1. Optimization in $\mathbb{R}$

Problem I. *Let $f(x)$ be a function on some interval I of the real line $\mathbb{R}$. Find the points of I at which $f(x)$ achieves its maximum and minimum values*

Definition. *If*

$$f(\bar{x}) \leq f(x); \quad \textit{for all } x \in I, \tag{1.1}$$

with equality for $x = \bar{x} \in I$, then $f(x)$ achieves its minimum value at $\bar{x}$.

(1.1) means that $f(x)$ achieves its global (or absolute) minimum at the point $\bar{x}$.

(Using $g(x) = -f(x)$, we have a similar definition for a maximum, ie. $f(\bar{x}) \geq f(x)$ for all $x \in I$.)

Definition. *$f(x)$ has a local minimum at a point $\bar{x}$ if*

$$f(\bar{x}) \leq f(x); \quad \textit{for all } x \in N \subset I, \tag{1.2}$$

with equality for $x = \bar{x}$, where N is the ε-neighborhood $|x - \bar{x}| < \varepsilon$, with ε small.

Theorem 1.1. *Let $f(x)$ be defined on an open interval $I = (a, b)$, $a < b$ and let f be C^2 (continuous second derivative) in some ε-neighborhood of $\bar{x} \in (a, b)$. If $f(x)$ has a local minimum at $\bar{x}$, then $f'(\bar{x}) = 0$.*

Proof. If $f(\bar{x})$ is a local minimum, then there exists a neighborhood N such that

$$f(\bar{x}+h) - f(\bar{x}) \geq 0, \forall \bar{x}+h \in N,$$

with equality for $h = 0$.

Using a Taylor expansion, we have

$$f(\bar{x}+h) = f(\bar{x}) + hf'(\bar{x}) + \frac{h^2}{2!} f''(\bar{x}+\theta h), \quad 0 < \theta < 1,$$

where $2! = 1 \times 2 = 2$. Then we have

$$hf'(\bar{x}) + \frac{h^2}{2!} f''(\bar{x}+\theta h) \geq 0 \quad \text{in } N,$$

where $f''(\bar{x}+\theta h)$ is bounded.

When $h > 0$ we can deduce that

$$f'(\bar{x}) + \frac{h}{2} f''(\bar{x}+\theta h) \geq 0.$$

As $h \to 0$, we obtain $f'(\bar{x}) \geq 0$.

When $h < 0$, a similar argument yield $f'(\bar{x}) \leq 0$, so $f'(\bar{x}) = 0$ is a necessary condition for a local minimum. $\square$

We have two theorems on sufficient conditions of a local minimum.

Theorem 1.2. *Let $f(x)$ be defined on an open interval $I = (a, b)$ and let f be C^1 (continuous first derivative) in some ε-neighborhood of $\bar{x} \in (a, b)$. If $f'(\bar{x}) = 0$, then $f(x)$ has a local minimum at $\bar{x}$ provided $f'(\bar{x} + h) < 0$ for $h < 0$ and $f'(\bar{x} + h) > 0$ for $h > 0$.*

Example. Let $f(x) = x^2$ be a function on $(-1, 1)$. $f'(x) = 2x$ so $\bar{x} = 0$ is a point satisfying all condition in Theorem 1.2. Therefore f has a local minimum at $\bar{x} = 0$.

Theorem 1.3. *Let $f(x)$ be C^m (continuous first m-derivative with integer $m \geq 1$) in some ε-neighborhood of $\bar{x} \in (a, b)$. If*

$$f'(\bar{x}) = f''(\bar{x}) = \cdots = f^{(m-1)}(\bar{x}) = 0, \textit{ and } f^{(m)}(\bar{x}) \neq 0,$$

then $f(\bar{x})$ has a local minimum at $\bar{x}$ provided (i) m is even and (ii) $f^{(m)}(\bar{x}) > 0$.

Proof. The Taylor expansion yields

$$f(\bar{x} + h) - f(\bar{x}) = hf'(\bar{x}) + \cdots + \frac{h^m}{m!} f^{(m)}(\bar{x} + \theta h)$$

where $0 < \theta < 1$ and $m! = 1 \times 2 \times \cdots \times m$.

By assumption in Theorem 1.3, we have

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

Since $f^{(m)}$ is continuous near $\bar{x}$ and $f^{(m)}(\bar{x}) > 0$, we know that $f^{(m)}(\bar{x} + \theta h) > 0$ in some small h. Thus $f(t+h) - f(\bar{x}) > 0$ for small h. This means that f has a local minimum at $\bar{x}$. $\square$

Critical points, end-points and points of discontinuity

In the last section, we need to understand the point $\bar{x}$ with $f'(\bar{x}) = 0$. But sometime $f'(\bar{x})$ does not exists.

For example, $f(x) = |x|$, with $-\infty < x < \infty$, has a global minimum at $x = 0$, but $f'(0)$ doesn't exist.

We need to extend the theory to the case that $f(x)$ is not differentiable.

We use the term critical point to mean a point at which either $f'(x) = 0$ or $f'(x)$ does not exists.

Theorem 1.4. *Let $f(x)$ be defined on (a,b) and differentiable on (a,b) except at $\bar{x}$. Consider a deleted neighborhood of $\bar{x}$, that is the set of points $\bar{x}-\varepsilon < x < \bar{x}$ and $\bar{x} < x < \bar{x}+\varepsilon$ for some small positive ε. Then $f(x)$ has a local minimum at $\bar{x}$ provide $f'(x) < 0$ through $\bar{x}-\varepsilon < x < \bar{x}$ and $f'(x) > 0$ through $\bar{x} < x < \bar{x}+\varepsilon$.*

The natural of the interval I is important too. When $f(x)$ is defined on a closed interval $[a,b]$. Then $f(x)$ has a global maximum or minimum on I, possible achieve at the end points a and b. For example, let $f(x) = x$ on $[0,2]$ has a minimum at 0 and maximum at 2.

Consider $f(x) = x^3$ on $I = (-1, 2)$, then it has neither a global maximum nor a minimum, but if I is $[-1, 2]$, then the function has a global minimum at $x = -1$ and a global maximum at $x = 2$.

Finally, we shall assume that any $f(x)$ we need to deal with is piecewise continuous. If this is the case, the isolated points of discontinuity also need to be examined because they too could be global maxima or minima.

For example, consider

$$f(x) = \begin{cases} -x^3 & -\infty < x < 1 \\ x - 3 & 1 \leq x < \infty \end{cases}$$

has a global minimum at $x = 1$, where $f(x) = -2$.

If we assume that f and f' are continuous on I except at a finite number of points at which $f'(x)$ does not exists or $f(x)$ is discontinuous, the procedure to find a global minimum is

(i) Find the stationary points ($f'(x) = 0$) and use Theorem 1.2 to determine any local minima

(ii) Examine $f(x)$ near every point at which $f(x)$ or $f'(x)$ is discontinuous to see if it is a local minimum

(iii) Compare all these value of $f(x)$ to find which is the smallest;

(iv) If I is closed, evaluate $f(a)$ and $f(b)$ and compare them with the smallest values in (iii). This will find the minimum, if there is one.

Exercises

Find the global minimum (when it exists) of each of the following functions:

1. $f(x) = 2x^3 - 9x^2 + 12$, for $-\infty < x < \infty$
2. Prove Theorem 1.2.

2. Optimization in $\mathbb{R}^n$ with $n \geq 2$

In this section, we turn to deal with minimizing problem of a function of several variables.

Let $x = (x_1, x_2, ..., x_n)$ be n variables.

Problem II. *Let $f(x)$ be a function on some region K of $\mathbb{R}^n$. Find the points of K at which $f(x)$ achieves its maximum and minimum values.*

Definition. *Let $f(x)$ be a given function of $x = (x_1, ..., x_n)$ defined in K. We say that f has a global minimum at $x = a$, with $a \in K$ if*

$$f(a) \leq f(x); \quad \textit{for all } x \in K, \tag{2.1}$$

with equality only for $x = a$.

Similarly, we define a local minima. Let N be an ε_0-neighborhood of $a = (a_1, ..., a_n)$, (ie. $y = (y_1, ..., y_n) \in N$, $|y - a| = \sqrt{|y_1 - a_1|^2 + \cdots + |y_n - a_n|^2} \leq \varepsilon_0$)

For any point y in N and any nonzero vector h, there is some sufficiently small ε such that $y = a + \varepsilon h$.

Definition. *Let $f(x)$ be a given function of $x = (x_1, ..., x_n)$ defined in K. We say that f has a local minimum at $x = a$, with $a \in K$ if*

$$f(a) \leq f(y) \quad \textit{for all } h \textit{ and } \varepsilon \textit{ with} \quad y = a + \varepsilon h \in N, \tag{2.2}$$

Now assume that in N, f has continuous fist and second order partial derivatives with respect to all its variables and that the third order derivatives are bounded. For $a \in \mathbb{R}^n$, we recall that the gradient of $f(x)$ at a is

$$\text{grad } f(a) = \left(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)\right)^T$$

and that the Hessian matrix $H(a)$ of $f(x)$ is the $n \times n$ matrix whose entries are the second derivatives of f evaluated at a; i.e.

$$H(a) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{n \times n}.$$

Let h be a vector in $\mathbb{R}^n$ (as $n \times 1$-matrix) i.e.

$$h = \begin{pmatrix} h_1 \\ . \\ . \\ . \\ h_n \end{pmatrix}$$

Then $h^T = (h_1, ..., h_n)$ is the transpose of h.

Note that $h^T \, \text{grad} f(a) = \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(a)$ and

$$h^T H(a) h = \sum_{i,j=1}^n h_i \frac{\partial^2 f}{\partial x_i \partial x_j} h_j$$

Then using Taylor series, we have

$$f(a + \varepsilon h) = f(a) + \varepsilon h^T \text{grad } f(a) + \frac{\varepsilon^2}{2} h^T H(a) h + O(\varepsilon^3)$$

for all $h \in \mathbb{R}^n$ and sufficiently small ε.

Theorem 2.1. *If f has a local minimum at $x = a$ inside a domain K and continuous second order derivatives, and the third order derivative of f is bounded, then*

$$grad\ f(a) = 0$$

Proof. Since a is a local minima,

$$f(a + \varepsilon h) \geq f(a)$$

for all $h \in \mathbb{R}^n$ and sufficiently small ε.

By the above Taylor expansion, we have

$$\varepsilon h^T + \frac{\varepsilon^2}{2} h^T H(a) h + O(\varepsilon^3) \geq 0$$

Divided by $\varepsilon > 0$ and letting $\varepsilon \to 0^+$, we have

$$h^T \mathrm{grad} f(a) \geq 0.$$

Similarly, divided by $\varepsilon < 0$ and letting $\varepsilon \to 0^-$, we have $\mathrm{grad} f(a) \leq 0$. Therefore we get

$$\mathrm{grad} f(a) = 0. \quad \square$$

Example: Let $f(x_1, x_2) = x_1^2 + x_2^2$ be defined on $\mathbb{R}^n$ and has a local minimum at $(0, 0)$. Then $\mathrm{grad} f(0, 0) = 0$.

Theorem 2.2. *Let f be defined an open domain K of $\mathbb{R}^n$ and have continuous second order derivatives. Then a sufficient condition for $f(x)$ to have a local minimum at a point $a \in K$ is that*

$$\textit{grad } f(a) = 0 \quad \textit{and} \quad h^T H(a) h > 0, \quad \forall h \in \mathbb{R}^n.$$

Hints. Taylor Expansion

$$f(a + \varepsilon h) - f(a) = \frac{\varepsilon^2}{2} h^T H(a) h + O(\varepsilon^3) > 0$$

for a sufficient small ε. $\square$

Theorem 2.3. *(From Linear algebra)*

The quadratic form $h^T H h > 0$ is positive if and only if $\det H$ *and all principal minors of H are positive (This means that all eigenvalues of H are positive).*

In $\mathbb{R}^3$, this would require

$$\frac{\partial^2 f}{\partial x_1^2} > 0, \quad \det \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} > 0$$

and

$$\det \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{pmatrix} > 0$$

The corresponding theorem for a local maximum requires that *gradf* $= 0$ and $h^T Hh < 0$ for all h. A quadrative form $h^T Hh$ is negative if and only $(-1)^n \det H > 0$ and the principal minors of H alternate in sign with $\frac{\partial^2 f}{\partial x_1^2} < 0$ (This means that all eigenvalues of H are negative.)

Theorem 2.4. *Let f be defined an open domain K of $\mathbb{R}^n$ and have continuous second order derivatives. Then a sufficient condition for $f(x)$ to have a local maximum at a point $a \in K$ is that*

$$\textit{grad } f(a) = 0 \quad \textit{and} \quad h^T H(a)h < 0, \quad \forall h \in \mathbb{R}^n.$$

Hints. Take $g(x) = -f(x)$. Then we can apply Theorem 2.2. $\square$

Example 2.5. *Minimize $f(x_1, x_2) = x_1^2 - x_2^2$ where K is $\mathbb{R}^2$.*

Solution. grad$f = (2x_1, -2x_2)^T$ and this is zero for $x_1 = x_2 = 0$ and

$$\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

Now $\det H = -4$ and $\frac{\partial^2 f}{\partial x_1^2} = 2$, so the quadratic form is neither positive nor negative and the origin is neither a maximum nor a minimum. $\square$

Exercise:

Find the local maxima and minima of the functions in $\mathbb{R}^2$:

1. $f(x_1, x_2) = (x_1^2 - 4)^2 + x_2^2$.
2. $f(x_1, x_2) = x_1^3 + x_2^3 + 3x_1^2 - 3x_2^2 - 8$.