Optimization Techniques
Lecture 3
Hale Gonce Köçken

Unconstrained Optimization,

Local and Global Minimizer,

Feasible Direction,

FONC and SOSC for a local minimizer

Unconstrained Optimization,

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega. \end{array}$$

The function $f : \mathbb{R}^n \to \mathbb{R}$ that we wish to minimize is a real-valued function, and is called the *objective function*, or *cost function*. The vector $\boldsymbol{x}$ is an n-vector of independent variables, that is, $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$. The variables $x_1, \ldots, x_n$ are often referred to as *decision variables*. The set Ω is a subset of $\mathbb{R}^n$, called the *constraint set* or *feasible set*.

The above problem is a general form of a *constrained* optimization problem, because the decision variables are constrained to be in the constraint set Ω. If $\Omega = \mathbb{R}^n$, then we refer to the problem as an *unconstrained* optimization problem.

The constraint "$\boldsymbol{x} \in \Omega$" is called a *set constraint*. Often, the constraint set Ω takes the form $\Omega = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$, where $\boldsymbol{h}$ and $\boldsymbol{g}$ are given functions. We refer to such constraints as *functional constraints*. The remainder of this chapter deals with general set constraints, including the special case where $\Omega = \mathbb{R}^n$. The case where $\Omega = \mathbb{R}^n$ is called the *unconstrained* case.

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega. \end{array}$$

$$minimizer \text{ of } f \ over\ \Omega$$
$$maximizing\ f \text{ is equivalent to } minimizing\ -f$$

Max versus min: Maximizing a function g is equivalent to minimizing −g, so there's no loss of generality in concentrating on minimization. This is the convention in much of optimization theory

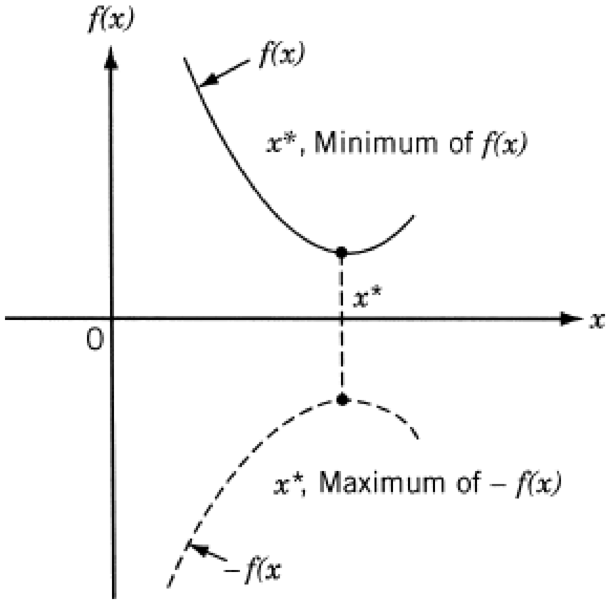


Figure 1.1 Minimum of $f(x)$ is same as maximum of $-f(x)$.

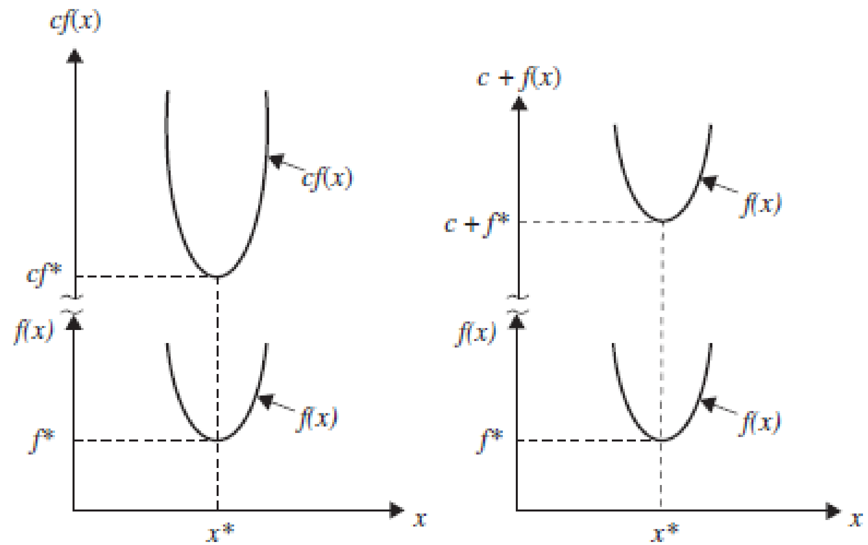


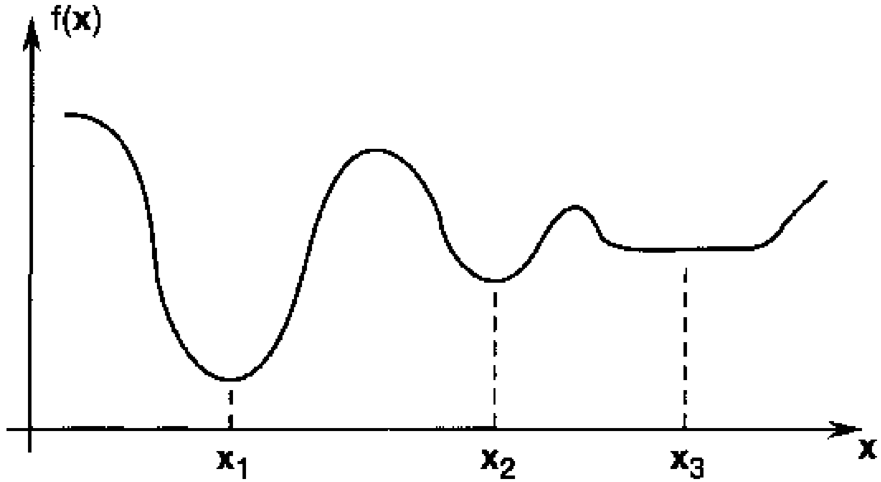
Figure 1.2 Optimum solution of $cf(x)$ or $c + f(x)$ same as that of $f(x)$.

Local and Global Minimizers

Definition 6.1 ***Local minimizer.*** Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function defined on some set $\Omega \subset \mathbb{R}^n$. A point $x^* \in \Omega$ is a *local minimizer* of f over Ω if there exists $\varepsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \Omega \setminus \{x^*\}$ and $\|x - x^*\| < \varepsilon$.

Global minimizer. A point $x^* \in \Omega$ is a *global minimizer* of f over Ω if $f(x) \geq f(x^*)$ for all $x \in \Omega \setminus \{x^*\}$.

If, in the above definitions, we replace "$\geq$" with "$>$", then we have a *strict local minimizer* and a *strict global minimizer*, respectively.



x_1: strict global minimizer;

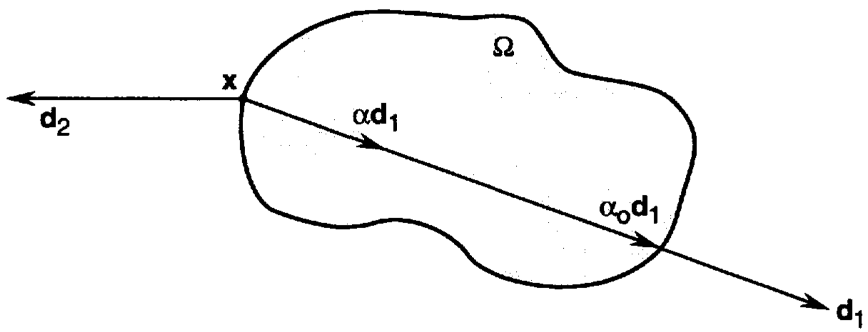
x_2: strict local minimizer;

x_3: local (not strict) minimizer

Feasible Direction

Given an optimization problem with constraint set Ω, a minimizer may lie either in the interior or on the boundary of Ω. To study the case where it lies on the boundary, we need the notion of *feasible directions.*

Definition 6.2 ***Feasible direction.*** A vector $\boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d} \neq \boldsymbol{0}$, is a *feasible direction* at $\boldsymbol{x} \in \Omega$ if there exists $\alpha_0 > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$.



$\boldsymbol{d}_1$ is a feasible direction,

$\boldsymbol{d}_2$ is not a feasible direction

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function and let $\boldsymbol{d}$ be a feasible direction at $\boldsymbol{x} \in \Omega$. The *directional derivative of f in the direction* $\boldsymbol{d}$, denoted $\partial f / \partial \boldsymbol{d}$, is the real-valued function defined by

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \lim_{\alpha \to 0} \frac{f(\boldsymbol{x} + \alpha \boldsymbol{d}) - f(\boldsymbol{x})}{\alpha}.$$

If $\|\boldsymbol{d}\| = 1$, then $\partial f/\partial \boldsymbol{d}$ is the rate of increase of f at $\boldsymbol{x}$ in the direction $\boldsymbol{d}$. To compute the above directional derivative, suppose that $\boldsymbol{x}$ and $\boldsymbol{d}$ are given. Then, $f(\boldsymbol{x} + \alpha \boldsymbol{d})$ is a function of α, and

$$f(x + \propto d) = F(\propto)$$

$$F'(\propto) = \frac{\delta f}{\delta x_1}\frac{dx_1}{d \propto} + \frac{\delta f}{\delta x_2}\frac{dx_2}{d \propto} + \cdots + \frac{\delta f}{\delta x_n}\frac{dx_n}{d \propto}$$

$$= \nabla f(x)^T . d$$

In summary, if $\boldsymbol{d}$ is a unit vector, that is, $\|\boldsymbol{d}\| = 1$, then $\langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \rangle$ is the rate of increase of f at the point $\boldsymbol{x}$ in the direction $\boldsymbol{d}$.

Theorem *First-Order Necessary Condition (FONC). Let Ω be a subset of $\mathbb{R}^n$ and $f \in C^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, and $\boldsymbol{x}^*$ is an interior point of Ω, then*

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}.$$

What happens if Ω is an open subset of R^n ?

FONC: *Let Ω be an open set and f is a continuously differentiable function over Ω ($f \in C^1$).*

If $x^* \in \Omega$ is a minimizer of f, then $\nabla f(x^*) = 0$.

$\nabla f(x^*) = 0.$

This is the **first-order necessary condition for optimality**. The condition is ``first-order" because it is derived using the first-order expansion. We emphasize that the result is valid when $f \in C^1$ and x^* is an interior point of Ω .

Definition: A point x^* satisfying the FONC condition $\nabla f(x^*) = 0$ is called a ***stationary point***.

Definition: A function *f* has **critical points** at all points x_0 where $f'(x_0) = 0$ or f(x) is not differentiable.

In several variables case, the same definition can be given as:

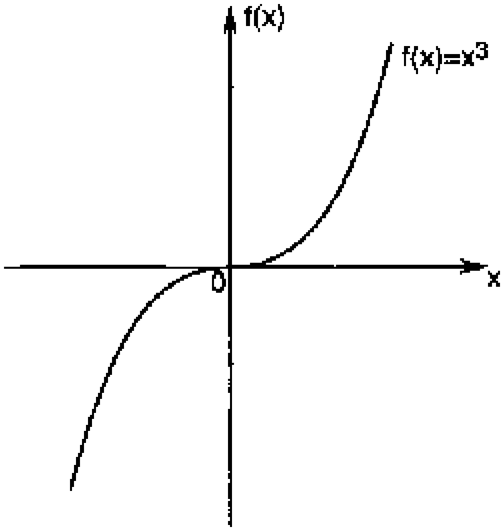
A function with several variables has **critical points** where the gradient is 0 or any of the partial derivatives is not defined.

Second Order Sufficient Condition (SOSC):

Let Ω be an open set and f is a twice continuously differentiable function over Ω ($f \in C^2$). The sufficient condition for a local minimizer ($x^* \in \Omega$) is that the Hessian matrix at x^* is positive semidefinite.

Proof of SOSC:

Example 6.5 Consider a function of one variable $f(x) = x^3$, $f : \mathbb{R} \to \mathbb{R}$.



The point 0 satisfies the FONC but is not a minimizer

This point is an inflection point.

Definition: An inflection point is a point on a curve at which the concavity changes.

A necessary condition for x to be an inflection point is $f''(x) = 0$. A sufficient condition requires $f''(x+\varepsilon)$ and $f''(x-\varepsilon)$ to have opposite signs in the neighborhood of x.

In multivariable calculus, a point of a function or surface which is a stationary point but not an extremum is called a saddle point.

In summary

FONC: Solve the following equation and obtain all the stationary points.

$\nabla f(x^*) = 0.$

SOSC: Form the Hessian at each of the stationary points:

If the Hessian is *positive* definite, then that point is a *strict* local minimum,

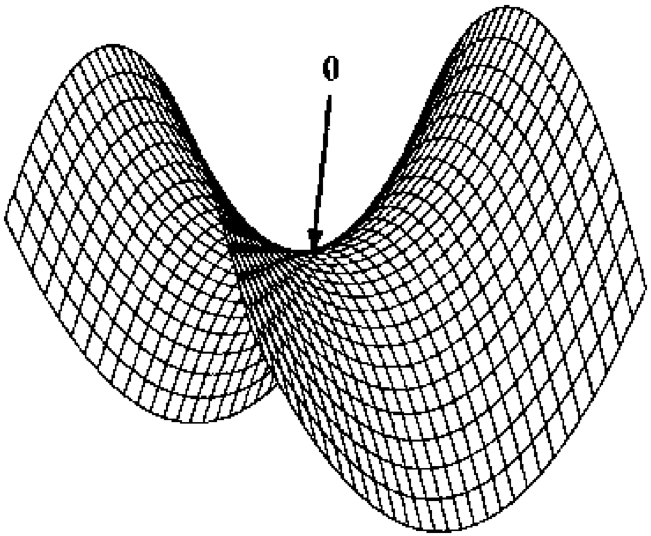
If the Hessian is *positive* semidefinite, then that point is a local minimum,

If the Hessian is *negative* definite, then that point is a *strict* local maximum,

If the Hessian is *negative* semidefinite, then that point is a local maximum.

If the Hessian is indefinite, then that point is a saddle point.

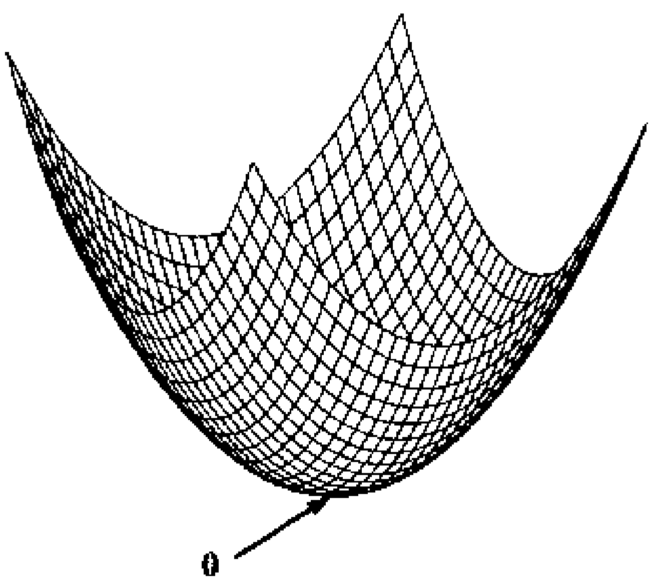
Example 6.6 Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(\boldsymbol{x}) = x_1^2 - x_2^2$.



The point $\mathbf{0}$ satisfies the FONC but not SOSC;

this point is not a minimizer

Example 6.7 Let $f(\boldsymbol{x}) = x_1^2 + x_2^2$.



$\boldsymbol{x} = [0, 0]^T$ is a strict global minimizer.

Example: A small startup company produces speakers and subwoofers for computers that they sell through a website. After extensive research, the company has developed a revenue function,

$R(x, y) = x(110 - 4.5x) + y(155 - 2y)$ thousand dollars

where x is the number of subwoofers produced and sold in thousands and y is the number of speakers produced and sold in thousands. The corresponding cost function is

$C(x, y) = 3x^2 + 3y^2 + 5xy - 5y + 50$ thousand dollars

Find the production levels that maximize revenue.

Find and classify all the critical points of the following functions:

1) $f(x_1, x_2) = x^3 - x^2y + 2y^2.$

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

3) $f(x, y) = 7x - 8y + 2xy - x^2 + y^3$

4) $f = x_1^3 + x_2^3 + x_3^3 - 3x_1^2 x_2 - 3x_2^2 x_3 - 3x_3^2 x_1 + 6x_1 + 6x_2 + 6x_3$

$$f(x,y) = x + \frac{1}{2x+y} - \frac{1}{x+v}$$

$$\nabla f = \begin{pmatrix} 1 - \frac{2}{(2x+y)^2} + \frac{1}{(x+y)^2} \\ -\frac{1}{(2x+y)^2} + \frac{1}{(x+y)^2} \end{pmatrix} \qquad H(f) = \begin{pmatrix} \frac{8}{(2x+y)^3} - \frac{2}{(x+y)^3}; & \frac{4}{(2x+y)^3} - \frac{2}{(x+y)^3} \\ \frac{4}{(2x+y)^3} - \frac{2}{(x+y)^3}; & \frac{2}{(2x+y)^3} - \frac{2}{(x+y)^3} \end{pmatrix}$$

Need to solve the equations $\nabla f(x,y) = 0$:

$$1 - \frac{2}{(2x+y)^2} + \frac{1}{(x+y)^2} = 0, \quad -\frac{1}{(2x+y)^2} + \frac{1}{(x+y)^2} = 0$$

Substracting the second equation from the first one we see that $(2x+y)^2 = 1$ and $(x+y)^2 = 1$ which give us 4 critical points

$$1)\ (0,1); \quad 2)\ (0,-1); \quad 3)\ (2,-3); \quad 4)\ (-2,3).$$

1) $H(f)|_{(0,1)} = \begin{pmatrix} 6 & 2 \\ 2 & 0 \end{pmatrix}$,

2) $H(f)|_{(0,-1)} = \begin{pmatrix} -6 & -2 \\ -2 & 0 \end{pmatrix}$

3) $H(f)|_{(2,-3)} = \begin{pmatrix} 10 & 6 \\ 6 & 4 \end{pmatrix}$,

4) $H(f)|_{(-2,3)} = \begin{pmatrix} -10 & -6 \\ -6 & -4 \end{pmatrix}$;

Practical Example

A container with an open top is to have 10 m^3 capacity and be made of thin sheet metal. Calculate the dimensions of the box if it is to use the minimum possible amount of metal.

Solution: Let A be the total area of metal used to make the box, and let x and y be the length and width and z the height. Then

$$A = 2xz + 2yz + xy$$

Also

$$xyz = 10$$

because the volume is 10 m^3. This implies that $z = \frac{10}{xy}$. Putting this into the formula for A gives A as a function of x and y only:

$$\begin{aligned} A &= 2x\left(\frac{10}{xy}\right) + 2y\left(\frac{10}{xy}\right) + xy \\ &= \frac{20}{y} + \frac{20}{x} + xy \end{aligned}$$

We shall apply our techniques to this function. Now

$$\frac{\partial A}{\partial x} = -\frac{20}{x^2} + y, \qquad \frac{\partial A}{\partial y} = -\frac{20}{y^2} + x$$

and for a stationary point we need $\partial A/\partial x = \partial A/\partial y = 0$. this gives

$$y = \frac{20}{x^2} \quad \text{and} \quad x = \frac{20}{y^2}.$$

Therefore

$$y = \frac{20}{(20/y^2)^2} = \frac{y^4}{20}$$

Since the zero root $y = 0$ is obviously not consistent with having a volume of 10 m^3 we reject $y = 0$ and conclude that $y^3 = 20$ so that $y = 20^{1/3} = 2.714$ metres. From $x = 20/y^2$ we conclude $x = 2.714$ metres also. To find z, use $z = \frac{10}{xy}$ so that $z = 1.357$ m.

We have to show that these values do indeed give a **minimum**. Now

$$\frac{\partial^2 A}{\partial x \partial y} = 1, \qquad \frac{\partial^2 A}{\partial x^2} = \frac{40}{x^3}, \qquad \frac{\partial^2 A}{\partial y^2} = \frac{40}{y^3}$$

So, when $(x, y) = (2.714, 2.714)$,

$$A_{xx}A_{yy} - A_{xy}^2 = (2)(2) - 1^2 = 3 > 0$$

so it is either a max or a min. But $A_{xx} > 0$ and $A_{yy} > 0$ so it is a minimum. Our conclusion is that the box should have length 2.714 m, width 2.714 m and height 1.357 m. The actual area of metal used will then (from the formula for A) be 22.1 m^2.