CH 4020: Optimization Techniques I
CH331: Engineering Elective 4 - Optimization

Spring 2022

Instructor: Dr. Kishalay Mitra

Global Optimization & Knowledge Unearthing Lab (GOKUL)

Department of Chemical Engineering

Indian Institute of Technology Hyderabad

(kishalay@che.iith.ac.in)

https://sites.google.com/site/kishalaymitra/

NLP

A general form for a nonlinear program (NLP) is

$$
\begin{aligned}
&\text{Minimize:} && f(\mathbf{x}) && \\
&\text{Subject to:} && a_i \le g_i(\mathbf{x}) \le b_i && i = 1, \ldots, m \qquad (4.1)\\
&\text{and} && l_j \le x_j \le u_j && j = 1, \ldots, n
\end{aligned}
$$

In this problem statement, $\mathbf{x}$ is a vector of n decision variables $(x_1, \ldots, x_n)$, f is the objective function, and the g_i are constraint functions. The a_i and b_i are specified lower and upper bounds on the constraint functions with $a_i \le b_i$, and l_j, u_j are lower and upper bounds on the variables with $l_j \le u_j$. If $a_i = b_i$, the ith constraint is an equality constraint. If the upper and lower limits on g_i correspond to $a_i = -\infty$ and $b_i = +\infty$, the constraint is unbounded. Similar comments apply to the variable bounds, with $l_j = u_j$ corresponding to a variable x_j whose value is fixed, and $l_j = -\infty$ and $u_j = +\infty$ specifying a free variable.

- **Nonlinear / Unconstrained / Bound constrained / linearly constrained / feasible / feasible region (set F) / infeasible (empty set F) / unbounded**

Solution Characterization

we consider optimization problems of the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \Omega, \end{aligned}$$

where f is a real-valued function and Ω, the feasible set, is a subset of E^n.

- **Perhaps the first question that arises here is whether a solution exists.**
- **The theorem of Weierstras states that if f is continuous and is compact, a solution exists.**
- **Given this result, our primary concern is with characterizing solution points and devising effective methods for finding them.**
- **A point x* called local optima if there is no other better point in its small neighborhood**

$$F(x^*) \leq F(x)$$

- **Other optima can exist outside this small neighborhood meaning NLP can have more than one local optima in the entire search space**
- **If the above relation is true for the point x* in the entire search space (for all x belongs to Feasible set F), x* can be global optima**

Solution Definition

Definition. A point $\mathbf{x}^* \in \Omega$ is said to be a *relative minimum point* or a *local minimum point* of f over Ω if there is an $\varepsilon > 0$ such that $f(\mathbf{x}) \geqslant f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega$ within a distance ε of $\mathbf{x}^*$ (that is, $\mathbf{x} \in \Omega$ and $|\mathbf{x} - \mathbf{x}^*| < \varepsilon$). If $f(\mathbf{x}) > f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega$, $\mathbf{x} \neq \mathbf{x}^*$, within a distance ε of $\mathbf{x}^*$, then $\mathbf{x}^*$ is said to be a *strict relative minimum point* of f over Ω.

Definition. A point $\mathbf{x}^* \in \Omega$ is said to be a *global minimum point* of f over Ω if $f(\mathbf{x}) \geqslant f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega$. If $f(\mathbf{x}) > f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega$, $\mathbf{x} \neq \mathbf{x}^*$, then $\mathbf{x}^*$ is said to be a *strict global minimum point* of f over Ω.

Practicality

- In formulating and attacking an optimization problem, we are, by definition, explicitly asking for a global minimum point of f over the entire feasible set.

- Practical reality (both from the theoretical and computational viewpoint) dictates that we must in many circumstances be content with a relative minimum point.

- In deriving necessary conditions based on the differential calculus or when searching for the minimum point by a convergent stepwise procedure, comparisons of the values of nearby points is all that is possible and attention focuses on relative minimum points.

- Global conditions and global solutions can, as a rule, only be found if the problem possesses certain convexity properties that essentially guarantee that any relative minimum is a global minimum.

Two Aspects

we consider optimization problems of the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \Omega, \end{aligned}$$

where f is a real-valued function and Ω, the feasible set, is a subset of E^n.

- **Our primary concern is with characterizing solution points (theoretical aspect) and devising effective methods for finding them (computational aspect).**
- **Characterizing solution / theoretical procedure – when can we say that the optimum is achieved? This is also required for building a computational procedure.**
- **Theoretical characterization –**
  - **Necessary condition (who can be a candidate minimum point)**
    - **Point doesn't satisfy this is ruled out of contention of being a minimum point**
  - **sufficient condition (out of the candidates who is the true minimum)**
    - **What additional conditions a point satisfying necessary condition should satisfy to become a local minimum**

Optimization Conditions (SV)

Theorem 2.1 Necessary Condition If a function $f(x)$ is defined in the interval $a \leq x \leq b$ and has a relative minimum at $x = x^*$, where $a < x^* < b$, and if the derivative $df(x)/dx = f'(x)$ exists as a finite number at $x = x^*$, then $f'(x^*) = 0$.

$$f'(x^*) = \lim_{h \to 0} \frac{f(x^* + h) - f(x^*)}{h}$$

exists as a definite number, which we want to prove to be zero. Since x^* is a relative minimum, we have

$$f(x^*) \leq f(x^* + h)$$

for all values of h sufficiently close to zero. Hence

$$\frac{f(x^* + h) - f(x^*)}{h} \geq 0 \quad \text{if } h > 0$$

$$\frac{f(x^* + h) - f(x^*)}{h} \leq 0 \quad \text{if } h < 0$$

Thus Eq. (2.1) gives the limit as h tends to zero through positive values as

$$f'(x^*) \geq 0 \tag{2.2}$$

while it gives the limit as h tends to zero through negative values as

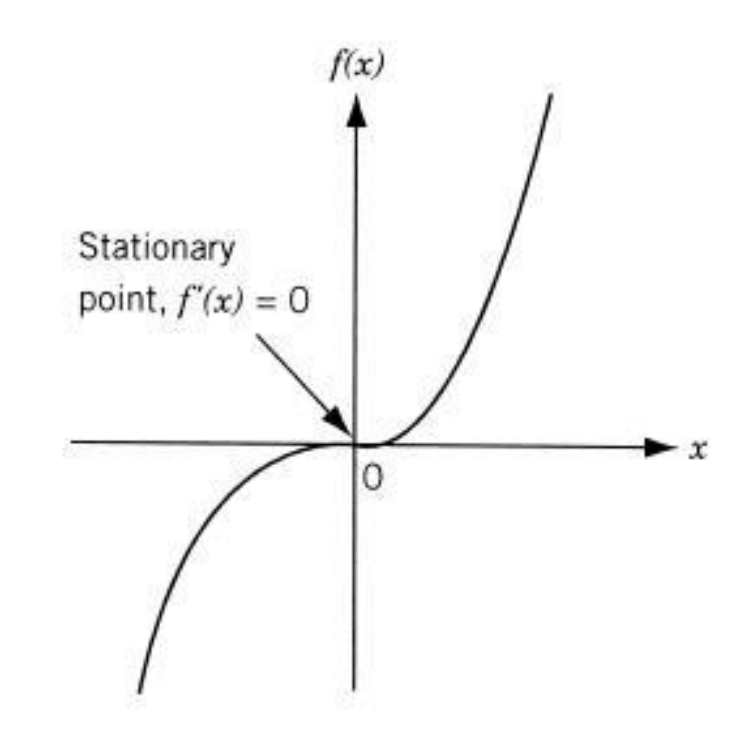
$$f'(x^*) \leq 0 \tag{2.3}$$

The only way to satisfy both Eqs. (2.2) and (2.3) is to have

$$f'(x^*) = 0 \tag{2.4}$$

Optimization Conditions ...

- Condition does not necessarily say that there will be maximum or minimum in case f′(x) = 0 (reverse of the theorem is not true)
- Theorem does not say what happens when minimum or maximum occurs at a point where derivative fails to exist - assumption is that derivative exists at that point



Optimization Conditions...

Theorem 2.2 Sufficient Condition Let $f'(x^*) = f''(x^*) = \cdots = f^{(n-1)}(x^*) = 0$, but $f^{(n)}(x^*) \neq 0$. Then $f(x^*)$ is (i) a minimum value of $f(x)$ if $f^{(n)}(x^*) > 0$ and n is even; (ii) a maximum value of $f(x)$ if $f^{(n)}(x^*) < 0$ and n is even; (iii) neither a maximum nor a minimum if n is odd.

Proof: Applying Taylor's theorem with remainder after n terms, we have

$$f(x^* + h) = f(x^*) + hf'(x^*) + \frac{h^2}{2!} f''(x^*) + \cdots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(x^*) + \frac{h^n}{n!} f^{(n)}(x^* + \theta h) \quad \text{for} \quad 0 < \theta < 1 \tag{2.5}$$

Since $f'(x^*) = f''(x^*) = \cdots = f^{(n-1)}(x^*) = 0$, Eq. (2.5) becomes

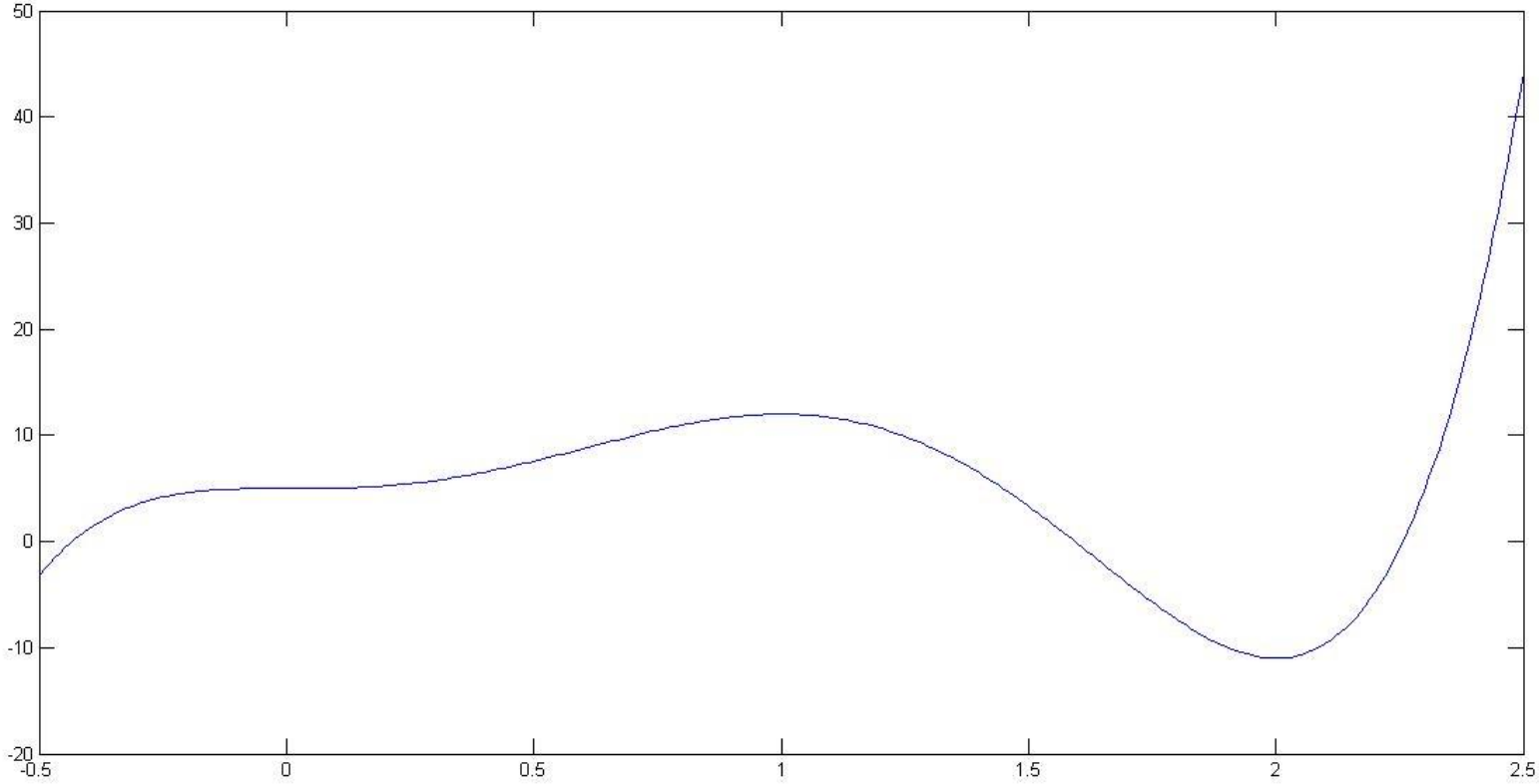
$$f(x^* + h) - f(x^*) = \frac{h^n}{n!} f^{(n)}(x^* + \theta h)$$

As $f^{(n)}(x^*) \neq 0$, there exists an interval around x^* for every point x of which the nth derivative $f^{(n)}(x)$ has the same sign, namely, that of $f^{(n)}(x^*)$. Thus for every point $x^* + h$ of this interval, $f^{(n)}(x^* + \theta h)$ has the sign of $f^{(n)}(x^*)$. When n is even, $h^n/n!$ is positive irrespective of whether h is positive or negative, and hence $f(x^* + h) - f(x^*)$ will have the same sign as that of $f^{(n)}(x^*)$. Thus x^* will be a relative minimum if $f^{(n)}(x^*)$ is positive and a relative maximum if $f^{(n)}(x^*)$ is negative. When n is odd, $h^n/n!$ changes sign with the change in the sign of h and hence the point x^* is neither a maximum nor a minimum. In this case the point x^* is called a *point of inflection*.

Finding Max / Min

Example 2.1 Determine the maximum and minimum values of the function

$$f(x) = 12x^5 - 45x^4 + 40x^3 + 5$$



Finding Max / Min

Example 2.1 Determine the maximum and minimum values of the function

$$f(x) = 12x^5 - 45x^4 + 40x^3 + 5$$

Finding Max / Min ...

Does $f(x) = x^4$ have an extremum? If so, what is the value of x^* and $f(x^*)$ at the extremum?

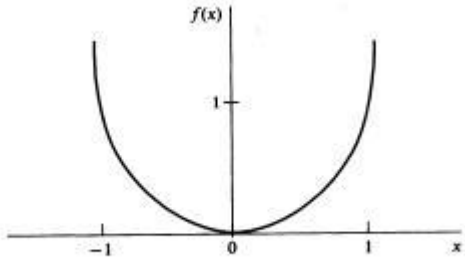


FIGURE E4.10

Continuous Function Example

- **When a local optimum becomes a global optimum? – notion of continuity and convexity!**

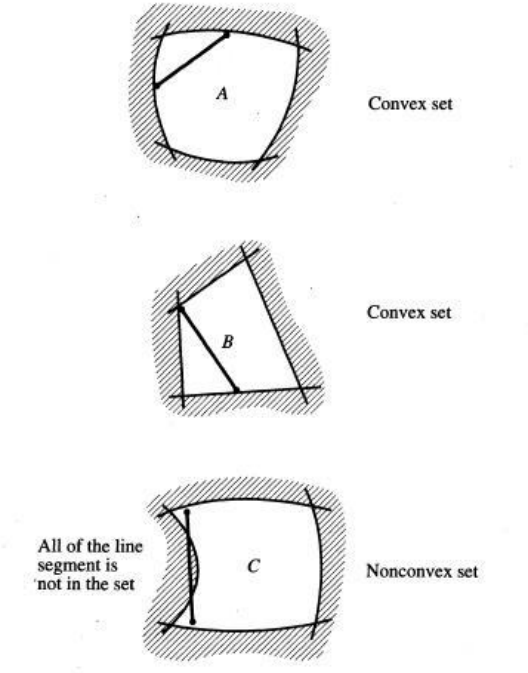
Are the following functions continuous? (a) $f(x) = 1/x$; (b) $f(x) = \ln x$. In each case specify the range of x for which $f(x)$ and $f'(x)$ are continuous.

Convex Set

In order to develop a theory directed toward characterizing global, rather than local, minimum points, it is necessary to introduce some sort of convexity assumptions.

Convex set

A set of points (or a *region*) is defined as a *convex set* in n-dimensional space if, for all pairs of points $\mathbf{x}_1$ and $\mathbf{x}_2$ in the set, the straight-line segment joining them is also entirely in the set. Figure 4.9 illustrates the concept in two dimensions.



Convex Function

In order to develop a theory directed toward characterizing global, rather than local, minimum points, it is necessary to introduce some sort of convexity assumptions.

Definition. A function f defined on a convex set Ω is said to be *convex* if, for every $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and every α, $0 \leqslant \alpha \leqslant 1$, there holds

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leqslant \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2).$$

If, for every α, $0 < \alpha < 1$, and $\mathbf{x}_1 \neq \mathbf{x}_2$, there holds

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) < \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2),$$

then f is said to be *strictly convex*.

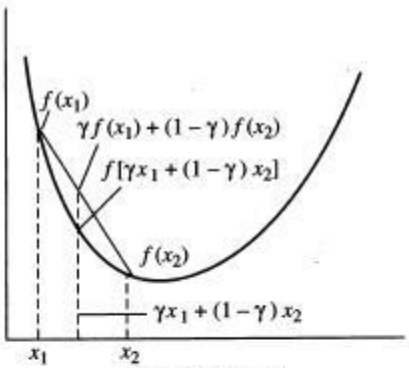
Convex Function

- **Function value of linear combination of any two points should under-estimate the linear combination of function values calculated at those two points (say, x_1 and x_2)**

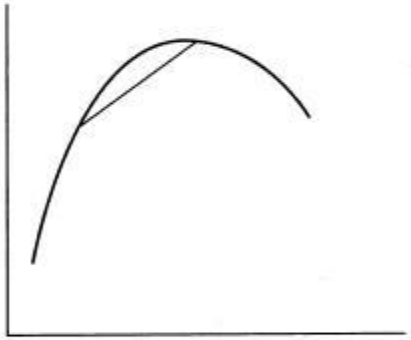
$$f[\gamma \mathbf{x}_1 + (1-\gamma)\mathbf{x}_2] \leq \gamma f(\mathbf{x}_1) + (1-\gamma) f(\mathbf{x}_2)$$

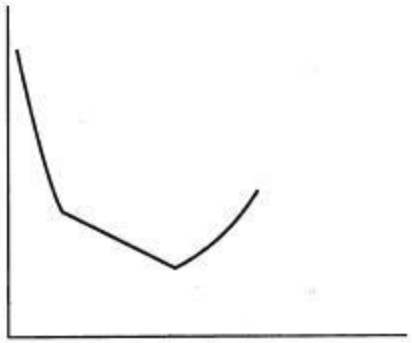
Next we turn to the definition of a concave function.

Definition. A function g defined on a convex set Ω is said to be *concave* if the function $f = -g$ is convex. The function g is *strictly concave* if $-g$ is strictly convex.



Convex function


Concave function


Convex function

Convex Function ...

Proposition 1. *Let f_1 and f_2 be convex functions on the convex set Ω. Then the function $f_1 + f_2$ is convex on Ω.*

Proposition 2. *Let f be a convex function over the convex set Ω. Then the function af is convex for any $a \geqslant 0$.*

Note that through repeated application of the above two propositions it follows that a positive combination $a_1 f_1 + a_2 f_2 + \ldots + a_m f_m$ of convex functions is again convex.

Proposition 3. *Let f be a convex function on a convex set Ω. The set $\Gamma_c = \{\mathbf{x} : \mathbf{x} \in \Omega, f(\mathbf{x}) \leqslant c\}$ is convex for every real number c.*

We note that, since the intersection of convex sets is also convex, the set of points simultaneously satisfying

$$f_1(\mathbf{x}) \leqslant c_1, \quad f_2(\mathbf{x}) \leqslant c_2, \ldots, f_m(\mathbf{x}) \leqslant c_m,$$

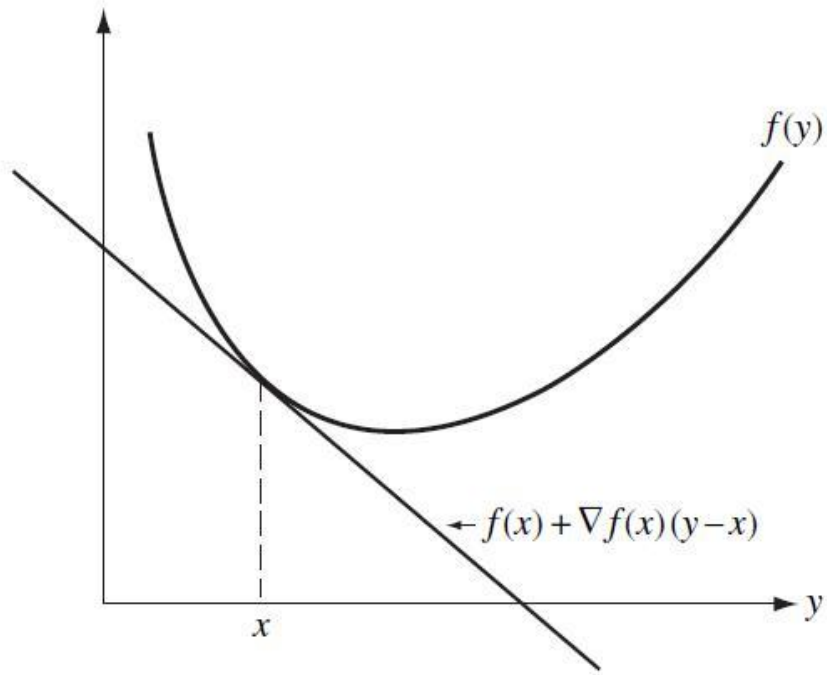
where each f_i is a convex function, defines a convex set. This is important in mathematical programming, since the constraint set is often defined this way.

Differentiable Convex Function

Proposition 4. *Let $f \in C^1$. Then f is convex over a convex set Ω if and only if*

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \tag{9}$$

for all $\mathbf{x}, \mathbf{y} \in \boldsymbol{\Omega}$.



Proposition 5. *Let $f \in C^2$. Then f is convex over a convex set Ω containing an interior point if and only if the Hessian matrix $\mathbf{F}$ of f is positive semidefinite throughout Ω.*

Differentiable Convex Function …

- ✓ **The Hessian matrix is the generalization of the concept of the curvature of a function.**
- ✓ **Positive definiteness of the Hessian is the generalization of positive curvature.**
- ✓ **Convex functions have positive (or at least nonnegative) curvature in every direction.**
- ✓ **Motivated by these observations, we sometimes refer to a function as being locally convex if its Hessian matrix is positive semi-definite in a small region.**
- ✓ **Locally strictly convex if the Hessian is positive definite in the region.**

Convex / Concave

from the concept of a convex set of points. A set of points $\mathbf{x}$ satisfying the relation

$$\mathbf{x}^T\mathbf{H}(\mathbf{x})\mathbf{x} \leq 1$$

is convex if the Hessian matrix $\mathbf{H}(\mathbf{x})$ is a real symmetric positive-semidefinite matrix. $\mathbf{H}(\mathbf{x})$ is another symbol for $\nabla^2 f(\mathbf{x})$, the matrix of second partial derivative of $f(\mathbf{x})$ with respect to each x_i

$$\mathbf{H}(\mathbf{x}) \equiv \mathbf{H} \equiv \nabla^2 f(\mathbf{x})$$

1. $\mathbf{H}$ is *positive-definite* if and only if $\mathbf{x}^T\mathbf{H}\mathbf{x}$ is > 0 for all $\mathbf{x} \neq \mathbf{0}$.
2. $\mathbf{H}$ is *negative-definite* if and only if $\mathbf{x}^T\mathbf{H}\mathbf{x}$ is < 0 for all $\mathbf{x} \neq \mathbf{0}$.
3. $\mathbf{H}$ is *positive-semidefinite* if and only if $\mathbf{x}^T\mathbf{H}\mathbf{x}$ is ≥ 0 for all $\mathbf{x} \neq \mathbf{0}$.
4. $\mathbf{H}$ is *negative-semidefinite* if and only if $\mathbf{x}^T\mathbf{H}\mathbf{x}$ is ≤ 0 for all $\mathbf{x} \neq \mathbf{0}$.
5. $\mathbf{H}$ is *indefinite* if $\mathbf{x}^T\mathbf{H}\mathbf{x} < 0$ for some $\mathbf{x}$ and > 0 for other $\mathbf{x}$.

Convex / Concave

- **For multivariate function**

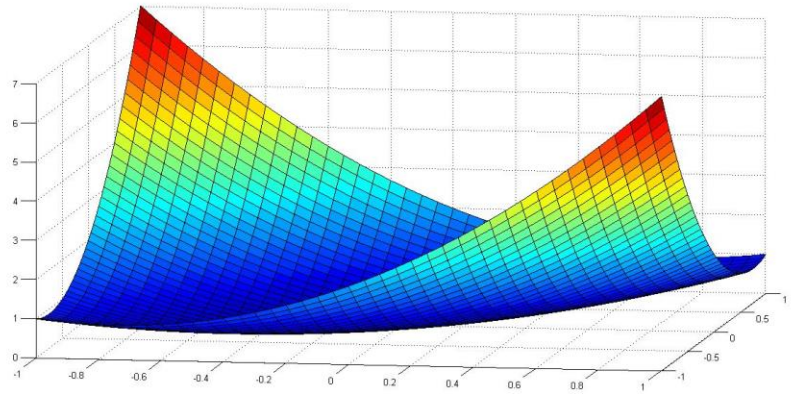
$f(\mathbf{x})$ is	$\mathbf{H}(\mathbf{x})$ is	All the eigenvalues of $\mathbf{H}(\mathbf{x})$ are
Strictly convex	Positive-definite	>0
Convex	Positive-semidefinite	≥ 0
Concave	Negative-semidefinite	≤ 0
Strictly concave	Negative-definite	<0

Convex / Concave

(a) $f(x) = 3x^2$
(b) $f(x) = 2x$
(c) $f(x) = -5x^2$
(d) $f(x) = 2x^2 - x^3$

determine if $f(x)$ is convex, concave, strictly convex, strictly concave, all, or none of these classes in the range $-\infty \leq x \leq \infty$.

Convex / Concave



Classify the function $f(\mathbf{x}) = 2x_1^2 - 3x_1x_2 + 2x_2^2$ using the categories in Table 4.1, or state that it does not belong in any of the categories.

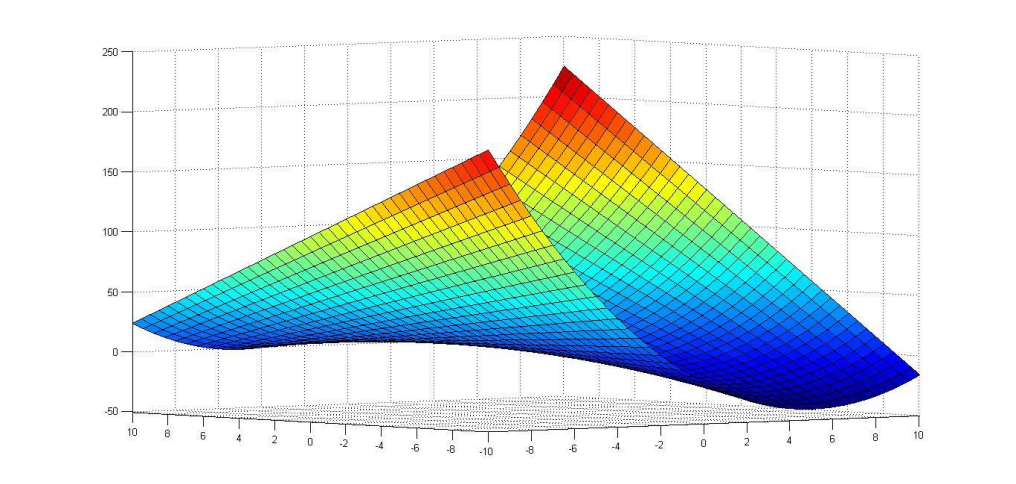
Solution

$$\frac{\partial f(x)}{\partial x_1} = 4x_1 - 3x_2 \qquad \frac{\partial^2 f(x)}{\partial x_1^2} = 4 \qquad \frac{\partial^2 f(x)}{\partial x_2^2} = 4$$

$$\frac{\partial f(x)}{\partial x_2} = -3x_1 + 4x_2 \qquad \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} = \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} = -3 \qquad \mathbf{H}(\mathbf{x}) = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix}$$

The eigenvalues of $\mathbf{H}$ are 7 and 1, hence $\mathbf{H}(\mathbf{x})$ is positive-definite. Consequently, $f(\mathbf{x})$ is strictly convex (as well as convex).

Convex / Concave



Repeat the analysis of Example 4.4 for $f(\mathbf{x}) = x_1^2 + x_1x_2 + 2x_2 + 4$

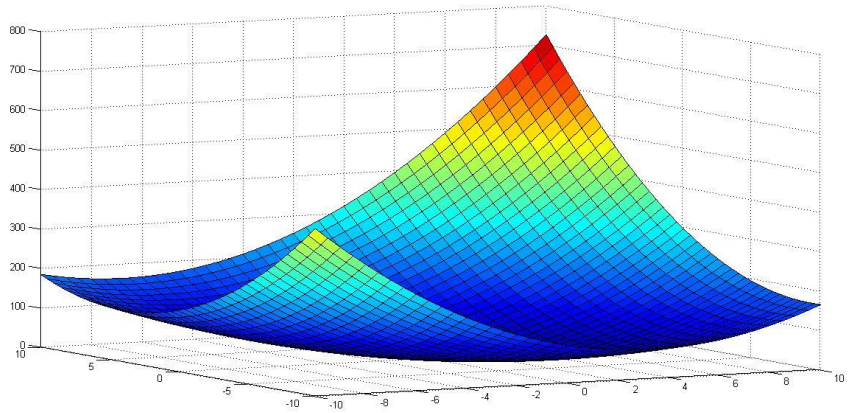
Convex / Concave

Determine if the following function

$$f(\mathbf{x}) = 2x_1 + 3x_2 + 6$$

is convex or concave.

Convex / Concave



Consider the following objective function: Is it convex?

$$f(\mathrm{x}) = 2x_1^2 + 2x_1x_2 + 1.5x_2^2 + 7x_1 + 8x_2 + 24$$

Optimization Conditions (MV)

Theorem 2.3 Necessary Condition If $f(\mathbf{X})$ has an extreme point (maximum or minimum) at $\mathbf{X} = \mathbf{X}^*$ and if the first partial derivatives of $f(\mathbf{X})$ exist at $\mathbf{X}^*$, then

$$\frac{\partial f}{\partial x_1}(\mathbf{X}^*) = \frac{\partial f}{\partial x_2}(\mathbf{X}^*) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{X}^*) = 0 \tag{2.9}$$

Proof: The proof given for Theorem 2.1 can easily be extended to prove the present theorem. However, we present a different approach to prove this theorem. Suppose that one of the first partial derivatives, say the kth one, does not vanish at $\mathbf{X}^*$. Then, by Taylor's theorem,

$$f(\mathbf{X}^* + \mathbf{h}) = f(\mathbf{X}^*) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(\mathbf{X}^*) + R_1(\mathbf{X}^*, \mathbf{h})$$

that is,

$$f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*) = h_k \frac{\partial f}{\partial x_k}(\mathbf{X}^*) + \frac{1}{2!} d^2 f(\mathbf{X}^* + \theta \mathbf{h}), \qquad 0 < \theta < 1$$

Since $d^2 f(\mathbf{X}^* + \theta \mathbf{h})$ is of order h_i^2, the terms of order $\mathbf{h}$ will dominate the higher-order terms for small $\mathbf{h}$. Thus the sign of $f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*)$ is decided by the sign of $h_k\, \partial f(\mathbf{X}^*)/\partial x_k$. Suppose that $\partial f(\mathbf{X}^*)/\partial x_k > 0$. Then the sign of $f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*)$ will be positive for $h_k > 0$ and negative for $h_k < 0$. This means that $\mathbf{X}^*$ cannot be an extreme point. The same conclusion can be obtained even if we assume that $\partial f(\mathbf{X}^*)/\partial x_k < 0$. Since this conclusion is in contradiction with the original statement that $\mathbf{X}^*$ is an extreme point, we may say that $\partial f/\partial x_k = 0$ at $\mathbf{X} = \mathbf{X}^*$. Hence the theorem is proved.

Optimization Conditions ...

Theorem 2.4 Sufficient Condition A sufficient condition for a stationary point $\mathbf{X}^*$ to be an extreme point is that the matrix of second partial derivatives (Hessian matrix) of $f(\mathbf{X})$ evaluated at $\mathbf{X}^*$ is (i) positive definite when $\mathbf{X}^*$ is a relative minimum point, and (ii) negative definite when $\mathbf{X}^*$ is a relative maximum point.

Proof: From Taylor's theorem we can write

$$f(\mathbf{X}^* + \mathbf{h}) = f(\mathbf{X}^*) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(\mathbf{X}^*) + \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{X}=\mathbf{X}^*+\theta\mathbf{h}},$$

$$0 < \theta < 1 \tag{2.10}$$

Since $\mathbf{X}^*$ is a stationary point, the necessary conditions give (Theorem 2.3)

$$\frac{\partial f}{\partial x_i} = 0, \qquad i = 1, 2, \ldots, n$$

Thus Eq. (2.10) reduces to

$$f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*) = \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{X}=\mathbf{X}^*+\theta\mathbf{h}}, \qquad 0 < \theta < 1$$

Optimization Conditions ...

Therefore, the sign of

$$f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*)$$

will be same as that of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{X}=\mathbf{X}^*+\theta\mathbf{h}}$$

Since the second partial derivative of $\partial^2 f(\mathbf{X})/\partial x_i \partial x_j$ is continuous in the neighborhood of $\mathbf{X}^*$,

$$\left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{X}=\mathbf{X}^*+\theta\mathbf{h}}$$

will have the same sign as $(\partial^2 f/\partial x_i \partial x_j)|\ \mathbf{X} = \mathbf{X}^*$ for all sufficiently small $\mathbf{h}$. Thus $f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*)$ will be positive, and hence X^* will be a relative minimum, if

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{X}=\mathbf{X}^*} \tag{2.11}$$

is positive. This quantity Q is a quadratic form and can be written in matrix form as

$$Q = \mathbf{h}^{\mathrm{T}} \mathbf{J} \mathbf{h}|_{\mathbf{X}=\mathbf{X}^*} \tag{2.12}$$

where

$$\mathbf{J}|_{\mathbf{X}=\mathbf{X}^*} = \left[\left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{X}=\mathbf{X}^*} \right] \tag{2.13}$$

is the matrix of second partial derivatives and is called the *Hessian matrix* of $f(\mathbf{X})$.

It is known from matrix algebra that the quadratic form of Eq. (2.11) or (2.12) will be positive for all $\mathbf{h}$ if and only if $[\mathbf{J}]$ is positive definite at $\mathbf{X} = \mathbf{X}^*$. This means that a sufficient condition for the stationary point $\mathbf{X}^*$ to be a relative minimum is that the Hessian matrix evaluated at the same point be positive definite. This completes the

Positive Definiteness

$f(\mathbf{x})$ is	H(x) is	All the eigenvalues of H(x) are
Strictly convex	Positive-definite	>0
Convex	Positive-semidefinite	≥ 0
Concave	Negative-semidefinite	≤ 0
Strictly concave	Negative-definite	<0

Another test that can be used to find the positive definiteness of a matrix **A** of order n involves evaluation of the determinants

$$A = |a_{11}|,$$

$$A_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},$$

$$A_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{32} \end{vmatrix}, \ldots,$$

$$A_n = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & & & & \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} \tag{2.15}$$

The matrix **A** will be positive definite if and only if all the values $A_1, A_2, A_3, \ldots, A_n$ are positive. The matrix **A** will be negative definite if and only if the sign of A_j is $(-1)^j$ for $j = 1, 2, \ldots, n$. If some of the A_j are positive and the remaining A_j are zero, the matrix **A** will be positive semidefinite.

Finding Max / Min ...

EXAMPLE 4.11 CALCULATION OF EXTREMA

Identify the stationary points of the following function (Fox, 1971), and determine if any extrema exist.

$$f(\mathbf{x}) = 4 + 4.5x_1 - 4x_2 + x_1^2 + 2x_2^2 - 2x_1x_2 + x_1^4 - 2x_1^2x_2$$

Solution. For this function, three stationary points can be located by setting $\nabla f(\mathbf{x}) = \mathbf{0}$:

$$\frac{\partial f(x)}{\partial x_1} = 4.5 + 2x_1 - 2x_2 + 4x_1^3 - 4x_1x_2 = 0 \qquad (a)$$

$$\frac{\partial f(x)}{\partial x_2} = -4 + 4x_2 - 2x_1 - 2x_1^2 = 0 \qquad (b)$$

The set of nonlinear equations (a) and (b) has to be solved, say by Newton's method, to get the pairs (x_1, x_2) as follows:

Point	Stationary point (x_1, x_2)	$f(\mathbf{x})$	Hessian matrix eigenvalues		Classification
B	(1.941, 3.854)	0.9855	37.03	0.97	Local minimum
A	(−1.053, 1.028)	−0.5134	10.5	3.5	Local minimum (also the global minimum)
C	(0.6117, 1.4929)	2.83	7.0	−2.56	Saddle point

Figure 4.17b shows contours for the objective function in this example. Note that the global minimum can only be identified by evaluating $f(\mathbf{x})$ for all the local minima.

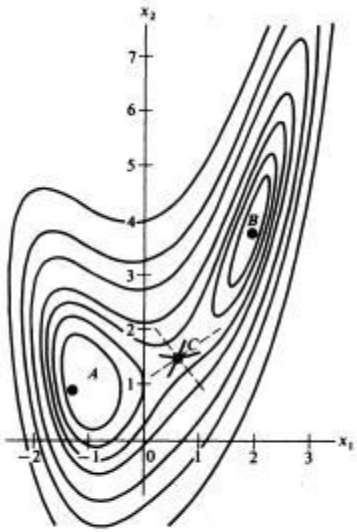


FIGURE 4.17b
A function of two variables with three stationary points and two extrema, A and B.

Semi-definite Case

Theorem 2.5 Let the partial derivatives of f of all orders up to the order $k \geq 2$ be continuous in the neighborhood of a stationary point $\mathbf{X}^*$, and

$$d^r f|_{\mathbf{X}=\mathbf{X}^*} = 0, \qquad 1 \leq r \leq k-1$$

$$d^k f|_{\mathbf{X}=\mathbf{X}^*} \neq 0$$

so that $d^k f|_{\mathbf{X}=\mathbf{X}^*}$ is the first nonvanishing higher-order differential of f at $\mathbf{X}^*$. If k is even, then (i) $\mathbf{X}^*$ is a relative minimum if $d^k f|_{\mathbf{X}=\mathbf{X}^*}$ is positive definite, (ii) $\mathbf{X}^*$ is a relative maximum if $d^k f|_{\mathbf{X}=\mathbf{X}^*}$ is negative definite, and (iii) if $d^k f|_{\mathbf{X}=\mathbf{X}^*}$ is semidefinite (but not definite), no general conclusion can be drawn. On the other hand, if k is odd, $\mathbf{X}^*$ is not an extreme point of $f(\mathbf{X})$.

Indefinite Case

In the case of a function of two variables, $f(x, y)$, the Hessian matrix may be neither positive nor negative definite at a point (x^*, y^*) at which

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$

In such a case, the point (x^*, y^*) is called a *saddle point*. The characteristic of a

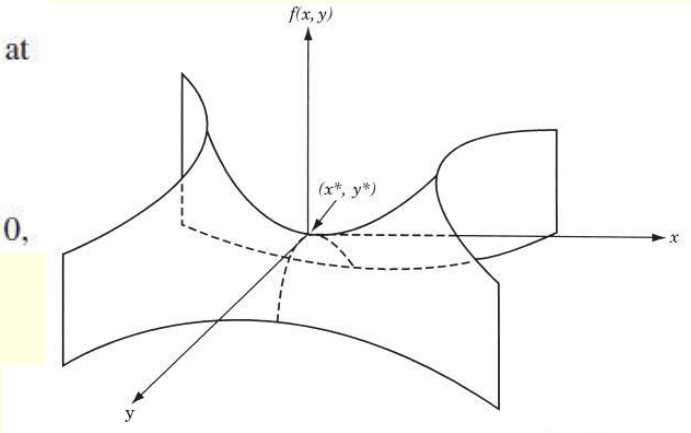
As an example, consider the function $f(x, y) = x^2 - y^2$. For this function,

$$\frac{\partial f}{\partial x} = 2x \quad \text{and} \quad \frac{\partial f}{\partial y} = -2y$$

These first derivatives are zero at $x^* = 0$ and $y^* = 0$. The Hessian matrix of f at (x^*, y^*) is given by

$$\mathbf{J} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

Since this matrix is neither positive definite nor negative definite, the point $(x^* = 0, y^* = 0)$ is a saddle point.



Indefinite Case ...

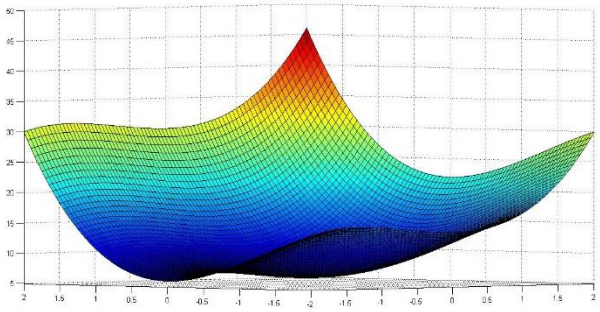
Example 2.5 Find the extreme points of the function

$$f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$$

SOLUTION The necessary conditions for the existence of an extreme point are

$$\frac{\partial f}{\partial x_1} = 3x_1^2 + 4x_1 = x_1(3x_1 + 4) = 0$$

$$\frac{\partial f}{\partial x_2} = 3x_2^2 + 8x_2 = x_2(3x_2 + 8) = 0$$



These equations are satisfied at the points

$$(0, 0), \quad (0, -\tfrac{8}{3}), \quad (-\tfrac{4}{3}, 0), \quad \text{and} \quad (-\tfrac{4}{3}, -\tfrac{8}{3})$$

To find the nature of these extreme points, we have to use the sufficiency conditions. The second-order partial derivatives of f are given by

$$\frac{\partial^2 f}{\partial x_1^2} = 6x_1 + 4$$

$$\frac{\partial^2 f}{\partial x_2^2} = 6x_2 + 8$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = 0$$

The Hessian matrix of f is given by

$$\mathbf{J} = \begin{bmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{bmatrix}$$

If $J_1 = |6x_1 + 4|$ and $J_2 = \begin{vmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{vmatrix}$, the values of J_1 and J_2 and the nature of the extreme point are as given below:

Point **X**	Value of J_1	Value of J_2	Nature of **J**	Nature of **X**	$f(\mathbf{X})$
(0, 0)	+4	+32	Positive definite	Relative minimum	6
$(0, -\frac{8}{3})$	+4	−32	Indefinite	Saddle point	418/27
$(-\frac{4}{3}, 0)$	−4	−32	Indefinite	Saddle point	194/27
$(-\frac{4}{3}, -\frac{8}{3})$	−4	+32	Negative definite	Relative maximum	50/3

MV Optn w Constraints

- **Method of Direct Substitution**

Example 2.6 Find the dimensions of a box of largest volume that can be inscribed in a sphere of unit radius.

SOLUTION Let the origin of the Cartesian coordinate system x_1, x_2, x_3 be at the center of the sphere and the sides of the box be $2x_1$, $2x_2$, and $2x_3$. The volume of the box is given by

$$f(x_1, x_2, x_3) = 8x_1x_2x_3 \qquad (E_1)$$

Since the corners of the box lie on the surface of the sphere of unit radius, x_1, x_2, and x_3 have to satisfy the constraint

$$x_1^2 + x_2^2 + x_3^2 = 1 \qquad (E_2)$$

This problem has three design variables and one equality constraint. Hence the equality constraint can be used to eliminate any one of the design variables from the objective function. If we choose to eliminate x_3, Eq. (E_2) gives

$$x_3 = (1 - x_1^2 - x_2^2)^{1/2} \qquad (E_3)$$

Thus the objective function becomes

$$f(x_1, x_2) = 8x_1x_2(1 - x_1^2 - x_2^2)^{1/2} \qquad (E_4)$$

which can be maximized as an unconstrained function in two variables.

The necessary conditions for the maximum of f give

$$\frac{\partial f}{\partial x_1} = 8x_2\left[(1 - x_1^2 - x_2^2)^{1/2} - \frac{x_1^2}{(1 - x_1^2 - x_2^2)^{1/2}}\right] = 0 \qquad (E_5)$$

$$\frac{\partial f}{\partial x_2} = 8x_1\left[(1 - x_1^2 - x_2^2)^{1/2} - \frac{x_2^2}{(1 - x_1^2 - x_2^2)^{1/2}}\right] = 0 \qquad (E_6)$$

Equations (E_5) and (E_6) can be simplified to obtain

$$1 - 2x_1^2 - x_2^2 = 0$$

$$1 - x_1^2 - 2x_2^2 = 0$$

from which it follows that $x_1^* = x_2^* = 1/\sqrt{3}$ and hence $x_3^* = 1/\sqrt{3}$. This solution gives the maximum volume of the box as

$$f_{\max} = \frac{8}{3\sqrt{3}}$$

To find whether the solution found corresponds to a maximum or a minimum, we apply the sufficiency conditions to $f(x_1, x_2)$ of Eq. (E_4). The second-order partial derivatives of f at (x_1^*, x_2^*) are given by

$$\frac{\partial^2 f}{\partial x_1^2} = -\frac{32}{\sqrt{3}} \text{ at } (x_1^*, x_2^*)$$

$$\frac{\partial^2 f}{\partial x_2^2} = -\frac{32}{\sqrt{3}} \text{ at } (x_1^*, x_2^*)$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = -\frac{16}{\sqrt{3}} \text{ at } (x_1^*, x_2^*)$$

Since

$$\frac{\partial^2 f}{\partial x_1^2} < 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 > 0$$

the Hessian matrix of f is negative definite at (x_1^*, x_2^*). Hence the point (x_1^*, x_2^*) corresponds to the maximum of f.

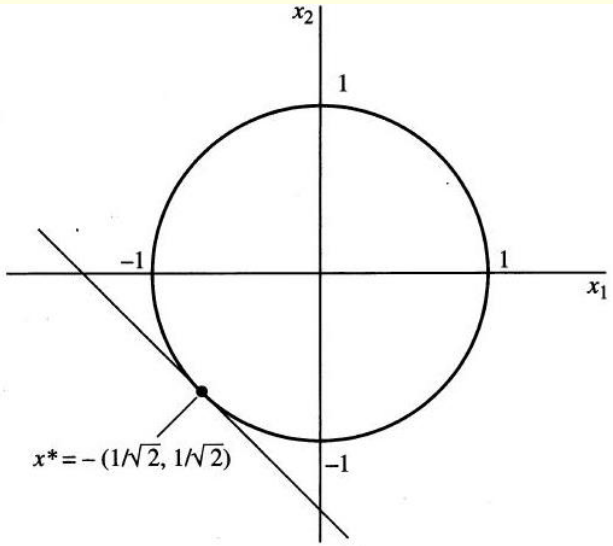
Graphical Interpretation

$$\text{Minimize:} \quad f(x_1, x_2) = x_1 + x_2$$

$$\text{Subject to:} \quad h(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0$$

$$\nabla f(\mathbf{x}^*) = [1, 1]$$

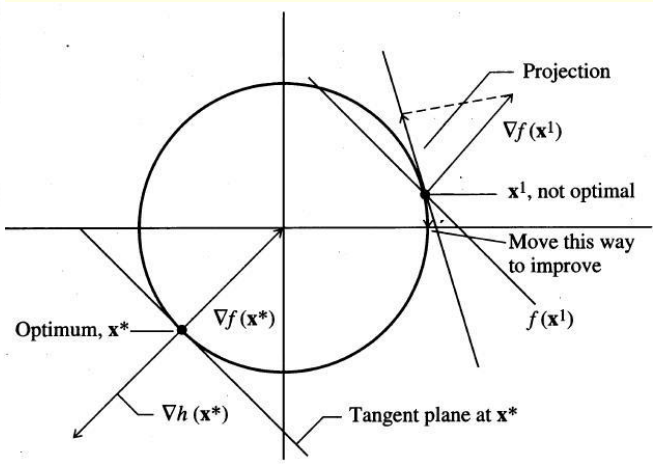
$$\nabla h(\mathbf{x}^*) = [2x_1, 2x_2]|_{x^*} = [-1.414, -1.414]$$



and are shown in Figure E8.1b. The gradient of the objective function $\nabla f(\mathbf{x}^*)$ is orthogonal to the tangent plane of the constraint at $\mathbf{x}^*$. In general $\nabla h(\mathbf{x}^*)$ is *always* orthogonal to this tangent plane, hence $\nabla f(\mathbf{x}^*)$ and $\nabla h(\mathbf{x}^*)$ are collinear, that is, they lie on the same line but point in opposite directions. This means the two vectors must be multiples of each other;

$$\nabla f(\mathbf{x}^*) = \lambda^* \nabla h(\mathbf{x}^*) \qquad (a)$$

where $\lambda^* = -1/1.414$ is called the *Lagrange multiplier* for the constraint $h = 0$.



must hold at *any* local optimum of *any* equality-constrained NLP involving smooth functions.

MV Optn w Constraints ...

- **Method of Lagrange Multipliers**

Minimize $f(x_1, x_2)$

subject to

$g(x_1, x_2) = 0$

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2)$$

By treating L as a function of the three variables x_1, x_2, and λ, the necessary conditions for its extremum are given by

$$\frac{\partial L}{\partial x_1}(x_1, x_2, \lambda) = \frac{\partial f}{\partial x_1}(x_1, x_2) + \lambda \frac{\partial g}{\partial x_1}(x_1, x_2) = 0$$

$$\frac{\partial L}{\partial x_2}(x_1, x_2, \lambda) = \frac{\partial f}{\partial x_2}(x_1, x_2) + \lambda \frac{\partial g}{\partial x_2}(x_1, x_2) = 0 \qquad (2.38)$$

$$\frac{\partial L}{\partial \lambda}(x_1, x_2, \lambda) = g(x_1, x_2) = 0$$

LM w Eq Constraints

Example 2.9 Find the solution of Example 2.7 using the Lagrange multiplier method:

$$\text{Minimize } f(x, y) = kx^{-1}y^{-2}$$

subject to

$$g(x, y) = x^2 + y^2 - a^2 = 0$$

LM w Eq Constraints ...

Necessary Conditions

$$\text{Minimize } f(\mathbf{X})$$

subject to

$$g_j(\mathbf{X}) = 0, \quad j = 1, 2, \ldots, m$$

$$L(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m)$$

n+m variables

$$= f(\mathbf{X}) + \lambda_1 g_1(\mathbf{X}) + \lambda_2 g_2(\mathbf{X}) + \cdots + \lambda_m g_m(\mathbf{X})$$

n+m equations

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \qquad i = 1, 2, \ldots, n$$

$$\frac{\partial L}{\partial \lambda_j} = g_j(\mathbf{X}) = 0, \qquad j = 1, 2, \ldots, m$$

$$\mathbf{X}^* = \begin{Bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{Bmatrix} \quad \text{and} \quad \lambda^* = \begin{Bmatrix} \lambda_1^* \\ \lambda_2^* \\ \vdots \\ \lambda_m^* \end{Bmatrix}$$

LM w Eq Constraints ...

Theorem 2.6 Sufficient Condition A sufficient condition for $f(\mathbf{X})$ to have a relative minimum at $\mathbf{X}^*$ is that the quadratic, Q, defined by

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 L}{\partial x_i \, \partial x_j} (\mathbf{X}^*, \boldsymbol{\lambda}^*) dx_i \, dx_j$$

evaluated at $\mathbf{X} = \mathbf{X}^*$ must be positive definite for all values of $d\mathbf{X}$ for which the constraints are satisfied.

X*, λ* coming from necessary conditions

Q would be negative definite for f(X) to be relative maximum

LM w Eq Constraints ...

It has been shown by Hancock [2.1] that a necessary condition for the quadratic form Q, defined by Eq. (2.43), to be positive (negative) definite for all admissible variations $d\mathbf{X}$ is that each root of the polynomial z_i, defined by the following determinantal equation, be positive (negative):

$$\begin{vmatrix} L_{11}-z & L_{12} & L_{13} & \cdots & L_{1n} & g_{11} & g_{21} & \cdots & g_{m1} \\ L_{21} & L_{22}-z & L_{23} & \cdots & L_{2n} & g_{12} & g_{22} & \cdots & g_{m2} \\ \vdots & & & & & & & & \\ L_{n1} & L_{n2} & L_{n3} & \cdots & L_{nn}-z & g_{1n} & g_{2n} & \cdots & g_{mn} \\ g_{11} & g_{12} & g_{13} & \cdots & g_{1n} & 0 & 0 & \cdots & 0 \\ g_{21} & g_{22} & g_{23} & \cdots & g_{2n} & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & & & \\ g_{m1} & g_{m2} & g_{m3} & \cdots & g_{mn} & 0 & 0 & \cdots & 0 \end{vmatrix} = 0 \qquad (2.44)$$

where

$$L_{ij} = \frac{\partial^2 L}{\partial x_i \, \partial x_j}(\mathbf{X}^*, \boldsymbol{\lambda}^*) \qquad (2.45)$$

$$g_{ij} = \frac{\partial g_i}{\partial x_j}(\mathbf{X}^*) \qquad (2.46)$$

If some of the roots are positive and some of them are negative, it is not an extreme point

LM w Eq Constraints ...

Example 2.10 Find the dimensions of a cylindrical tin (with top and bottom) made up of sheet metal to maximize its volume such that the total surface area is equal to $A_0 = 24\pi$.

LM w Eq Constraints ...

Equations (E_1) and (E_2) lead to

$$\lambda = -\frac{x_1 x_2}{2x_1 + x_2} = -\frac{1}{2}x_1$$

that is,

$$x_1 = \tfrac{1}{2}x_2 \qquad (E_4)$$

and Eqs. (E_3) and (E_4) give the desired solution as

$$x_1^* = \left(\frac{A_0}{6\pi}\right)^{1/2}, \; x_2^* = \left(\frac{2A_0}{3\pi}\right)^{1/2}, \text{ and } \lambda^* = -\left(\frac{A_0}{24\pi}\right)^{1/2}$$

This gives the maximum value of f as

$$f^* = \left(\frac{A_0^3}{54\pi}\right)^{1/2}$$

If $A_0 = 24\pi$, the optimum solution becomes

$$x_1^* = 2, \quad x_2^* = 4, \quad \lambda^* = -1, \quad \text{and} \quad f^* = 16\pi$$

To see that this solution really corresponds to the maximum of f, we apply the sufficiency condition of Eq. (2.44). In this case

$$L_{11} = \left.\frac{\partial^2 L}{\partial x_1^2}\right|_{(\mathbf{X}^*, \lambda^*)} = 2\pi x_2^* + 4\pi\lambda^* = 4\pi$$

$$L_{12} = \left.\frac{\partial^2 L}{\partial x_1 \partial x_2}\right|_{(\mathbf{X}^*, \lambda^*)} = L_{21} = 2\pi x_1^* + 2\pi\lambda^* = 2\pi$$

LM w Eq Constraints ...

$$L_{22} = \left.\frac{\partial^2 L}{\partial x_2^2}\right|_{(\mathbf{X}^*, \lambda^*)} = 0$$

$$g_{11} = \left.\frac{\partial g_1}{\partial x_1}\right|_{(\mathbf{X}^*, \lambda^*)} = 4\pi x_1^* + 2\pi x_2^* = 16\pi$$

$$g_{12} = \left.\frac{\partial g_1}{\partial x_2}\right|_{(\mathbf{X}^*, \lambda^*)} = 2\pi x_1^* = 4\pi$$

Thus Eq. (2.44) becomes

$$\begin{vmatrix} 4\pi - z & 2\pi & 16\pi \\ 2\pi & 0 - z & 4\pi \\ 16\pi & 4\pi & 0 \end{vmatrix} = 0$$

that is,

$$272\pi^2 z + 192\pi^3 = 0$$

This gives

$$z = -\tfrac{12}{17}\pi$$

Since the value of z is negative, the point (x_1^*, x_2^*) corresponds to the maximum of f.

LM Sensitivity

$$\text{Minimize } f(\mathbf{X})$$

subject to

$$\underset{\sim}{g}(\mathbf{X}) = b \quad \text{or} \quad g(\mathbf{X}) = b - \underset{\sim}{g}(\mathbf{X}) = 0$$

where b is a constant. The necessary conditions to be satisfied f
problem are

$$\frac{\partial f}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i} = 0, \qquad i = 1, 2, \ldots, n$$

$$g = 0$$

$$db = d\underset{\sim}{g} = \sum_{i=1}^{n} \frac{\partial \underset{\sim}{g}}{\partial x_i} dx_i$$

Equation (2.49) can be rewritten as

$$\frac{\partial f}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i} = \frac{\partial f}{\partial x_i} - \lambda \frac{\partial \underset{\sim}{g}}{\partial x_i} = 0$$

or

$$\frac{\partial \underset{\sim}{g}}{\partial x_i} = \frac{\partial f / \partial x_i}{\lambda}, \qquad i = 1, 2, \ldots, n$$

Substituting Eq. (2.53) into Eq. (2.51), we obtain

$$db = \sum_{i=1}^{n} \frac{1}{\lambda} \frac{\partial f}{\partial x_i} dx_i = \frac{df}{\lambda}$$

since

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$

Equation (2.54) gives

$$\lambda = \frac{df}{db} \quad \text{or} \quad \lambda^* = \frac{df^*}{db}$$

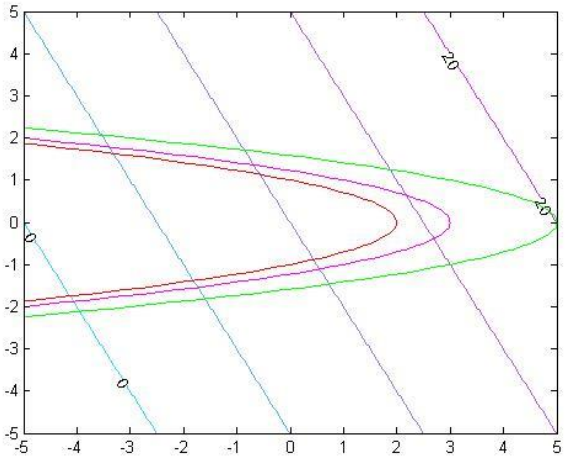
or

$$df^* = \lambda^* db$$

LM Sensitivity ...

Example 2.11 Find the maximum of the function $f(\mathbf{X}) = 2x_1 + x_2 + 10$ subject to $g(\mathbf{X}) = x_1 + 2x_2^2 = 3$ using the Lagrange multiplier method. Also find the effect of changing the right-hand side of the constraint on the optimum value of f.

LM Sensitivity ...



The application of the sufficiency condition of Eq. (2.44) yields

$$\begin{vmatrix} L_{11}-z & L_{12} & g_{11} \\ L_{21} & L_{22}-z & g_{12} \\ g_{11} & g_{12} & 0 \end{vmatrix} = 0$$

$$\begin{vmatrix} -z & 0 & -1 \\ 0 & -4\lambda - z & -4x_2 \\ -1 & -4x_2 & 0 \end{vmatrix} = \begin{vmatrix} -z & 0 & -1 \\ 0 & -8-z & -0.52 \\ -1 & -0.52 & 0 \end{vmatrix} = 0$$

$$0.2704z + 8 + z = 0$$

$$z = -6.2972$$

Hence $\mathbf{X}^*$ will be a maximum of f with $f^* = f(\mathbf{X}^*) = 16.07$.

LM w Inq Cons

Minimize $f(\mathbf{X})$

$$g_j(\mathbf{X}) \le 0, \quad j = 1, 2, \ldots, m$$

Minimize $f(\mathbf{X})$

subject to

$$G_j(\mathbf{X}, \mathbf{Y}) = g_j(\mathbf{X}) + y_j^2 = 0, \qquad j = 1, 2, \ldots, m$$

where $\mathbf{Y} = \{y_1, y_2, \ldots, y_m\}^{\mathrm{T}}$ is the vector of slack variables.

$$L(\mathbf{X}, \mathbf{Y}, \boldsymbol{\lambda}) = f(\mathbf{X}) + \sum_{j=1}^{m} \lambda_j G_j(\mathbf{X}, \mathbf{Y})$$

(necessary conditions):

$$\frac{\partial L}{\partial x_i}(\mathbf{X}, \mathbf{Y}, \boldsymbol{\lambda}) = \frac{\partial f}{\partial x_i}(\mathbf{X}) + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(\mathbf{X}) = 0, \qquad i = 1, 2, \ldots, n$$

$$\frac{\partial L}{\partial \lambda_j}(\mathbf{X}, \mathbf{Y}, \boldsymbol{\lambda}) = G_j(\mathbf{X}, \mathbf{Y}) = g_j(\mathbf{X}) + y_j^2 = 0, \qquad j = 1, 2, \ldots, m$$

$$\boxed{\frac{\partial L}{\partial y_j}(\mathbf{X}, \mathbf{Y}, \boldsymbol{\lambda}) = 2\lambda_j y_j = 0,} \qquad j = 1, 2, \ldots, m$$

n+2m equations with n+2m unknowns

X*, λ*, y* coming from necessary conditions

LM w Inq Cons ...

Either $\lambda^*_j = 0$
$y^*_j \neq 0 \Rightarrow$ Constraint INACTIVE (J2)

Or $y^*_j = 0$
$\lambda^*_j \neq 0 \Rightarrow$ Constraint ACTIVE (J1)

(necessary conditions):

$$\frac{\partial L}{\partial x_i}(\mathbf{X}, \mathbf{Y}, \boldsymbol{\lambda}) = \frac{\partial f}{\partial x_i}(\mathbf{X}) + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(\mathbf{X}) = 0, \qquad i = 1, 2, \ldots, n$$

$$\frac{\partial L}{\partial \lambda_j}(\mathbf{X}, \mathbf{Y}, \boldsymbol{\lambda}) = G_j(\mathbf{X}, \mathbf{Y}) = g_j(\mathbf{X}) + y_j^2 = 0, \qquad j = 1, 2, \ldots, m$$

$$\frac{\partial L}{\partial y_j}(\mathbf{X}, \mathbf{Y}, \boldsymbol{\lambda}) = 2\lambda_j y_j = 0, \qquad j = 1, 2, \ldots, m$$

$$\frac{\partial f}{\partial x_i} + \sum_{j \in J_1} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \qquad i = 1, 2, \ldots, n$$

Similarly, Eqs. (2.63) can be written as

$$g_j(\mathbf{X}) = 0, \qquad j \in J_1$$

$$g_j(\mathbf{X}) + y_j^2 = 0, \qquad j \in J_2$$

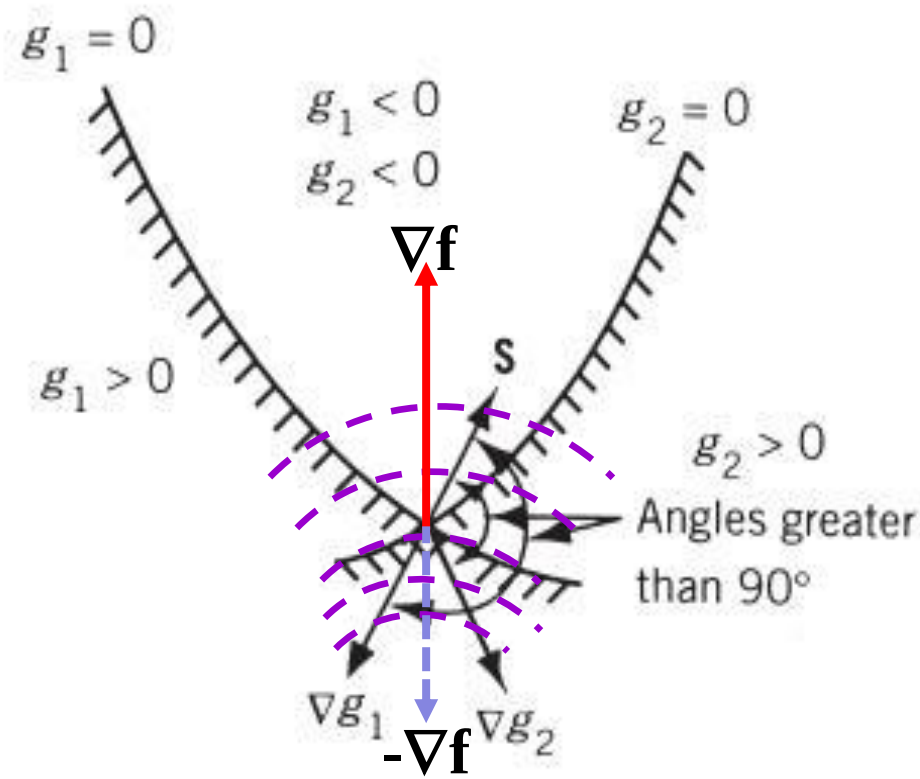
n+m equations with
n+m unknowns

LM w Inq Cons ...

$$-\frac{\partial f}{\partial x_i} = \lambda_1 \frac{\partial g_1}{\partial x_i} + \lambda_2 \frac{\partial g_2}{\partial x_i} + \ldots + \lambda_p \frac{\partial g_p}{\partial x_i}, \qquad i = 1, 2, \ldots, n$$

$$-\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 + \cdot \cdot \cdot + \lambda_p \nabla g_p$$

$$\nabla f = \begin{Bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{Bmatrix} \quad \text{and} \quad \nabla g_j = \begin{Bmatrix} \partial g_j/\partial x_1 \\ \partial g_j/\partial x_2 \\ \vdots \\ \partial g_j/\partial x_n \end{Bmatrix}$$



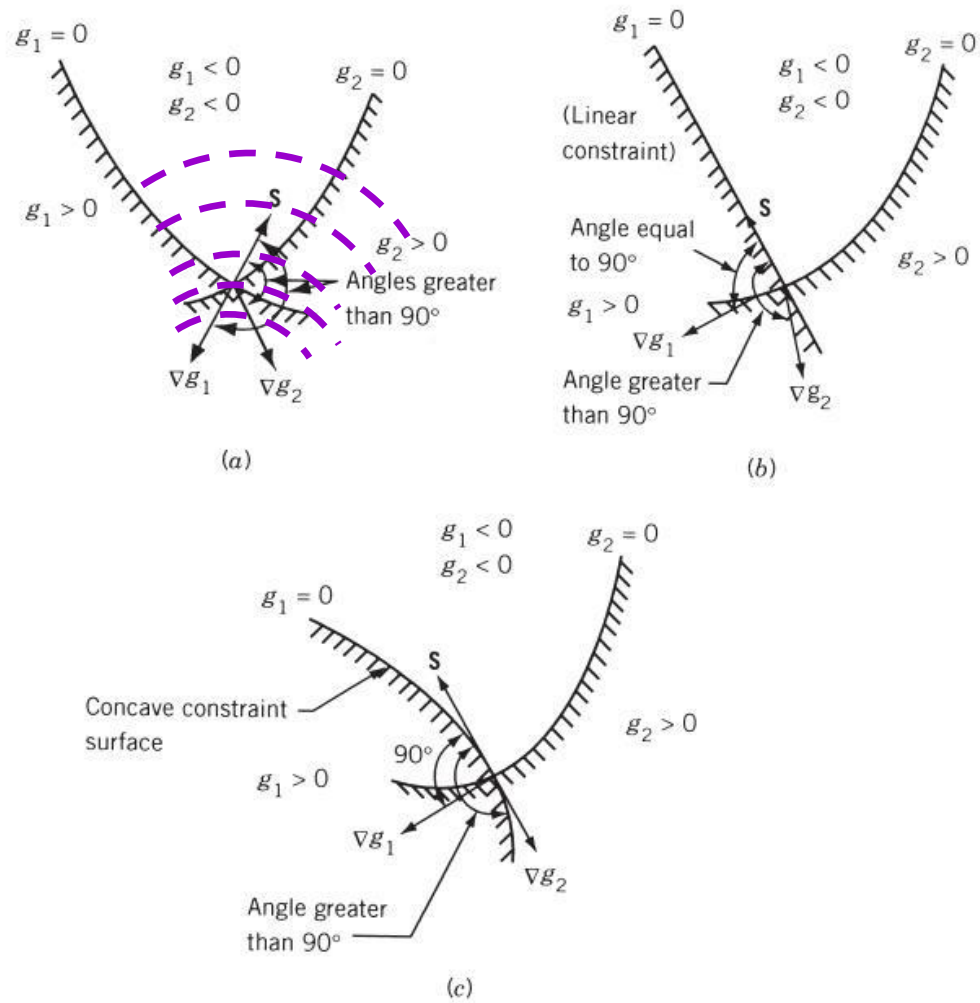
LM w Inq Cons ...

A vector **S** is called a *feasible direction* from a point **X** if at least a small step can be taken along **S** that does not immediately leave the feasible region. Thus for problems with sufficiently smooth constraint surfaces, vector **S** satisfying the relation

$$\mathbf{S}^T \nabla g_j < 0$$

can be called a feasible direction. On the other hand, if the constraint is either linear or concave, as shown in Fig. 2.8b and *c*, any vector satisfying the relation

$$\mathbf{S}^T \nabla g_j \leq 0$$



LM w Inq Cons ...

$$-\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 \quad \Rightarrow \quad -\mathbf{S}^T \nabla f = \lambda_1 \mathbf{S}^T \nabla g_1 + \lambda_2 \mathbf{S}^T \nabla g_2$$

λ_i are positive for minimization problems
λ_i are negative for maximization problems

Sationarity

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \qquad i = 1, 2, \ldots, n$$

Complementary slackness

$$\lambda_j g_j = 0,^{\ddagger} \qquad j = 1, 2, \ldots, m$$

Primal Feasibility

$$g_j \le 0, \qquad j = 1, 2, \ldots, m$$

Dual Feasibility

$$\lambda_j \ge 0, \qquad j = 1, 2, \ldots, m$$

KKT Conditions

Optimality Necessary Condition

Minimization problems & $g_j \le 0 \Rightarrow \lambda_j \ge 0$
Minimization problems & $g_j \ge 0 \Rightarrow \lambda_j \le 0$
Maximization problems & $g_j \le 0 \Rightarrow \lambda_j \le 0$
Maximization problems & $g_j \ge 0 \Rightarrow \lambda_j \ge 0$

Convex Programming

In convex programming, objective function and constraints are convex

Minimize $f(\mathbf{X})$

$$g_j(\mathbf{X}) \le 0, \quad j = 1, 2, \ldots, m$$

Minimize $f(\mathbf{X})$

subject to

$$G_j(\mathbf{X}, \mathbf{Y}) = g_j(\mathbf{X}) + y_j^2 = 0, \qquad j = 1, 2, \ldots, m$$

where $\mathbf{Y} = \{y_1, y_2, \ldots, y_m\}^T$ is the vector of slack variables.

$$L(\mathbf{X}, \mathbf{Y}, \lambda) = f(\mathbf{X}) + \sum_{j=1}^{m} \lambda_j [g_j(\mathbf{X}) + y_j^2]$$

KKT condition $\Rightarrow \lambda_j y_j = 0$

For $\lambda_j > 0$, $\lambda_j g_j$ is convex as g_j is convex $\Rightarrow$ L (X, Y, λ) is convex

As Lagrange function convex, there will be only 1 optima $\Rightarrow$ local optima will be the global optima & necessary condition to find that will be the sufficient condition also

Useable / Feasible

$$\text{Minimize } f(x_1, x_2) = x_1^2 + x_2^2$$

subject to

$$x_1 + 2x_2 \leq 15$$

$$1 \leq x_i \leq 10; \;\; i = 1, 2$$

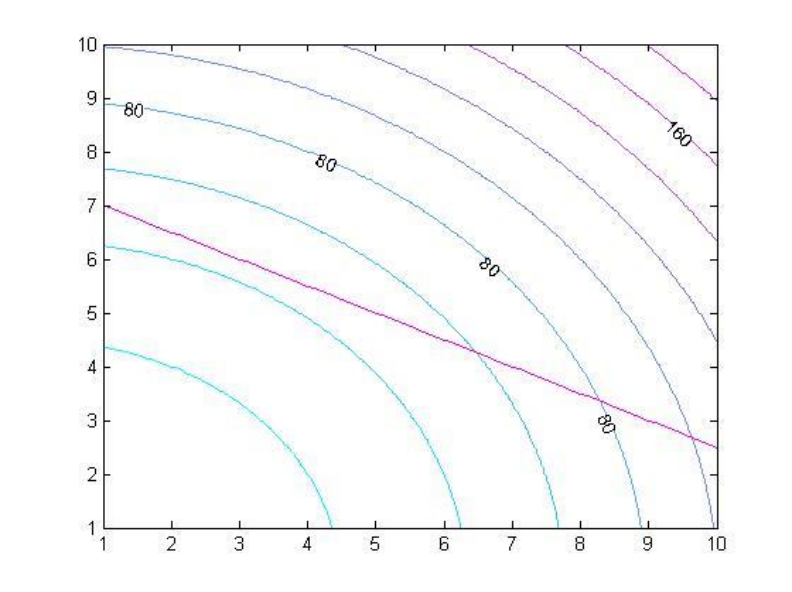
Derive the conditions to be satisfied at the point $\mathbf{X}_1 = \{1, 7\}^T$ by the search direction $\mathbf{S} = \{s_1, s_2\}^T$ if it is a (a) usable direction, and (b) feasible direction.

$$g_2(\mathbf{X}) = 1 - x_1 \leq 0$$

$$g_3(\mathbf{X}) = 1 - x_2 \leq 0$$

$$g_4(\mathbf{X}) = x_1 - 10 \leq 0$$

$$g_5(\mathbf{X}) = x_2 - 10 \leq 0$$



Useable / Feasible ...

point $\mathbf{X}_1 = \{1, 7\}^T$ are given by

$$\nabla f = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{Bmatrix}_{\mathbf{X}_1} = \begin{Bmatrix} 2x_1 \\ 2x_2 \end{Bmatrix}_{\mathbf{X}_1} = \begin{Bmatrix} 2 \\ 14 \end{Bmatrix}$$

$$\nabla g_1 = \begin{Bmatrix} \dfrac{\partial g_1}{\partial x_1} \\ \dfrac{\partial g_1}{\partial x_2} \end{Bmatrix}_{\mathbf{X}_1} = \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}$$

$$\nabla g_2 = \begin{Bmatrix} \dfrac{\partial g_2}{\partial x_1} \\ \dfrac{\partial g_2}{\partial x_2} \end{Bmatrix}_{\mathbf{X}_1} = \begin{Bmatrix} -1 \\ 0 \end{Bmatrix}$$

For the search direction $\mathbf{S} = \{s_1, s_2\}^T$, the usability and feasibility conditions can be expressed as

(a) Usability condition:

$$\mathbf{S}^T \nabla f \le 0 \quad \text{or} \quad (s_1 \quad s_2) \begin{Bmatrix} 2 \\ 14 \end{Bmatrix} \le 0 \quad \text{or} \quad 2s_1 + 14s_2 \le 0 \qquad (E_1)$$

(b) Feasibility conditions:

$$\mathbf{S}^T \nabla g_1 \le 0 \quad \text{or} \quad (s_1 \quad s_2) \begin{Bmatrix} 1 \\ 2 \end{Bmatrix} \le 0 \quad \text{or} \quad s_1 + 2s_2 \le 0 \qquad (E_2)$$

$$\mathbf{S}^T \nabla g_2 \le 0 \quad \text{or} \quad (s_1 \quad s_2) \begin{Bmatrix} -1 \\ 0 \end{Bmatrix} \le 0 \quad \text{or} \quad -s_1 \le 0 \qquad (E_3)$$

Note: Any two numbers for s_1 and s_2 that satisfy the inequality (E_1) will constitute a usable direction $\mathbf{S}$. For example, $s_1 = 1$ and $s_2 = -1$ gives the usable direction $\mathbf{S} = \{1, -1\}^T$. This direction can also be seen to be a feasible direction because it satisfies the inequalities (E_2) and (E_3).

KKT Example

Minimize

$$f(x_1, x_2, x_3) = (x_1^2 + 1000) + (x_2^2 + 1000) + (x_3^2 + 1000) + 20(x_1 - 50) + 20(x_1 + x_2 - 100)$$
$$= x_1^2 + x_2^2 + x_3^2 + 40x_1 + 20x_2$$

The constraints can be stated as

$$g_1(x_1, x_2, x_3) = x_1 - 50 \geq 0$$
$$g_2(x_1, x_2, x_3) = x_1 + x_2 - 100 \geq 0$$
$$g_3(x_1, x_2, x_3) = x_1 + x_2 + x_3 - 150 \geq 0$$

$$\frac{\partial f}{\partial x_i} + \lambda_1 \frac{\partial g_1}{\partial x_i} + \lambda_2 \frac{\partial g_2}{\partial x_i} + \lambda_3 \frac{\partial g_3}{\partial x_i} = 0, \qquad i = 1, 2, 3$$

$$\lambda_j g_j = 0, \qquad j = 1, 2, 3$$

$$2x_1 + 40 + \lambda_1 + \lambda_2 + \lambda_3 = 0 \qquad (E_1)$$
$$2x_2 + 20 + \lambda_2 + \lambda_3 = 0 \qquad (E_2)$$
$$2x_3 + \lambda_3 = 0 \qquad (E_3)$$

$$\lambda_1(x_1 - 50) = 0 \qquad (E_4)$$
$$\lambda_2(x_1 + x_2 - 100) = 0 \qquad (E_5)$$
$$\lambda_3(x_1 + x_2 + x_3 - 150) = 0 \qquad (E_6)$$
$$g_j \geq 0, \qquad j = 1, 2, 3$$

KKT Example ...

Minimize

$$f(x_1, x_2, x_3) = (x_1^2 + 1000) + (x_2^2 + 1000) + (x_3^2 + 1000) + 20(x_1 - 50) + 20(x_1 + x_2 - 100)$$
$$= x_1^2 + x_2^2 + x_3^2 + 40x_1 + 20x_2$$

The constraints can be stated as

$$g_1(x_1, x_2, x_3) = x_1 - 50 \geq 0$$
$$g_2(x_1, x_2, x_3) = x_1 + x_2 - 100 \geq 0$$
$$g_3(x_1, x_2, x_3) = x_1 + x_2 + x_3 - 150 \geq 0$$

$$x_1 - 50 \geq 0 \qquad (E_7)$$
$$x_1 + x_2 - 100 \geq 0 \qquad (E_8)$$
$$x_1 + x_2 + x_3 - 150 \geq 0 \qquad (E_9)$$
$$\lambda_j \leq 0, \quad j = 1, 2, 3$$

$$\lambda_1 \leq 0 \qquad (E_{10})$$
$$\lambda_2 \leq 0 \qquad (E_{11})$$
$$\lambda_3 \leq 0 \qquad (E_{12})$$

KKT Example ...

$$2x_1 + 40 + \lambda_1 + \lambda_2 + \lambda_3 = 0 \qquad (E_1)$$

$$2x_2 + 20 + \lambda_2 + \lambda_3 = 0 \qquad (E_2)$$

$$2x_3 + \lambda_3 = 0 \qquad (E_3)$$

$$\lambda_1(x_1 - 50) = 0 \qquad (E_4)$$

$$\lambda_2(x_1 + x_2 - 100) = 0 \qquad (E_5)$$

$$\lambda_3(x_1 + x_2 + x_3 - 150) = 0 \qquad (E_6)$$

$$x_1 - 50 \geq 0 \qquad (E_7)$$

$$x_1 + x_2 - 100 \geq 0 \qquad (E_8)$$

$$x_1 + x_2 + x_3 - 150 \geq 0 \qquad (E_9)$$

$$\lambda_1 \leq 0 \qquad (E_{10})$$

$$\lambda_2 \leq 0 \qquad (E_{11})$$

$$\lambda_3 \leq 0 \qquad (E_{12})$$

Case 1: $\boldsymbol{\lambda_1 = 0}$.

Equations (E_1) to (E_3) give

$$x_3 = -\frac{\lambda_3}{2}$$

$$x_2 = -10 - \frac{\lambda_2}{2} - \frac{\lambda_3}{2}$$

$$x_1 = -20 - \frac{\lambda_2}{2} - \frac{\lambda_3}{2}$$

Substituting Eqs. (E_{13}) in Eqs. (E_5) and (E_6), we obtain

$$\lambda_2(-130 - \lambda_2 - \lambda_3) = 0$$

$$\lambda_3(-180 - \lambda_2 - \tfrac{3}{2}\lambda_3) = 0$$

KKT Example ...

Case 1: $\lambda_1 = 0$.

Equations (E_1) to (E_3) give

$$x_3 = -\frac{\lambda_3}{2}$$

$$x_2 = -10 - \frac{\lambda_2}{2} - \frac{\lambda_3}{2}$$

$$x_1 = -20 - \frac{\lambda_2}{2} - \frac{\lambda_3}{2}$$

Substituting Eqs. (E_{13}) in Eqs. (E_5) and (E_6), we obtain

$$\lambda_2(-130 - \lambda_2 - \lambda_3) = 0$$

$$\lambda_3(-180 - \lambda_2 - \tfrac{3}{2}\lambda_3) = 0$$

$$x_1 - 50 \geq 0 \qquad (E_7)$$

$$x_1 + x_2 - 100 \geq 0 \qquad (E_8)$$

$$x_1 + x_2 + x_3 - 150 \geq 0 \qquad (E_9)$$

$$\lambda_1 \leq 0 \qquad (E_{10})$$

$$\lambda_2 \leq 0 \qquad (E_{11})$$

$$\lambda_3 \leq 0 \qquad (E_{12})$$

The four possible solutions of Eqs. (E_{14}) are

1. $\lambda_2 = 0$, $-180 - \lambda_2 - \frac{3}{2}\lambda_3 = 0$. These equations, along with Eqs. (E_{13}), yield the solution

2. $\lambda_3 = 0$, $-130 - \lambda_2 - \lambda_3 = 0$. The solution of these equations leads to

3. $\lambda_2 = 0$, $\lambda_3 = 0$. Equations (E_{13}) give

4. $-130 - \lambda_2 - \lambda_3 = 0$, $-180 - \lambda_2 - \frac{3}{2}\lambda_3 = 0$. The solution of these equations and Eqs. (E_{13}) yields

KKT Example ...

$$2x_1 + 40 + \lambda_1 + \lambda_2 + \lambda_3 = 0 \qquad (E_1)$$

$$2x_2 + 20 + \lambda_2 + \lambda_3 = 0 \qquad (E_2)$$

$$2x_3 + \lambda_3 = 0 \qquad (E_3)$$

$$\lambda_1(x_1 - 50) = 0 \qquad (E_4)$$

$$\lambda_2(x_1 + x_2 - 100) = 0 \qquad (E_5)$$

$$\lambda_3(x_1 + x_2 + x_3 - 150) = 0 \qquad (E_6)$$

$$x_1 - 50 \geq 0 \qquad (E_7)$$

$$x_1 + x_2 - 100 \geq 0 \qquad (E_8)$$

$$x_1 + x_2 + x_3 - 150 \geq 0 \qquad (E_9)$$

$$\lambda_1 \leq 0 \qquad (E_{10})$$

$$\lambda_2 \leq 0 \qquad (E_{11})$$

$$\lambda_3 \leq 0 \qquad (E_{12})$$

Case 2: $x_1 = 50$.

Substitution of Eqs. (E_{15}) in Eqs. (E_5) and (E_6) leads to

KKT Example ...

Case 2: $x_1 = 50$.

In this case, Eqs. (E_1) to (E_3) give

$$\lambda_3 = -2x_3$$
$$\lambda_2 = -20 - 2x_2 - \lambda_3 = -20 - 2x_2 + 2x_3$$
$$\lambda_1 = -40 - 2x_1 - \lambda_2 - \lambda_3 = -120 + 2x_2$$

Substitution of Eqs. (E_{15}) in Eqs. (E_5) and (E_6) leads to

$$(-20 - 2x_2 + 2x_3)(x_1 + x_2 - 100) = 0$$
$$(-2x_3)(x_1 + x_2 + x_3 - 150) = 0$$

$$x_1 - 50 \geq 0 \qquad (E_7)$$
$$x_1 + x_2 - 100 \geq 0 \qquad (E_8)$$
$$x_1 + x_2 + x_3 - 150 \geq 0 \qquad (E_9)$$

$$\lambda_1 \leq 0 \qquad (E_{10})$$
$$\lambda_2 \leq 0 \qquad (E_{11})$$
$$\lambda_3 \leq 0 \qquad (E_{12})$$

Once again, it can be seen that there are four possible solutions to Eqs. (E_{16}), as indicated below:

1. $-20 - 2x_2 + 2x_3 = 0,\ x_1 + x_2 + x_3 - 150 = 0$: The solution of these equations yields

2. $-20 - 2x_2 + 2x_3 = 0,\ -2x_3 = 0$: These equations lead to the solution

3. $x_1 + x_2 - 100 = 0,\ -2x_3 = 0$: These equations give

4. $x_1 + x_2 - 100 = 0,\ x_1 + x_2 + x_3 - 150 = 0$: The solution of these equations yields