*)d^2 + R$$

$$\Delta f = f'(x^*)d. \longleftarrow$$

It must be non-negative

$$\boxed{f'(x^*) = 0}$$

Optimality Conditions for Functions of a Single Variable

- Stationary Points
  - Since the points satisfying *f'(**x**)* = 0 can be local minima or maxima, or neither minimum nor maximum (inflection points), they are called stationary points.
- Sufficient Condition

$$\Delta f(x) = f'(x^*)d + \frac{1}{2}f''(x^*)d^2 + R$$

$$\Delta f(x) = \frac{1}{2}f''(x^*)d^2 + R$$

- This term can be positive for all *d* ≠ 0 if $\boxed{f''(x^*)>0}$

Optimality Conditions for Functions of a Single Variable

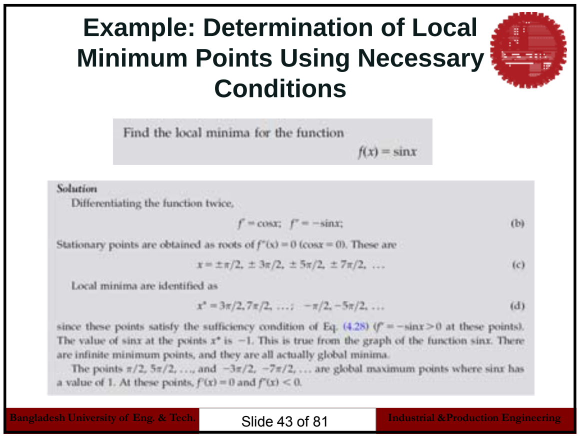
- Second-Order Necessary Condition
  - If *f*''(***x****) = 0, we cannot conclude that ***x**** is not a minimum point.

$$\boxed{f''(x^*) \geq 0}$$

- Any point violating it cannot be a local minimum.

Summary

1. The necessary conditions must be satisfied at the minimum point; otherwise, it cannot be a minimum.
2. The necessary conditions may also be satisfied by points that are not minima. A point satisfying the necessary conditions is simply a candidate local minimum.
3. If the sufficient condition is satisfied at a candidate point, then it is indeed a minimum point.



Example: Determination of Local Minimum Points Using Necessary Conditions

Find the local minima for the function

$$f(x) = \sin x$$

Solution

Differentiating the function twice,

$$f' = \cos x; \quad f'' = -\sin x; \qquad \text{(b)}$$

Stationary points are obtained as roots of $f'(x) = 0$ ($\cos x = 0$). These are

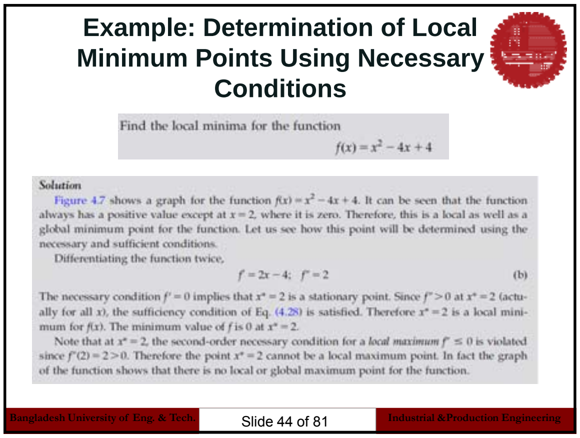
$$x = \pm\pi/2, \pm 3\pi/2, \pm 5\pi/2, \pm 7\pi/2, \ldots \qquad \text{(c)}$$

Local minima are identified as

$$x^* = 3\pi/2, 7\pi/2, \ldots; \quad -\pi/2, -5\pi/2, \ldots \qquad \text{(d)}$$

since these points satisfy the sufficiency condition of Eq. (4.28) ($f'' = -\sin x > 0$ at these points). The value of $\sin x$ at the points x^* is -1. This is true from the graph of the function $\sin x$. There are infinite minimum points, and they are all actually global minima.

The points $\pi/2, 5\pi/2, \ldots,$ and $-3\pi/2, -7\pi/2, \ldots$ are global maximum points where $\sin x$ has a value of 1. At these points, $f'(x) = 0$ and $f''(x) < 0$.



Example: Determination of Local Minimum Points Using Necessary Conditions

Find the local minima for the function

$$f(x) = x^2 - 4x + 4$$

Solution

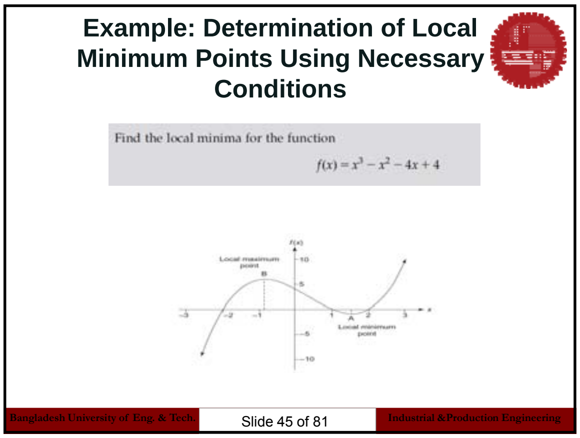
Figure 4.7 shows a graph for the function $f(x) = x^2 - 4x + 4$. It can be seen that the function always has a positive value except at $x = 2$, where it is zero. Therefore, this is a local as well as a global minimum point for the function. Let us see how this point will be determined using the necessary and sufficient conditions.

Differentiating the function twice,

$$f' = 2x - 4; \quad f'' = 2 \qquad \text{(b)}$$

The necessary condition $f' = 0$ implies that $x^* = 2$ is a stationary point. Since $f'' > 0$ at $x^* = 2$ (actually for all x), the sufficiency condition of Eq. (4.28) is satisfied. Therefore $x^* = 2$ is a local minimum for $f(x)$. The minimum value of f is 0 at $x^* = 2$.

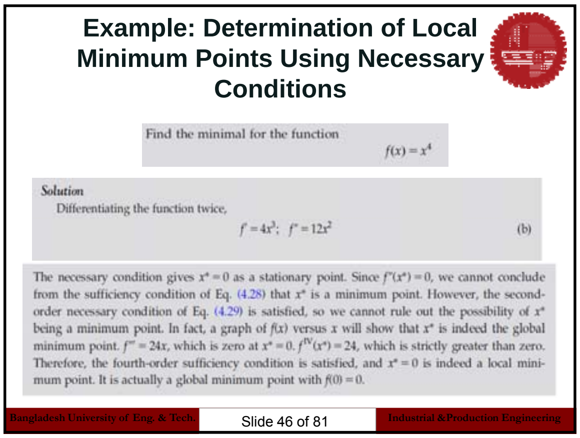
Note that at $x^* = 2$, the second-order necessary condition for a *local maximum* $f'' \le 0$ is violated since $f''(2) = 2 > 0$. Therefore the point $x^* = 2$ cannot be a local maximum point. In fact the graph of the function shows that there is no local or global maximum point for the function.



Example: Determination of Local Minimum Points Using Necessary Conditions

Find the local minima for the function

$$f(x) = x^3 - x^2 - 4x + 4$$



Example: Determination of Local Minimum Points Using Necessary Conditions

Find the minimal for the function

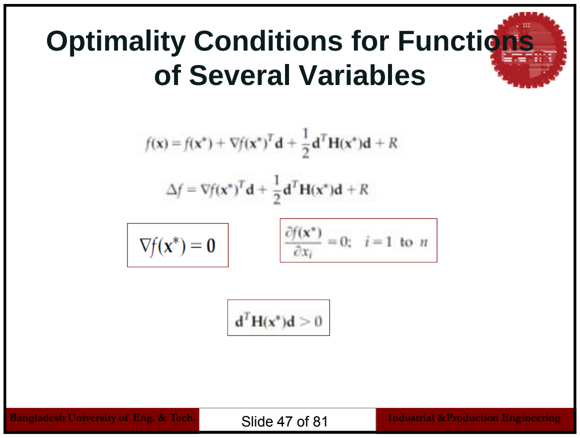
$$f(x) = x^4$$

Solution

Differentiating the function twice,

$$f' = 4x^3; \quad f'' = 12x^2 \qquad \text{(b)}$$

The necessary condition gives $x^* = 0$ as a stationary point. Since $f''(x^*) = 0$, we cannot conclude from the sufficiency condition of Eq. (4.28) that x^* is a minimum point. However, the second-order necessary condition of Eq. (4.29) is satisfied, so we cannot rule out the possibility of x^* being a minimum point. In fact, a graph of $f(x)$ versus x will show that x^* is indeed the global minimum point. $f''' = 24x$, which is zero at $x^* = 0$. $f^{IV}(x^*) = 24$, which is strictly greater than zero. Therefore, the fourth-order sufficiency condition is satisfied, and $x^* = 0$ is indeed a local minimum point. It is actually a global minimum point with $f(0) = 0$.



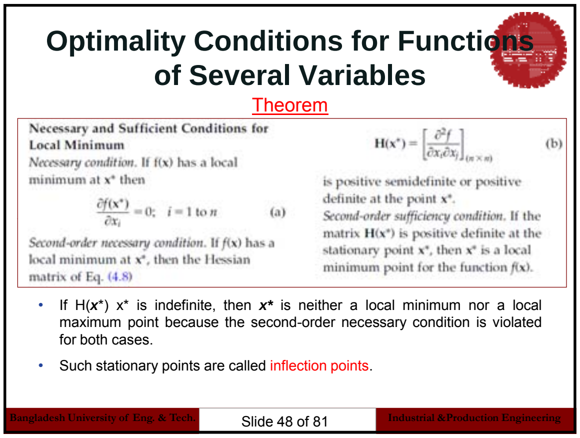
Optimality Conditions for Functions of Several Variables

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T \mathbf{d} + \frac{1}{2}\mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} + R$$

$$\Delta f = \nabla f(\mathbf{x}^*)^T \mathbf{d} + \frac{1}{2}\mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} + R$$

$$\boxed{\nabla f(\mathbf{x}^*) = 0} \qquad \boxed{\frac{\partial f(\mathbf{x}^*)}{\partial x_i} = 0; \quad i = 1 \text{ to } n}$$

$$\boxed{\mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} > 0}$$



Optimality Conditions for Functions of Several Variables

Theorem

Necessary and Sufficient Conditions for Local Minimum

Necessary condition. If f(x) has a local minimum at $\mathbf{x}^*$ then

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_i} = 0; \quad i = 1 \text{ to } n \qquad \text{(a)}$$

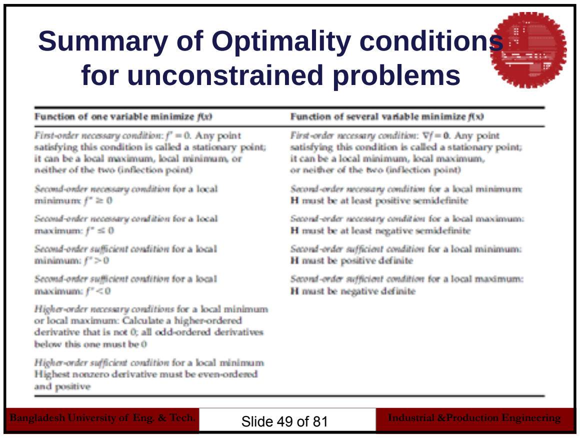
Second-order necessary condition. If $f(\mathbf{x})$ has a local minimum at $\mathbf{x}^*$, then the Hessian matrix of Eq. (4.8)

$$\mathbf{H}(\mathbf{x}^*) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{(n \times n)} \qquad \text{(b)}$$

is positive semidefinite or positive definite at the point $\mathbf{x}^*$.

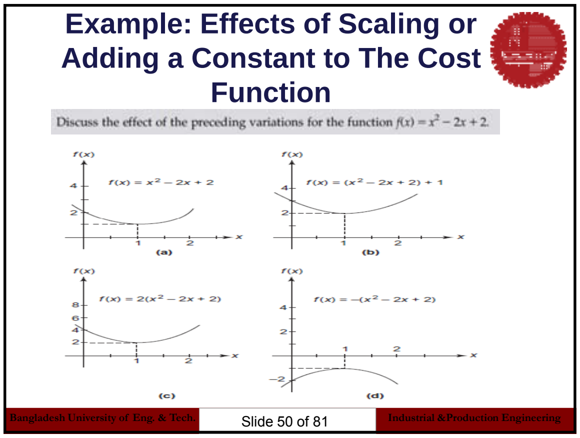
Second-order sufficiency condition. If the matrix $\mathbf{H}(\mathbf{x}^*)$ is positive definite at the stationary point $\mathbf{x}^*$, then $\mathbf{x}^*$ is a local minimum point for the function $f(\mathbf{x})$.

- If H(***x****) x* is indefinite, then ***x**** is neither a local minimum nor a local maximum point because the second-order necessary condition is violated for both cases.
- Such stationary points are called inflection points.



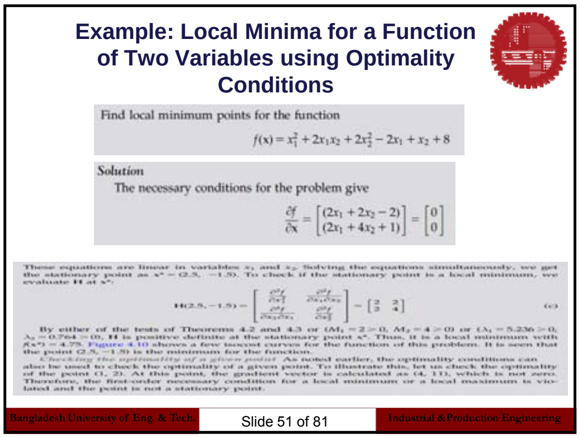
Summary of Optimality conditions for unconstrained problems

Function of one variable minimize $f(x)$	Function of several variable minimize $f(\mathbf{x})$
First-order necessary condition: $f' = 0$. Any point satisfying this condition is called a stationary point; it can be a local maximum, local minimum, or neither of the two (inflection point)	*First-order necessary condition:* $\nabla f = \mathbf{0}$. Any point satisfying this condition is called a stationary point; it can be a local minimum, local maximum, or neither of the two (inflection point)
Second-order necessary condition for a local minimum: $f'' \geq 0$	*Second-order necessary condition* for a local minimum: **H** must be at least positive semidefinite
Second-order necessary condition for a local maximum: $f'' \leq 0$	*Second-order necessary condition* for a local maximum: **H** must be at least negative semidefinite
Second-order sufficient condition for a local minimum: $f'' > 0$	*Second-order sufficient condition* for a local minimum: **H** must be positive definite
Second-order sufficient condition for a local maximum: $f'' < 0$	*Second-order sufficient condition* for a local maximum: **H** must be negative definite
Higher-order necessary conditions for a local minimum or local maximum: Calculate a higher-ordered derivative that is not 0; all odd-ordered derivatives below this one must be 0	
Higher-order sufficient condition for a local minimum Highest nonzero derivative must be even-ordered and positive	



Example: Effects of Scaling or Adding a Constant to The Cost Function

Discuss the effect of the preceding variations for the function $f(x) = x^2 - 2x + 2$.



Example: Local Minima for a Function of Two Variables using Optimality Conditions

Find local minimum points for the function

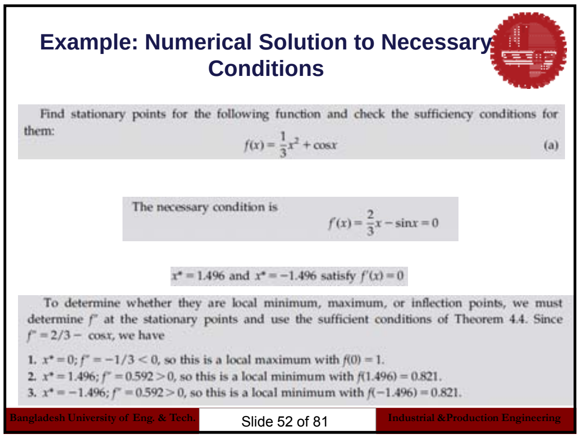
$$f(\mathbf{x}) = x_1^2 + 2x_1x_2 + 2x_2^2 - 2x_1 + x_2 + 8$$

Solution

The necessary conditions for the problem give

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} (2x_1 + 2x_2 - 2) \\ (2x_1 + 4x_2 + 1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

These equations are linear in variables x_1 and x_2. Solving the equations simultaneously, we get the stationary point as $\mathbf{x}^* = (2.5, -1.5)$. To check if the stationary point is a local minimum, we evaluate **H** at $\mathbf{x}^*$:



Example: Numerical Solution to Necessary Conditions

Find stationary points for the following function and check the sufficiency conditions for them:

$$f(x) = \frac{1}{3}x^2 + \cos x \qquad \text{(a)}$$

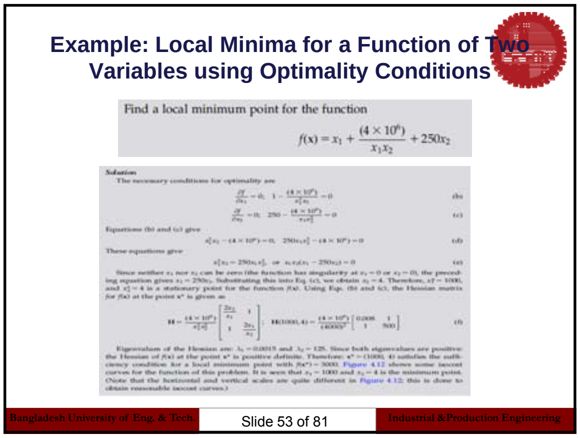
The necessary condition is

$$f'(x) = \frac{2}{3}x - \sin x = 0$$

$x^* = 1.496$ and $x^* = -1.496$ satisfy $f'(x) = 0$

To determine whether they are local minimum, maximum, or inflection points, we must determine f'' at the stationary points and use the sufficient conditions of Theorem 4.4. Since $f'' = 2/3 - \cos x$, we have

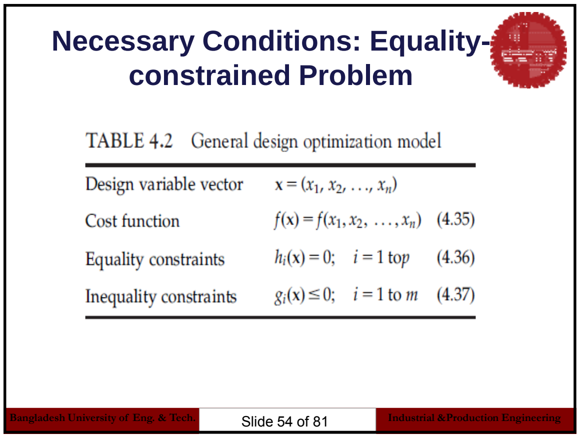
1. $x^* = 0$; $f'' = -1/3 < 0$, so this is a local maximum with $f(0) = 1$.
2. $x^* = 1.496$; $f'' = 0.592 > 0$, so this is a local minimum with $f(1.496) = 0.821$.
3. $x^* = -1.496$; $f'' = 0.592 > 0$, so this is a local minimum with $f(-1.496) = 0.821$.



Example: Local Minima for a Function of Two Variables using Optimality Conditions

Find a local minimum point for the function

$$f(\mathbf{x}) = x_1 + \frac{(4 \times 10^6)}{x_1x_2} + 250x_2$$



Necessary Conditions: Equality-constrained Problem

TABLE 4.2 General design optimization model

Design variable vector	$\mathbf{x} = (x_1, x_2, \ldots, x_n)$	
Cost function	$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$	(4.35)
Equality constraints	$h_i(\mathbf{x}) = 0;\quad i = 1 \text{ to} p$	(4.36)
Inequality constraints	$g_i(\mathbf{x}) \leq 0;\quad i = 1 \text{ to } m$	(4.37)

Lagrange Multipliers

**REGULAR POINT* Consider the constrained optimization problem of minimizing $f(\mathbf{x})$ subject to the constraints $h_i(\mathbf{x}) = 0$, $i = 1$ to p. A point $\mathbf{x}^*$ satisfying the constraints $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ is said to be a *regular point* of the feasible set if $f(\mathbf{x}^*)$ is differentiable and gradient vectors of all constraints at the point $\mathbf{x}^*$ are linearly independent. *Linear independence* means that no two gradients are parallel to each other, and no gradient can be expressed as a linear combination of the others (refer to Appendix A for more discussion on the linear independence of a set of vectors). When inequality constraints are also included in the problem definition, then for a point to be regular, gradients of all of the active constraints must also be linearly independent.

Lagrange Multipliers and their Geometrical Meaning

Minimize

$$f(x_1, x_2) = (x_1 - 1.5)^2 + (x_2 - 1.5)^2$$

subject to an equality constraint:

$$h(x_1, x_2) = x_1 + x_2 - 2 = 0$$

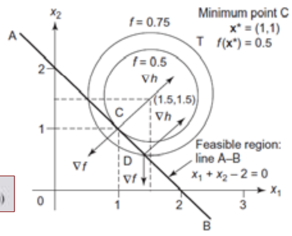
$$x_2 = \phi(x_1)$$

$$x_2 = \phi(x_1) = -x_1 + 2$$

Minimize

$$f(x_1, \phi(x_1))$$

$$f(x_1) = (x_1 - 1.5)^2 + (-x_1 + 2 - 1.5)^2$$



Lagrange Multipliers and their Geometrical Meaning

Minimize

$$f(x_1, x_2) = (x_1 - 1.5)^2 + (x_2 - 1.5)^2$$

subject to an equality constraint:

$$h(x_1, x_2) = x_1 + x_2 - 2 = 0$$

Minimize

$$f(x_1, \phi(x_1))$$

$$\frac{df(x_1, x_2)}{dx_1} = \frac{\partial f(x_1, x_2)}{\partial x_1} + \frac{\partial f(x_1, x_2)}{\partial x_2}\frac{dx_2}{dx_1} = 0$$

$$x_2 = \phi(x_1)$$

$$\frac{\partial f(x_1^*, x_2^*)}{\partial x_1} + \frac{\partial f(x_1^*, x_2^*)}{\partial x_2}\frac{d\phi}{dx_1} = 0$$

Lagrange Multipliers and their Geometrical Meaning

$$\frac{dh(x_1^*, x_2^*)}{dx_1} = \frac{\partial h(x_1^*, x_2^*)}{\partial x_1} + \frac{\partial h(x_1^*, x_2^*)}{\partial x_2}\frac{d\phi}{dx_1} = 0$$

$$\frac{d\phi}{dx_1} = -\frac{\partial h(x_1^*, x_2^*)/\partial x_1}{\partial h(x_1^*, x_2^*)/\partial x_2}$$

$$\frac{\partial f(x_1^*, x_2^*)}{\partial x_1} - \frac{\partial f(x_1^*, x_2^*)}{\partial x_2}\left(\frac{\partial h(x_1^*, x_2^*)/\partial x_1}{\partial h(x_1^*, x_2^*/\partial x_2)}\right) = 0$$

$$v = -\frac{\partial f(x_1^*, x_2^*)/\partial x_2}{\partial h\,(x_1^*, x_2^*)/\partial x_2}$$

$$\boxed{\frac{\partial f(x_1^*, x_2^*)}{\partial x_1} + v\frac{\partial h(x_1^*, x_2^*)}{\partial x_1} = 0} \qquad \boxed{\frac{\partial f(x_1^*, x_2^*)}{\partial x_2} + v\frac{\partial h(x_1^*, x_2^*)}{\partial x_2} = 0}$$

$$\boxed{h(x_1,\ x_2) = 0}$$

Necessary conditions

Lagrange Multipliers and their Geometrical Meaning

$$L(x_1, x_2, v) = f(x_1, x_2) + vh(x_1, x_2) \longrightarrow$$

Lagrangian/Lagrange function

$$\frac{\partial L(x_1^*, x_2^*)}{\partial x_1} = 0, \quad \frac{\partial L(x_1^*, x_2^*)}{\partial x_2} = 0$$

$$\boxed{\nabla f(\mathbf{x}^*) = -v\nabla h(\mathbf{x}^*)}$$

At the candidate minimum : gradients of the cost and constraint functions are proportional to each other, and the Lagrange multiplier *v* is the proportionality constant.

Lagrange multiplier for the equality constraint is free in sign; that is, the sign is determined by the form of the constraint function.

Lagrange Multiplier Theorem

Cost function $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$ (4.35)
Equality constraints $h_i(\mathbf{x}) = 0; \quad i = 1$ to p (4.36)

Lagrange Multiplier Theorem Consider the optimization problem defined in Eqs. (4.35) and (4.36):

Minimize $f(\mathbf{x})$

subject to equality constraints

$$h_i(\mathbf{x}) = 0, \quad i = 1 \text{ to } p$$

Let $\mathbf{x}^*$ be a regular point that is a local minimum for the problem. Then there exist unique Lagrange multipliers v_j^*, $j = 1$ to p such that

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_i} + \sum_{j=1}^{p} v_j^* \frac{\partial h_j(\mathbf{x}^*)}{\partial x_i} = 0; \quad i = 1 \text{ to } n \qquad (4.38)$$

$$h_j(\mathbf{x}^*) = 0; \quad j = 1 \text{ to } p \qquad (4.39)$$

It is convenient to write these conditions in terms of a *Lagrange function*, defined as

$$L(\mathbf{x}, \mathbf{v}) = f(\mathbf{x}) + \sum_{j=1}^{p} v_j h_j(\mathbf{x}) = f(\mathbf{x}) + \mathbf{v}^T\mathbf{h}(\mathbf{x}) \qquad (4.40)$$

Then Eq. (4.38) becomes

$$\nabla L(\mathbf{x}^*, \mathbf{v}^*) = \mathbf{0}, \quad \text{or} \quad \frac{\partial L(\mathbf{x}^*, \mathbf{v}^*)}{\partial x_i} = 0; \quad i = 1 \text{ to } n \qquad (4.41)$$

Differentiating $L(\mathbf{x}, \mathbf{v})$ with respect to v_j, we recover the equality constraints as

$$\frac{\partial L(\mathbf{x}^*, \mathbf{v}^*)}{\partial v_j} = 0 \Rightarrow h_j(\mathbf{x}^*) = 0; \quad j = 1 \text{ to } p \qquad (4.42)$$

Lagrange Multiplier Theorem

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_i} = -\sum_{j=1}^{p} v_j^* \frac{\partial h_j(\mathbf{x}^*)}{\partial x_i}; \quad i = 1 \text{ to } n$$

- The gradient of the cost function is a linear combination of the gradients of the constraints at the candidate minimum point.

Necessary Conditions for a General Constrained Problem

- Note that the inequality constraint may be active or inactive at the minimum point.
- How do we determine the status of an inequality at the minimum point?

Example: Active Inequality

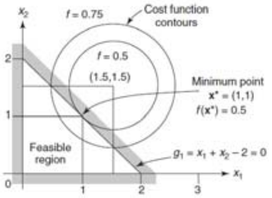
Minimize

$$f(x) = (x_1 - 1.5)^2 + (x_2 - 1.5)^2$$

subject to

$$g_1(\mathbf{x}) = x_1 + x_2 - 2 \le 0$$

$$g_2(\mathbf{x}) = -x_1 \le 0; \quad g_3(\mathbf{x}) = -x_2 \le 0$$



Example: Inactive Inequality

Minimize

$$f(\mathbf{x}) = (x_1 - 0.5)^2 + (x_2 - 0.5)^2$$

subject to

$$g_1(\mathbf{x}) = x_1 + x_2 - 2 \le 0$$

$$g_2(\mathbf{x}) = -x_1 \le 0; \quad g_3(\mathbf{x}) = -x_2 \le 0$$

Example: Infeasible Problem

Minimize

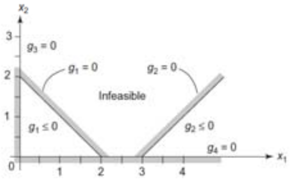
$$f(\mathbf{x}) = (x_1 - 2)^2 + (x_2 - 2)^2$$

subject to the constraints:

$$g_1(\mathbf{x}) = x_1 + x_2 - 2 \le 0$$

$$g_2(\mathbf{x}) = -x_1 + x_2 + 3 \le 0$$

$$g_3(\mathbf{x}) = -x_1 \le 0; \quad g_4(\mathbf{x}) = -x_2 \le 0$$

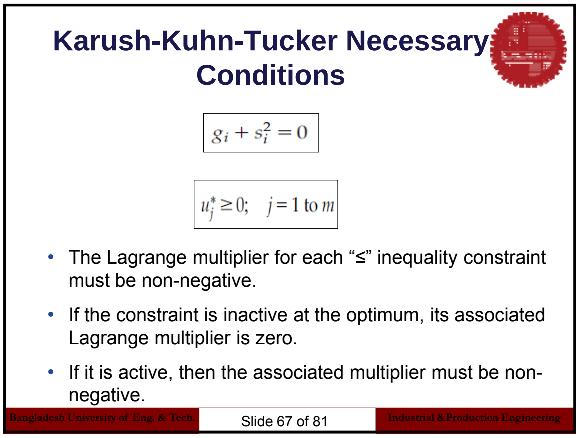


Karush-Kuhn-Tucker Necessary Conditions

Design variable vector	$\mathbf{x} = (x_1, x_2, \ldots, x_n)$	
Cost function	$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$	(4.35)
Equality constraints	$h_i(\mathbf{x}) = 0; \quad i = 1$ to p	(4.36)
Inequality constraints	$g_i(\mathbf{x}) \le 0; \quad i = 1$ to m	(4.37)

$$\boxed{g_i(\mathbf{x}) + s_i = 0}$$

- When $s_i = 0$, inequality is called an active (tight) constraint.
- When $s_i > 0$, the constraint is a strict inequality (inactive).

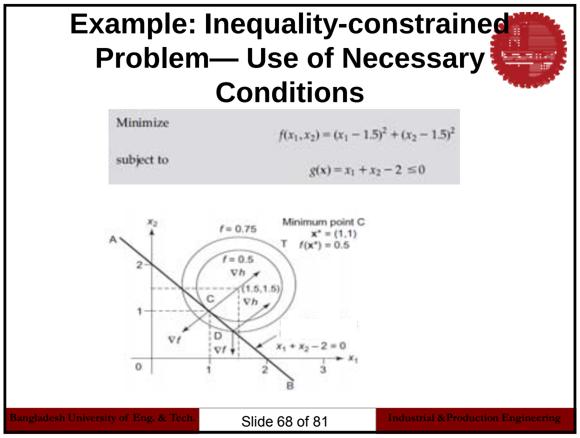


Karush-Kuhn-Tucker Necessary Conditions

$$g_i + s_i^2 = 0$$

$$u_j^* \geq 0; \quad j = 1 \text{ to } m$$

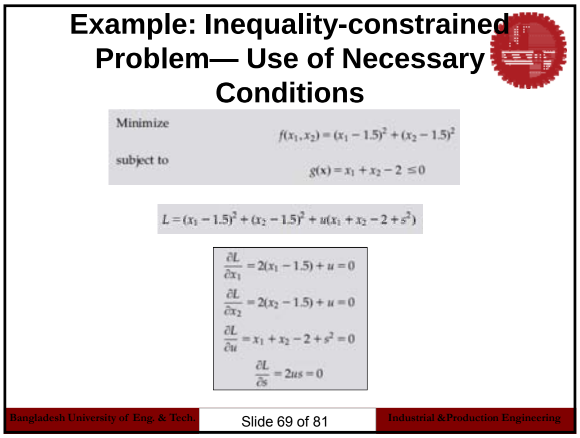
- The Lagrange multiplier for each “≤” inequality constraint must be non-negative.
- If the constraint is inactive at the optimum, its associated Lagrange multiplier is zero.
- If it is active, then the associated multiplier must be non-negative.



Example: Inequality-constrained Problem— Use of Necessary Conditions

Minimize $f(x_1, x_2) = (x_1 - 1.5)^2 + (x_2 - 1.5)^2$

subject to $g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0$



Example: Inequality-constrained Problem— Use of Necessary Conditions

Minimize $f(x_1, x_2) = (x_1 - 1.5)^2 + (x_2 - 1.5)^2$

subject to $g(\mathbf{x}) = x_1 + x_2 - 2 \leq 0$

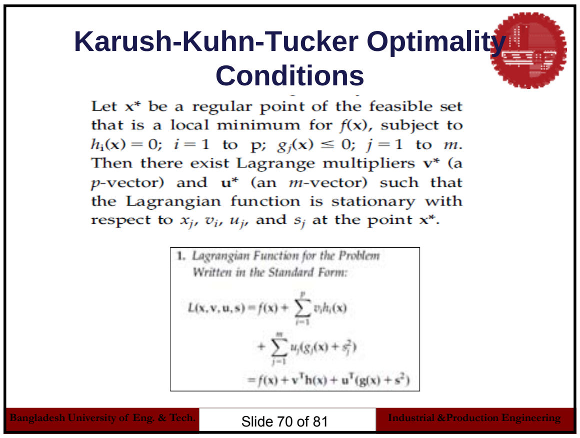
$$L = (x_1 - 1.5)^2 + (x_2 - 1.5)^2 + u(x_1 + x_2 - 2 + s^2)$$

$$\frac{\partial L}{\partial x_1} = 2(x_1 - 1.5) + u = 0$$

$$\frac{\partial L}{\partial x_2} = 2(x_2 - 1.5) + u = 0$$

$$\frac{\partial L}{\partial u} = x_1 + x_2 - 2 + s^2 = 0$$

$$\frac{\partial L}{\partial s} = 2us = 0$$

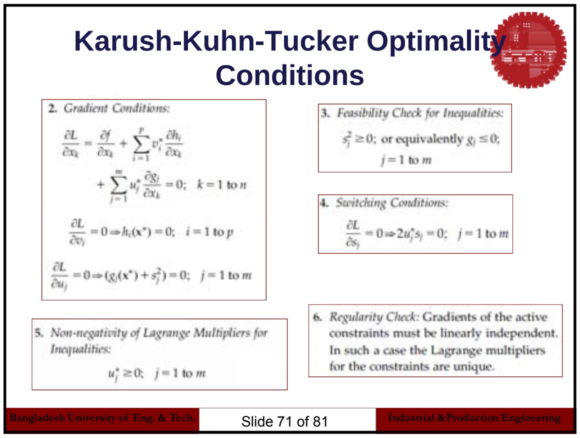


Karush-Kuhn-Tucker Optimality Conditions

Let $\mathbf{x}^*$ be a regular point of the feasible set that is a local minimum for $f(\mathbf{x})$, subject to $h_i(\mathbf{x}) = 0$; $i = 1$ to p; $g_j(\mathbf{x}) \leq 0$; $j = 1$ to m. Then there exist Lagrange multipliers $\mathbf{v}^*$ (a p-vector) and $\mathbf{u}^*$ (an m-vector) such that the Lagrangian function is stationary with respect to x_j, v_i, u_j, and s_j at the point $\mathbf{x}^*$.

1. *Lagrangian Function for the Problem Written in the Standard Form:*

$$L(\mathbf{x}, \mathbf{v}, \mathbf{u}, \mathbf{s}) = f(\mathbf{x}) + \sum_{i=1}^{p} v_i h_i(\mathbf{x}) + \sum_{j=1}^{m} u_j (g_j(\mathbf{x}) + s_j^2) = f(\mathbf{x}) + \mathbf{v}^T \mathbf{h}(\mathbf{x}) + \mathbf{u}^T (\mathbf{g}(\mathbf{x}) + \mathbf{s}^2)$$



Karush-Kuhn-Tucker Optimality Conditions

2. *Gradient Conditions:*

$$\frac{\partial L}{\partial x_k} = \frac{\partial f}{\partial x_k} + \sum_{i=1}^{p} v_i^* \frac{\partial h_i}{\partial x_k} + \sum_{j=1}^{m} u_j^* \frac{\partial g_j}{\partial x_k} = 0; \quad k = 1 \text{ to } n$$

$$\frac{\partial L}{\partial v_i} = 0 \Rightarrow h_i(\mathbf{x}^*) = 0; \quad i = 1 \text{ to } p$$

$$\frac{\partial L}{\partial u_j} = 0 \Rightarrow (g_j(\mathbf{x}^*) + s_j^2) = 0; \quad j = 1 \text{ to } m$$

3. *Feasibility Check for Inequalities:*

$s_j^2 \geq 0$; or equivalently $g_j \leq 0$; $j = 1$ to m

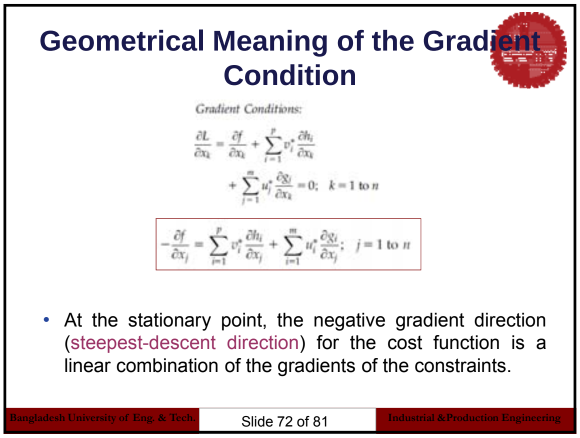
4. *Switching Conditions:*

$$\frac{\partial L}{\partial s_j} = 0 \Rightarrow 2u_j^* s_j = 0; \quad j = 1 \text{ to } m$$

5. *Non-negativity of Lagrange Multipliers for Inequalities:*

$$u_j^* \geq 0; \quad j = 1 \text{ to } m$$

6. *Regularity Check:* Gradients of the active constraints must be linearly independent. In such a case the Lagrange multipliers for the constraints are unique.



Geometrical Meaning of the Gradient Condition

Gradient Conditions:

$$\frac{\partial L}{\partial x_k} = \frac{\partial f}{\partial x_k} + \sum_{i=1}^{p} v_i^* \frac{\partial h_i}{\partial x_k} + \sum_{j=1}^{m} u_j^* \frac{\partial g_j}{\partial x_k} = 0; \quad k = 1 \text{ to } n$$

$$-\frac{\partial f}{\partial x_j} = \sum_{i=1}^{p} v_i^* \frac{\partial h_i}{\partial x_j} + \sum_{i=1}^{m} u_i^* \frac{\partial g_i}{\partial x_j}; \quad j = 1 \text{ to } n$$

- At the stationary point, the negative gradient direction (steepest-descent direction) for the cost function is a linear combination of the gradients of the constraints.

Important Observations about KKT Conditions

1. The KKT conditions are not applicable at the points that are not regular.
2. Any point that does not satisfy the KKT conditions cannot be a local minimum point unless it is an irregular point.
3. The points satisfying the KKT conditions can be constrained or unconstrained.
4. If there are equality constraints and no inequalities are active (i.e., *u* = 0), then the points satisfying the KKT conditions are only stationary.
5. If some inequality constraints are active and their multipliers are positive, then the points satisfying the KKT conditions cannot be local maxima for the cost function.

Summary of the KKT Solution Approach

1. The conditions can be used to check whether a given point is a candidate minimum.
2. Several cases defined by the switching conditions must be considered and solved.
3. For each solution case, remember to:
   - Check all inequality constraints for feasibility (e.g., $g \le 0$ or $s \ge 0$).
   - Calculate all of the Lagrange multipliers.
   - Ensure that the Lagrange multipliers for all of the inequality constraints are nonnegative.

Example: Solution to KKT Necessary Conditions

Write the KKT necessary conditions and solve them for the problem

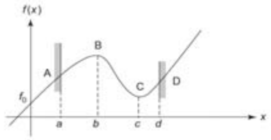
Minimize

$$f(x) = \frac{1}{3}x^3 - \frac{1}{2}(b+c)x^2 + bcx + f_0 \qquad \text{(a)}$$

subject to

$$a \le x \le d \qquad \text{(b)}$$

where $0 < a < b < c < d$ and f_0 are specified constants (created by Y. S. Ryu).



Example: Solution to KKT Necessary Conditions

$$g_1 = a - x \le 0; \quad g_2 = x - d \le 0$$

$$L = \frac{1}{3}x^3 - \frac{1}{2}(b+c)x^2 + bcx + f_0 + u_1(a - x + s_1^2) + u_2(x - d + s_2^2)$$

$$\frac{\partial L}{\partial x} = x^2 - (b+c)x + bc - u_1 + u_2 = 0$$
$$(a-x) + s_1^2 = 0, \; s_1^2 \ge 0; \quad (x-d) + s_2^2 = 0, s_2^2 \ge 0$$
$$u_1 s_1 = 0; \quad u_2 s_2 = 0$$
$$u_1 \ge 0; \quad u_2 \ge 0$$

Case 1: $u_1 = 0, u_2 = 0.$ *Case 2:* $u_1 = 0, s_2 = 0.$ *Case 3:* $s_1 = 0, u_2 = 0.$ *Case 4:* $s_1 = 0, s_2 = 0.$

Example: Solution to KKT Necessary Conditions

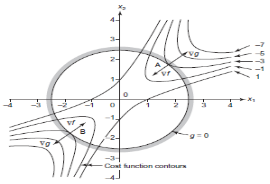
Solve the KKT condition for the problem:

Minimize

$$f(x) = x_1^2 + x_2^2 - 3x_1x_2$$

subject to

$$g = x_1^2 + x_2^2 - 6 \le 0$$



Example: Solution to KKT Necessary Conditions

Solve the KKT condition for the problem:

Minimize

$$f(x) = x_1^2 + x_2^2 - 3x_1x_2$$

subject to

$$g = x_1^2 + x_2^2 - 6 \le 0$$

$$L = x_1^2 + x_2^2 - 3x_1x_2 + u(x_1^2 + x_2^2 - 6 + s^2)$$

$$\frac{\partial L}{\partial x_1} = 2x_1 - 3x_2 + 2ux_1 = 0$$
$$\frac{\partial L}{\partial x_2} = 2x_2 - 3x_1 + 2ux_2 = 0$$
$$x_1^2 + x_2^2 - 6 + s^2 = 0, \; s^2 \ge 0, \; u \ge 0$$
$$us = 0$$

Solution to optimality conditions using MATLAB

x(1) = x_1, x(2) = x_2, x(3) = u, and x(4) = s.

```
Function F=kktsystem(x)
F=[2*x(1) - 3*x(2)+2*x(3)*x(1);
2*x(2) - 3*x(1)+2*x(3)*x(2);
x(1)^2+x(2)^2 - 6+x(4)^2;
x(3)*x(4)];
```

```
x0=[1;1;1;1];
options=optimset('Display','iter')
x=fsolve(@kktsystem,x0,options)
```

Limitation of the KKT Solution Approach

- Addition of an inequality to the problem formulation doubles the number of KKT solution cases.

Assignment

- Arora Chapter 4:
  - Problems 4.27, 4.28, 4.29, 4.42, 4.50, 4.52, 4.59, 4.61, 4.82.