Optimization Principles for Univariate Functions

Maximum and Minimum values of univariate functions

Let $f : \mathcal{D} \rightarrow \Re$. Now f has

- An *absolute maximum* (or global maximum) value at point $c \in \mathcal{D}$ if

$$f(x) \leq f(c), \ \forall x \in \mathcal{D}$$

- An *absolute minimum* (or global minimum) value at $c \in \mathcal{D}$ if

$$f(x) \geq f(c), \ \forall x \in \mathcal{D}$$

- A *local maximum value* at c if there is an open interval $\mathcal{I}$ containing c in which $f(c) \geq f(x), \ \forall x \in \mathcal{I}$
- A *local minimum value* at c if there is an open interval $\mathcal{I}$ containing c in which $f(c) \leq f(x), \ \forall x \in \mathcal{I}$
- A *local extreme value* at c, if $f(c)$ is either a local maximum or local minimum value of f in an open interval $\mathcal{I}$ with $c \in \mathcal{I}$

First Derivative Test

First derivative test for local extreme value of f, when f is differentiable at the extremum.

Claim

If $f(c)$ is a local extreme value and if f is differentiable at $x = c$, then $f'(c) = 0$.

[1] By virtue of the *squeeze* or *sandwich theorem*

First Derivative Test

First derivative test for local extreme value of f, when f is differentiable at the extremum.

Claim

If $f(c)$ is a local extreme value and if f is differentiable at $x = c$, then $f'(c) = 0$.

Proof: Suppose $f(c) \geq f(x)$ for all x in an open interval $\mathcal{I}$ containing c and that $f'(c)$ exists. Then the difference quotient $\frac{f(c+h)-f(c)}{h} \leq 0$ for small $h \geq 0$ (so that $c + h \in \mathcal{I}$). This inequality remains true as $h \to 0$ from the right. In the limit, $f'(c) \leq 0$.

[1] By virtue of the *squeeze* or *sandwich theorem*

First Derivative Test

First derivative test for local extreme value of f, when f is differentiable at the extremum.

Claim

If $f(c)$ is a local extreme value and if f is differentiable at $x = c$, then $f'(c) = 0$.

Proof: Suppose $f(c) \geq f(x)$ for all x in an open interval $\mathcal{I}$ containing c and that $f'(c)$ exists. Then the difference quotient $\frac{f(c+h)-f(c)}{h} \leq 0$ for small $h \geq 0$ (so that $c + h \in \mathcal{I}$). This inequality remains true as $h \to 0$ from the right. In the limit, $f'(c) \leq 0$. Also, the difference quotient $\frac{f(c+h)-f(c)}{h} \geq 0$ for small $h \leq 0$ (so that $c + h \in \mathcal{I}$). This inequality remains true as $h \to 0$ from the left. In the limit, $f'(c) \geq 0$. Since $f'(c) \leq 0$ as well as $f'(c) \geq 0$, we must have $f'(c) = 0$[1]. □

[1] By virtue of the *squeeze* or *sandwich theorem*

The Extreme Value Theorem

A most fundamental theorems in calculus concerning continuous functions on closed intervals.

Claim

A continuous function $f(x)$ on a closed and bounded interval $[a, b]$ attains a minimum value $f(c)$ for some $c \in [a, b]$ and a maximum value $f(d)$ for some $d \in [a, b]$. That is, a continuous function on a closed, bounded interval attains a minimum and a maximum value.

The Extreme Value Theorem (contd.)

We must point out that either or both of the values c and d may be attained at the end points of the interval $[a, b]$. Based on theorem (1), the extreme value theorem can extended as:

Claim

A continuous function $f(x)$ on a closed and bounded interval $[a, b]$ attains a minimum value $f(c)$ for some $c \in [a, b]$ and a maximum value $f(d)$ for some $d \in [a, b]$. If $a < c < b$ and $f'(c)$ exists, then $f'(c) = 0$. If $a < d < b$ and $f'(d)$ exists, then $f'(d) = 0$.

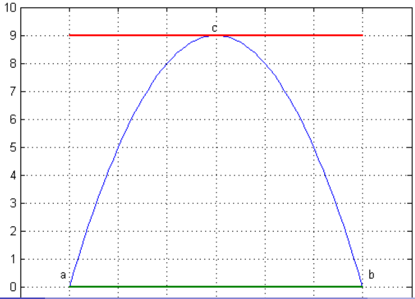
Proof sketch: In 4 parts. In $\Re^n$, one additionally needs compactness of the set in order to get this result.

Rolle's Theorem

Claim

If f is continuous on $[a, b]$ *and differentiable at all* $x \in (a, b)$ *and if* $f(a) = f(b)$*, then* $f'(c) = 0$ *for some* $c \in (a, b)$*.*

This result can be easily proved using the Extreme value theorem. Figure 1 illustrates Rolle's theorem with an example function $f(x) = 9 - x^2$ on the interval $[-3, +3]$.



Mean Value Theorem

A generalization of the Rolle's theorem and proved using the Rolle's theorem:

Claim

If f is continuous on $[a, b]$ and differentiable at all $x \in (a, b)$, then there is some $c \in (a, b)$ such that, $f'(c) = \frac{f(b)-f(a)}{b-a}$.

Mean Value Theorem

A generalization of the Rolle's theorem and proved using the Rolle's theorem:

Claim

If f is continuous on $[a, b]$ *and differentiable at all* $x \in (a, b)$, *then there is some* $c \in (a, b)$ *such that,* $f'(c) = \frac{f(b)-f(a)}{b-a}$.

Proof: Define $g(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x - a)$ on $[a, b]$. We note rightaway that $g(a) = g(b)$ and $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$. Applying Rolle's theorem on $g(x)$, we know that there exists $c \in (a, b)$ such that $g'(c) = 0$. Which implies that $f'(c) = \frac{f(b)-f(a)}{b-a}$. □

Mean Value Theorem (contd.)

Figure 2 illustrates the mean value theorem for $f(x) = 9 - x^2$ on the interval $[-3, 1]$. We observe that the tangent at $x = -1$ is parallel to the secant joining -3 to 1. That is, $f'(-1) = \frac{f(1)-f(-3)}{4}$ One could think of the *mean value theorem* as a slanted version of Rolle's theorem.

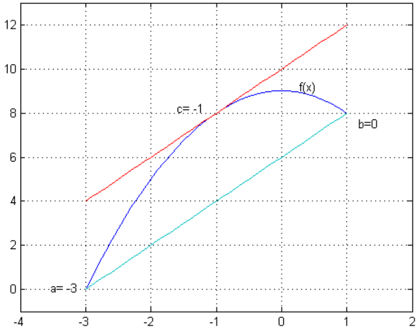


Figure 2

Corollary and Approximations

A natural corollary of Mean Value Theorem is as follows:

Corollary

Let f be continuous on $[a, b]$ *and differentiable on* (a, b) *with* $m \leq f'(x) \leq M, \ \forall x \in (a, b)$. *Then,* $m(x-t) \leq f(x) - f(t) \leq M(x-t)$, *if* $a \leq t \leq x \leq b$.

Corollary and Approximations (contd.)

Let $\mathcal{D}$ be the domain of function f. We define

1. the linear approximation of a differentiable function $f(x)$ as $L_a(x) = f(a) + f'(a)(x - a)$ for some $a \in \mathcal{D}$. We note that $L_a(x)$ and its first derivative at a agree with $f(a)$ and $f'(a)$ respectively.
2. the quadratic approximatin of a twice differentiable function $f(x)$ as the parabola $Q_a(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2$. We note that $Q_a(x)$ and its first and second derivatives at a agree with $f(a)$, $f'(a)$ and $f''(a)$ respectively.
3. the cubic approximation of a thrice differentiable function $f(x)$ is $C_a(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 + \frac{1}{6}f'''(a)(x - a)^3$. $C_a(x)$ and its first, second and third derivatives at a agree with $f(a)$, $f'(a)$, $f''(a)$ and $f'''(a)$ respectively.

Convexity and Concavity of Approximations

The parabola given by $Q_a(x)$ is strictly convex if $f''(a) > 0$ and is strictly concave if $f''(a) < 0$. The coefficient of x^2 in $Q_a(x)$ is $\frac{1}{2}f''(a)$. Figure 3 illustrates the linear, quadratic and cubic approximations to the function $f(x) = \frac{1}{x}$ with $a = 1$.

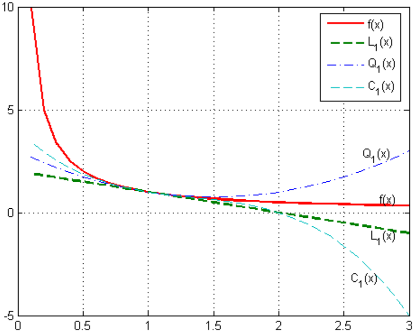


Figure 3:

Taylor's Theorem and n^{th} degree polynomial approximation

The n^{th} degree polynomial approximation of a function is used to prove a generalization of the mean value theorem, called the *Taylor's theorem*.

Claim

The Taylor's theorem states that if f and its first n derivatives $f', f'', \ldots, f^{(n)}$ *are continuous on the closed interval* $[a, b]$, *and differentiable on* (a, b), *then there exists a number* $c \in (a, b)$ *such that*

$$f(b) = f(a) + f'(a)(b-a) + \frac{1}{2!}f''(a)(b-a)^2 + \ldots + \frac{1}{n!}f^{(n)}(a)(b-a)^n + \frac{1}{(n+1)!}f^{(n+1)}(c)(b-a)^{n+1}$$

Proof:

Define

$$p_n(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \ldots + \frac{1}{n!}f^{(n)}(a)(x-a)^n$$

and

$$\phi_n(x) = p_n(x) + \Gamma(x-a)^{n+1}$$

The polynomials $p_n(x)$ as well as $\phi_n(x)$ and their first n derivatives match f and its first n derivatives at $x = a$. We will choose a value of Γ so that

$$f(b) = p_n(b) + \Gamma(b-a)^{n+1}$$

This requires that $\Gamma = \frac{f(b)-p_n(b)}{(b-a)^{n+1}}$.

Taylor's Theorem and n^{th} degree polynomial approximation

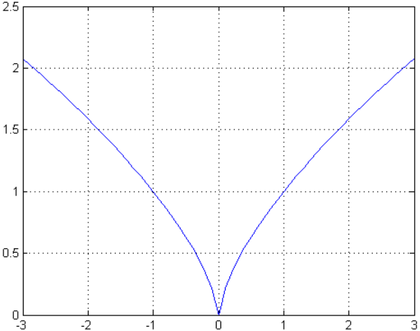
Define the function $g(x) = f(x) - \phi_n(x)$ that measures the difference between function f and the approximating function $\phi_n(x)$ for each $x \in [a, b]$.

- Since $g(a) = g(b) = 0$ and since g and g' are both continuous on $[a, b]$, we can apply the Rolle's theorem to conclude that there exists $c_1 \in [a, b]$ such that $g'(c_1) = 0$.
- Similarly, since $g'(a) = g'(c_1) = 0$, and since g' and g'' are continuous on $[a, c_1]$, we can apply the Rolle's theorem to conclude that there exists $c_2 \in [a, c_1]$ such that $g''(c_2) = 0$.
- In this way, Rolle's theorem can be applied successively to $g'', g''', \ldots, g^{(n+1)}$ to imply the existence of $c_i \in (a, c_{i-1})$ such that $g^{(i)}(c_i) = 0$ for $i = 3, 4, \ldots, n+1$. Note however that $g^{(n+1)}(x) = f^{(n+1)}(x) - 0 - (n+1)!\Gamma$ which gives us another representation 'of Γ as $\frac{f^{(n+1)}(c_{n+1})}{(n+1)!}$.

Thus

Mean Value, Taylor's Theorem and words of caution

Note that if f fails to be differentiable at even one number in the interval, then the conclusion of the mean value theorem may be false. For example, if $f(x) = x^{2/3}$, then $f'(x) = \frac{2}{3\sqrt[3]{x}}$ and the theorem does not hold in the interval $[-3, 3]$, since f is not differentiable at s0 as can be seen in Figure 4.



Sufficient Conditions for Increasing and decreasing functions

A function f is said to be ...

- *increasing* on an interval $\mathcal{I}$ in its domain $\mathcal{D}$ if $f(t) < f(x)$ whenever $t < x$.
- *decreasing* on an interval $\mathcal{I} \in \mathcal{D}$ if $f(t) > f(x)$ whenever $t < x$.

Consequently:

Claim

Let $\mathcal{I}$ be an interval and suppose f is continuous on $\mathcal{I}$ and differentiable on $int(\mathcal{I})$. Then:

1. *if $f'(x) > 0$ for all $x \in int(\mathcal{I})$, then f is increasing on $\mathcal{I}$;*
2. *if $f'(x) < 0$ for all $x \in int(\mathcal{I})$, then f is decreasing on $\mathcal{I}$;*
3. *if $f'(x) = 0$ for all $x \in int(\mathcal{I})$, iff, f is constant on $\mathcal{I}$.*

Proof

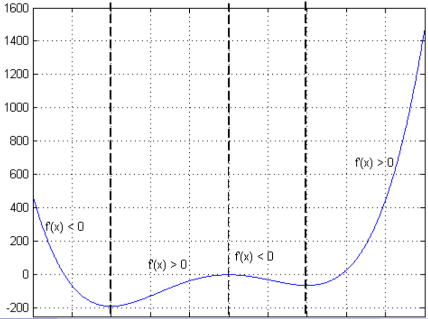
Proof:
Let $t \in \mathcal{I}$ and $x \in \mathcal{I}$ with $t < x$. By virtue of the mean value theorem, $\exists c \in (t, x)$ such that $f'(c) = \frac{f(x)-f(t)}{x-t}$.

- If $f'(x) > 0$ for all $x \in int(\mathcal{I})$, $f'(c) > 0$, which implies that $f(x) - f(t) > 0$ and we can conclude that f is increasing on $\mathcal{I}$.
- If $f'(x) < 0$ for all $x \in int(\mathcal{I})$, $f'(c) < 0$, which implies that $f(x) - f(t) < 0$ and we can conclude that f is decreasing on $\mathcal{I}$.
- If $f'(x) = 0$ for all $x \in int(\mathcal{I})$, $f'(c) = 0$, which implies that $f(x) - f(t) = 0$, and since x and t are arbitrary, we can conclude that f is constant on $\mathcal{I}$.

□

Illustration

Figure 5 illustrates the intervals in $(-\infty, \infty)$ on which the function $f(x) = 3x^4 + 4x^3 - 36x^2$ is decreasing and increasing. First we note that $f(x)$ is differentiable everywhere on $(-\infty, \infty)$ and compute $f'(x) = 12(x^3 + x^2 - 6x) = 12(x-2)(x+3)x$, which is negative in the intervals $(-\infty, -3]$ and $[0, 2]$ and positive in the intervals $[-3, 0]$ and $[2, \infty)$. We observe that f is decreasing in the intervals $(-\infty, -3]$ and $[0, 2]$ and while it is increasing in the intervals $[-3, 0]$ and $[2, \infty)$.



Another sufficient condition for increasing/decreasing function

A related sufficient condition for a function f to be increasing/decreasing on an interval $\mathcal{I}$:

Claim

Let $\mathcal{I}$ be an interval and suppose f is continuous on $\mathcal{I}$ and differentiable on int($\mathcal{I}$). Then:

1. *if $f'(x) \geq 0$ for all $x \in int(\mathcal{I})$, and if $f'(x) = 0$ at only finitely many $x \in \mathcal{I}$, then f is increasing on $\mathcal{I}$;*
2. *if $f'(x) \leq 0$ for all $x \in int(\mathcal{I})$, and if $f'(x) = 0$ at only finitely many $x \in \mathcal{I}$, then f is decreasing on $\mathcal{I}$.*

For example, the derivative of the function $f(x) = 6x^5 - 15x^4 + 10x^3$ vanishes at 0, and 1 and $f'(x) > 0$ elsewhere. So $f(x)$ is increasing on $(-\infty, \infty)$.

Necessary conditions for increasing/decreasing function

The conditions for increasing and decreasing properties of $f(x)$ in theorem 7 are not necesssary. Figure 6 shows that for the function $f(x) = x^5$, though $f(x)$ is increasing in $(-\infty, \infty)$, $f'(0) = 0$.

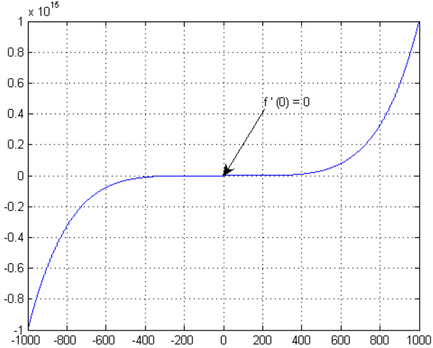


Figure 6:

Necessary conditions for increasing/decreasing function (contd.)

We have a slightly different necessary condition..

Claim

Let $\mathcal{I}$ be an interval, and suppose f is continuous on $\mathcal{I}$ and differentiable in int$(\mathcal{I})$. Then:

1. *if f is increasing on $\mathcal{I}$, then $f'(x) \geq 0$ for all $x \in$ int$(\mathcal{I})$;*
2. *if f is decreasing on $\mathcal{I}$, then $f'(x) \leq 0$ for all $x \in$ int$(\mathcal{I})$.*

Proof:

Necessary conditions for increasing/decreasing function (contd.)

We have a slightly different necessary condition..

Claim

Let $\mathcal{I}$ be an interval, and suppose f is continuous on $\mathcal{I}$ and differentiable in $int(\mathcal{I})$. Then:

1. *if f is increasing on $\mathcal{I}$, then $f'(x) \geq 0$ for all $x \in int(\mathcal{I})$;*
2. *if f is decreasing on $\mathcal{I}$, then $f'(x) \leq 0$ for all $x \in int(\mathcal{I})$.*

Proof: Suppose f is increasing on $\mathcal{I}$, and let $x \in int(\mathcal{I})$. Then $\frac{f(x+h)-f(x)}{h} > 0$ for all h such that $x+h \in int(\mathcal{I})$. This implies that $f'(x) = \lim\limits_{h\to 0} \frac{f(x+h)-f(x)}{h} \geq 0$. For the case when f is decreasing on $\mathcal{I}$, it can be similarly proved that $f'(x) = \lim\limits_{h\to 0} \frac{f(x+h)-f(x)}{h} \leq 0$. $\square$

Critical Point

This concept will help us derive the general condition for local extrema.

Definition

[Critical Point]: *A point c in the domain $\mathcal{D}$ of f is called a critical point of f if either $f'(c) = 0$ or $f'(c)$ does not exist.*

The following general condition for local extrema extends the result in theorem 1 to general non-differentiable functions.

Claim

If $f(c)$ is a local extreme value, then c is a critical number of f.

The converse of theorem 10 does not hold (see Figure 6); 0 is a critical number ($f'(0) = 0$), although $f(0)$ is not a local extreme value.

Critical Point and Local Extreme Value

Given a critical point c, the following test helps determine if $f(c)$ is a local extreme value:

Procedure

[Local Extreme Value]: *Let c be an isolated critical point of f*

1. $f(c)$ *is a local minimum if* $f(x)$ *is decreasing in an interval* $[c-\epsilon_1, c]$ *and increasing in an interval* $[c, c+\epsilon_2]$ *with* $\epsilon_1, \epsilon_2 > 0$.
2. $f(c)$ *is a local maximum if* $f(x)$ *is increasing in an interval* $[c-\epsilon_1, c]$ *and decreasing in an interval* $[c, c+\epsilon_2]$ *with* $\epsilon_1, \epsilon_2 > 0$.

Given a critical point c, *first derivative test* (sufficient condition) helps determine if $f(c)$ is a local extreme value:

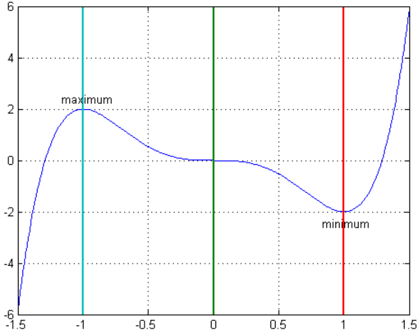
Procedure

[First derivative test]: *Let c be an isolated critical point of f*

1. *$f(c)$ is a local minimum if the sign of $f'(x)$ changes from negative in $[c - \epsilon_1, c]$ to positive in $[c, c + \epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$.*
2. *$f(c)$ is a local maximum if $f(x)$ the sign of $f'(x)$ changes from positive in $[c - \epsilon_1, c]$ to negative in $[c, c + \epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$.*
3. *If $f'(x)$ is positive in an interval $[c - \epsilon_1, c]$ and also positive in an interval $[c, c - \epsilon_2]$, or $f'(x)$ is negative in an interval $[c - \epsilon_1, c]$ and also negative in an interval $[c, c - \epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$, then $f(c)$ is not a local extremum.*

First Derivative Test: Critical Point and Local Extreme Value

As an example, the function $f(x) = 3x^5 - 5x^3$ has the derivative $f'(x) = 15x^2(x+1)(x-1)$. The critical points are 0, 1 and -1. Of the three, the sign of $f'(x)$ changes at 1 and -1, which are local minimum and maximum respectively. The sign does not change at 0, which is therefore not a local supremum.



First Derivative Test: Critical Point and Local Extreme Value

As another example, consider the function

$$f(x) = \begin{cases} -x & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Then,

$$f'(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x > 0 \end{cases}$$

Note that $f(x)$ is discontinuous at $x = 0$, and therefore $f'(x)$ is not defined at $x = 0$. All numbers $x \geq 0$ are critical numbers. $f(0) = 0$ is a local minimum, whereas $f(x) = 1$ is a local minimum as well as a local maximum $\forall x > 0$.

Strict Convexity and Extremum

- A differentiable function f is said to be *strictly convex* (or *strictly concave up*) on an open interval $\mathcal{I}$, *iff*, $f'(x)$ is increasing on $\mathcal{I}$.
- Recall from theorem 7, the graphical interpretation of the first derivative $f'(x)$; $f'(x) > 0$ implies that $f(x)$ is increasing at x.
- Similarly, $f'(x)$ is increasing when $f''(x) > 0$. This gives us a sufficient condition for the strict convexity of a function:

Claim

If at all points in an open interval $\mathcal{I}$, $f(x)$ is doubly differentiable and if $f''(x) > 0, \ \forall x \in \mathcal{I}$, then the slope of the function is always increasing with x and the graph is strictly convex. This is illustrated in Figure 8.

- On the other hand, if the function is strictly convex and doubly differentiable in $\mathcal{I}$, then $f''(x) \geq 0, \ \forall x \in \mathcal{I}$.

Strict Convexity and Extremum (Illustrated)

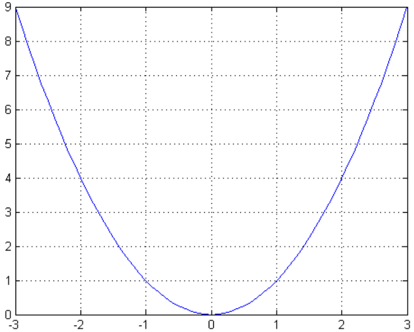


Figure 8:

Strict Convexity and Extremum: Slopeless interpretation (SI)

Claim

A function f is strictly convex on an open interval $\mathcal{I}$, iff

$$f(ax_1 + (1-a)x_2) < af(x_1) + (1-a)f(x_2) \quad (1)$$

whenver $x_1, x_2 \in \mathcal{I}$, $x_1 \neq x_2$ and $0 < a < 1$.

SI: Necessity when f is differentiable

First we will prove the **necessity**.

[2] For the case $x_2 < x_1$, the proof is very similar.

SI: Necessity when f is differentiable

First we will prove the **necessity**. Suppose f' is increasing on $\mathcal{I}$. Let $0 < a < 1$, $x_1, x_2 \in \mathcal{I}$ and $x_1 \neq x_2$. Without loss of generality assume that $x_1 < x_2$[2]. Then, $x_1 < ax_1 + (1-a)x_2 < x_2$ and therefore $ax_1 + (1-a)x_2 \in \mathcal{I}$. By the mean value theorem, there exist s and t with $x_1 < s < ax_1 + (1-a)x_2 < t < x_2$, such that $f(ax_1 + (1-a)x_2) - f(x_1) = f'(s)(x_2 - x_1)(1-a)$ and $f(x_2) - f(ax_1 + (1-a)x_2) = f'(t)(x_2 - x_1)a$. Therefore,

$$\begin{aligned} (1-a)f(x_1) - f(ax_1 + (1-a)x_2) + af(x_2) &= \\ a\left[f(x_2) - f(ax_1 + (1-a)x_2)\right] - (1-a)\left[f(ax_1 + (1-a)x_2) - f(x_1)\right] &= \\ a(1-a)(x_2 - x_1)\left[f'(t) - f'(s)\right] \end{aligned}$$

Since $f(x)$ is strictly convex on $\mathcal{I}$, $f'(x)$ is increasing $\mathcal{I}$ and therefore, $f'(t) - f'(s) > 0$. Moreover, $x_2 - x_1 > 0$ and $0 < a < 1$. This implies that $(1-a)f(x_1) - f(ax_1 + (1-a)x_2) + af(x_2) > 0$, or equivalently, $f(ax_1 + (1-a)x_2) < af(x_1) + (1-a)f(x_2)$, which is what we wanted to prove in 1.

[2] For the case $x_2 < x_1$, the proof is very similar.

SI: Sufficiency when f is differentiable

Suppose the inequality in 1 holds. Therefore, $\lim\limits_{a \to 0} \frac{f(x_2 + a(x_1 - x_2)) - f(x_2)}{a} \leq f(x_1) - f(x_2)$. That is,

$$f'(x_2)(x_1 - x_2) \leq f(x_1) - f(x_2) \tag{2}$$

Similarly, we can show that

$$f'(x_1)(x_2 - x_1) \leq f(x_2) - f(x_1) \tag{3}$$

Adding the left and right hand sides of inequalities in (2) and (3), and multiplying the resultant inequality by -1 gives us

$$\left(f'(x_2) - f'(x_1)\right)(x_2 - x_1) \geq 0 \tag{4}$$

SI: Sufficiency when f is differentiable (contd)

Using the mean value theorem, $\exists z = x_1 + t(x_2 - x_1)$ for $t \in (0,1)$ such that

$$f(x_2) - f(x_1) = f'(z)(x_2 - x_1) \tag{5}$$

Since 4 holds for any $x_1, x_2 \in \mathcal{I}$, it also hold for $x_2 = z$. Therefore,

$$(f'(z) - f'(x_1))(x_2 - x_1) = \frac{1}{t}(f'(z) - f'(x_1))(z - x_1) \geq 0$$

Additionally using 5, we get

$$f(x_2) - f(x_1) = (f'(z) - f'(x_1))(x_2 - x_1) + f'(x_1)(x_2 - x_1) \geq f'(x_1)(x_2 - x_1) \tag{6}$$

SI: Sufficiency when f is differentiable (contd)

Suppose equality holds in 4 for some $x_1 \neq x_2$. Then equality holds in 6 for the same x_1 and x_2. That is, $f(x_2) - f(x_1) = f'(x_1)(x_2 - x_1)$. Applying 6 we can conclude that

$$f(x_1) + af'(x_1)(x_2 - x_1) \leq f(x_1 + a(x_2 - x_1)) \quad (7)$$

From 1 and **??**, we can derive that

$$f(x_1 + a(x_2 - x_1)) < (1 - a)f(x_1) + af(x_2) = f(x_1) + af'(x_1)(x_2 - x_1) \quad (8)$$

However, equations 7 and 8 contradict each other. Therefore, equality in 4 cannot hold for any $x_1 \neq x_2$, implying that

$$\left(f'(x_2) - f'(x_1)\right)(x_2 - x_1) > 0$$

that is, $f'(x)$ is increasing and therefore f is convex on $\mathcal{I}$.

Strict Concavity

- A differentiable function f is said to be *strictly concave* on an open interval $\mathcal{I}$, *iff*, $f'(x)$ is decreasing on $\mathcal{I}$.
- Recall from theorem 7, the graphical interpretation of the first derivative $f'(x)$; $f'(x) < 0$ implies that $f(x)$ is decreasing at x.
- Similarly, $f'(x)$ is monotonically decreasing when $f''(x) > 0$. This gives us a sufficient condition for the concavity of a function:

Claim

If at all points in an open interval $\mathcal{I}$, $f(x)$ is doubly differentiable and if $f''(x) < 0, \ \forall x \in \mathcal{I}$, then the slope of the function is always decreasing with x and the graph is strictly concave.

Strict Concavity

On the other hand, if the function is strictly concave and doubly differentiable in $\mathcal{I}$, then $f''(x) \leq 0, \ \forall x \in \mathcal{I}$. This is illustrated in Figure 9.

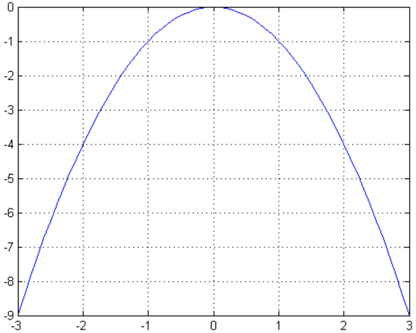


Figure 9:

Strict Concavity (slopeless interpretation)

There is also a slopeless interpretation of concavity as stated in the following theorem:

Claim

A differentiable function f is strictly concave on an open interval $\mathcal{I}$, iff

$$f(ax_1 + (1-a)x_2) > af(x_1) + (1-a)f(x_2) \tag{9}$$

whenver $x_1, x_2 \in \mathcal{I}$, $x_1 \neq x_2$ and $0 < a < 1$.

The proof is similar to that for theorem 12.

Convex & Concave Regions and Inflection Point

Figure 10 illustrates a function $f(x) = x^3 - x + 2$, whose slope decreases as x increases to 0 ($f'(x) < 0$) and then the slope increases beyond $x = 0$ ($f'(x) > 0$). The point 0, where the $f'(x)$ changes sign is called the *inflection point*; the graph is strictly concave for $x < 0$ and strictly convex for $x > 0$.

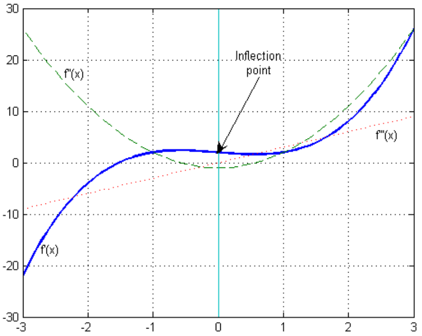


Figure 10

Convex & Concave Regions and Inflection Point

Along similar lines, we can diagnose the function
$f(x) = \frac{1}{20}x^5 - \frac{7}{12}x^4 + \frac{7}{6}x^3 - \frac{15}{2}x^2$
It is strictly concave on $(-\infty, -1]$ and $[3, 5]$ and strictly convex on $[-1, 3]$ and $[5, \infty]$.
The inflection points for this function are at $x = -1$, $x = 3$ and $x = 5$.

First Derivative Test: Restated using Strict Convexity

The *first derivative test* for local extrema can be restated in terms of strict convexity and concavity of functions.

Procedure

[First derivative test in terms of strict convexity]: *Let c be a critical number of f and $f'(c) = 0$. Then,*

1. *$f(c)$ is a local minimum if the graph of $f(x)$ is strictly convex on an open interval containing c.*
2. *$f(c)$ is a local maximum if the graph of $f(x)$ is strictly concave on an open interval containing c.*

Strict Convexity: Restated using Second Derivative

If the second derivative $f''(c)$ exists, then the strict convexity conditions for the critical number can be stated in terms of the sign of of $f''(c)$, making use of theorems 11 and 13. This is called the *second derivative test*.

Procedure

[Second derivative test]: *Let c be a critical number of f where*

$$f'(c) = 0 \text{ and } f''(c) \text{ exists.}$$

1. *If $f''(c) > 0$ then $f(c)$ is a local minimum.*
2. *If $f''(c) < 0$ then $f(c)$ is a local maximum.*
3. *If $f''(c) = 0$ then $f(c)$ could be a local maximum, a local minimum, neither or both. That is, the test fails.*

Convexity, Minima and Maxima: Illustrations

- If $f(x) = x^4$, then $f'(0) = 0$ and $f''(0) = 0$ and we can see that $f(0)$ is a local minimum.
- If $f(x) = -x^4$, then $f'(0) = 0$ and $f''(0) = 0$ and we can see that $f(0)$ is a local maximum.
- If $f(x) = x^3$, then $f'(0) = 0$ and $f''(0) = 0$ and we can see that $f(0)$ is neither a local minimum nor a local maximum. $(0, 0)$ is an inflection point in this case.

Convexity, Minima and Maxima: Illustrations (contd.)

- If $f(x) = x + 2\sin x$, then $f'(x) = 1 + 2\cos x$. $f'(x) = 0$ for $x = \frac{2\pi}{3}, \frac{4\pi}{3}$, which are the critical numbers. $f''\left(\frac{2\pi}{3}\right) = -2\sin\frac{2\pi}{3} = -\sqrt{3} < 0 \Rightarrow f\left(\frac{2\pi}{3}\right) = \frac{2\pi}{3} + \sqrt{3}$ is a local maximum value. On the other hand, $f''\left(\frac{4\pi}{3}\right) = \sqrt{3} > 0 \Rightarrow f\left(\frac{4\pi}{3}\right) = \frac{4\pi}{3} - \sqrt{3}$ is a local minimum value.
- If $f(x) = x + \frac{1}{x}$, then $f'(x) = 1 - \frac{1}{x^2}$. The critical numbers are $x = \pm 1$. Note that $x = 0$ is not a critical number, even though $f'(0)$ does not exist, because 0 is not in the domain of f. $f''(x) = \frac{2}{x^3}$. $f''(-1) = -2 < 0$ and therefore $f(-1) = -2$ is a local maximum. $f''(1) = 2 > 0$ and therefore $f(1) = 2$ is a local minimum.

Global Extrema on Closed Intervals

Recall the extreme value theorem (theorem 2). An outcome of the extreme value theorem is that

- if either of c or d lies in (a, b), then it is a critical number of f;
- else each of c and d must lie on one of the boundaries of $[a, b]$.

This gives us a procedure for finding the maximum and minimum of a continuous function f on a closed bounded interval $\mathcal{I}$:

Procedure

[Finding extreme values on closed, bounded intervals]:

1. *Find the critical points in int*$(\mathcal{I})$.
2. *Compute the values of f at the critical points and at the endpoints of the interval.*
3. *Select the least and greatest of the computed values.*

Global Extrema on Closed Intervals (contd)

To compute the maximum and minimum values of
$f(x) = 4x^3 - 8x^2 + 5x$ on the interval $[0, 1]$, we first compute
$f'(x) = 12x^2 - 16x + 5$ which is 0 at $x = \frac{1}{2}, \frac{5}{6}$.
Values at the critical points are $f(\frac{1}{2}) = 1$, $f(\frac{5}{6}) = \frac{25}{27}$.
The values at the end points are $f(0) = 0$ and $f(1) = 1$.
Therefore, the minimum value is $f(0) = 0$ and the maximum value is $f(1) = f(\frac{1}{2}) = 1$.
In this context, it is relevant to discuss the one-sided derivatives of a function at the endpoints of the closed interval on which it is defined.

Global Extrema on Closed Intervals (contd)

Definition

[One-sided derivatives at endpoints]: *Let f be defined on a closed bounded interval $[a, b]$. The (right-sided) derivative of f at $x = a$ is defined as*

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

Similarly, the (left-sided) derivative of f at $x = b$ is defined as

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

Essentially, each of the one-sided derivatives defines one-sided slopes at the endpoints.

Global Extrema on Closed Intervals (contd)

Based on these definitions, the following result can be derived.

Claim

If f is continuous on $[a, b]$ and $f'(a)$ exists as a real number or as $\pm\infty$, then we have the following necessary conditions for extremum at a.

- *If $f(a)$ is the maximum value of f on $[a, b]$, then $f'(a) \leq 0$ or $f'(a) = -\infty$.*
- *If $f(a)$ is the minimum value of f on $[a, b]$, then $f'(a) \geq 0$ or $f'(a) = \infty$.*

If f is continuous on $[a, b]$ and $f'(b)$ exists as a real number or as $\pm\infty$, then we have the following necessary conditions for extremum at b.

- *If $f(b)$ is the maximum value of f on $[a, b]$, then $f'(b) \geq 0$ or $f'(b) = \infty$.*
- *If $f(b)$ is the minimum value of f on $[a, b]$, then $f'(b) \leq 0$ or $f'(b) = -\infty$.*

Global Extrema on Closed Intervals (contd)

The following theorem gives a useful procedure for finding extrema on closed intervals.

Claim

If f is continuous on $[a, b]$ and $f'(x)$ exists for all $x \in (a, b)$. Then,

- *If $f'(x) \leq 0, \ \forall x \in (a, b)$, then the minimum value of f on $[a, b]$ is either $f(a)$ or $f(b)$. If, in addition, f has a critical number $c \in (a, b)$, then $f(c)$ is the maximum value of f on $[a, b]$.*
- *If $f'(x) \geq 0, \ \forall x \in (a, b)$, then the maximum value of f on $[a, b]$ is either $f(a)$ or $f(b)$. If, in addition, f has a critical number $c \in (a, b)$, then $f(c)$ is the minimum value of f on $[a, b]$.*

Global Extrema on Open Intervals

The next theorem is very useful for finding global extrema values on open intervals.

Claim

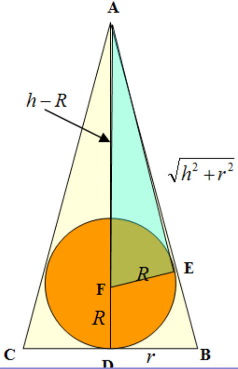
Let $\mathcal{I}$ be an open interval and let $f'(x)$ exist $\forall x \in \mathcal{I}$.

- *If $f'(x) \geq 0, \ \forall x \in \mathcal{I}$, and if there is a number $c \in \mathcal{I}$ where $f'(c) = 0$, then $f(c)$ is the global minimum value of f on $\mathcal{I}$.*
- *If $f'(x) \leq 0, \ \forall x \in \mathcal{I}$, and if there is a number $c \in \mathcal{I}$ where $f'(c) = 0$, then $f(c)$ is the global maximum value of f on $\mathcal{I}$.*

For example, let $f(x) = \frac{2}{3}x - \sec x$ and $\mathcal{I} = (\frac{-\pi}{2}, \frac{\pi}{2})$.
$f'(x) = \frac{2}{3} - \sec x \tan x = \frac{2}{3} - \frac{\sin x}{\cos^2 x} = 0 \Rightarrow x = \frac{\pi}{6}$. Further,
$f''(x) = -\sec x(\tan^2 x + \sec^2 x) < 0$ on $(\frac{-\pi}{2}, \frac{\pi}{2})$. Therefore, f attains the maximum value $f(\frac{\pi}{6}) = \frac{\pi}{9} - \frac{2}{\sqrt{3}}$ on $\mathcal{I}$.

Global Extrema on Open Intervals (contd)

As another example, let us find the dimensions of the cone with minimum volume that can contain a sphere with radius R. Let h be the height of the cone and r the radius of its base. The objective to be minimized is the volume $f(r,h) = \frac{1}{3}\pi r^2 h$. The constraint betwen r and h is shown in Figure 11. The traingle AEF is similar to traingle ADB and therefore, $\frac{h-R}{R} = \frac{\sqrt{h^2+r^2}}{r}$.



Global Extrema on Open Intervals (contd)

Our first step is to reduce the volume formula to involve only one of r^2[3] or h.

The algebra involved will be the simplest if we solved for h.

The constraint gives us $r^2 = \frac{R^2 h}{h-2R}$. Substituting this expression for r^2 into the volume formula, we get $g(h) = \frac{\pi R^2}{3} \frac{h^2}{(h-2R)}$ with the domain given by $\mathcal{D} = \left\{ h | 2R < h < \infty \right\}$.

Note that $\mathcal{D}$ is an open interval.

$g' = \frac{\pi R^2}{3} \frac{2h(h-2R)-h^2}{(h-2R)^2} = \frac{\pi R^2}{3} \frac{h(h-4R)}{(h-2R)^2}$ which is 0 in its domain $\mathcal{D}$ if and only if $h = 4R$.

$g'' = \frac{\pi R^2}{3} \frac{2(h-2R)^3 - 2h(h-4R)(h-2R)^2}{(h-2R)^4} = \frac{\pi R^2}{3} \frac{2(h^2-4Rh+4R^2-h^2+4Rh)}{(h-2R)^3} = \frac{\pi R^2}{3} \frac{8R^2}{(h-2R)^3}$, which is greater than 0 in $\mathcal{D}$.

Therefore, g (and consequently f) has a unique minimum at $h = 4R$ and correspondingly, $r^2 = \frac{R^2 h}{h-2R} = 2R^2$.

---

[3]Since r appears in the volume formula only in terms of r^2.

References

- Online Gradient Descent: Efficient Algorithm for Regret Minimization - Zinkevich 2005
- Yu-Hong Dai, Roger Fletcher. New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. http://link.springer.com/content/pdf/10.1007%2Fs10107-005-0595-2.pdf
- Kalai-Vempala 2005
- Regret bound for Adagrad: Duchi, Hazan, Singer 2010