Notes 1

CONVEX FUNCTIONS

A function f defined in an interval I is called a *convex function* if it satisfies

$$f((1-\lambda)x+\lambda y) \le (1-\lambda)f(x)+\lambda f(y),\ \forall x,y \in I, \lambda \in [0,1].$$

Observe that $z=(1-\lambda)x+\lambda y$ is a point on the line segment connecting x and y. As λ increases from 0 to 1, z runs from x to y. The line segment in $\mathbb{R}^2$ connecting $(x, f(x))$ and $(y, f(y))$ is given by the graph of the linear function

$$\begin{aligned} l(z) &= \Big(\frac{f(y)-f(x)}{y-x}\Big)(z-x)+f(x) \\ &= \Big(\frac{f(x)-f(y)}{x-y}\Big)(z-y)+f(y). \end{aligned}$$

It is readily checked that f is convex if and only if

$$f(z) \le l(z),$$

for any z lying between x and y. (Here l depends on x and y). This condition has a clear geometric meaning. Namely, the line segment connecting $(x, f(x))$ and $(y, f(y))$ always lies above the graph of f over the interval with endpoints x and y.

Proposition 1.1. *Let f be defined in an open interval I. The following conditions are equivalent:*

(a) f is convex on I

(b) for $x, y, z \in I$, with $x < z < y$,

$$\frac{f(z)-f(x)}{z-x} \le \frac{f(y)-f(x)}{y-x}, \tag{1.1}$$

(c) for $x, y, z \in I$, with $x < z < y$,

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(y)-f(z)}{y-z}. \tag{1.2}$$

Proof. (i) $\Rightarrow$ (ii): Assume f is convex. Let $x < y$ be two points in I. Each z in $[x, y]$ can be expressed in the form $z = (1-\lambda)x+\lambda y$ for a unique $\lambda \in [0,1]$. By the definition of convexity we have

$$\begin{aligned}\frac{f(z)-f(x)}{z-x} &= \frac{f((1-\lambda)x+\lambda y)-f(x)}{(1-\lambda)x+\lambda y-x}\\ &\le \frac{(1-\lambda)f(x)+\lambda f(y)-f(x)}{\lambda(y-x)}\\ &= \frac{f(y)-f(x)}{y-x},\end{aligned}$$

and (1.1) follows. Similarly (i) $\Rightarrow$ (iii).

(ii) $\Rightarrow$ (i): Assume (ii) holds. Let $x, y \in I$ with $x < y$, and $\lambda \in [0,1]$. Let $z = \lambda x + (1-\lambda)y$. Then (1.1) implies

$$(y-x)(f(z)-f(x)) \le (z-x)(f(y)-f(x)),$$

i.e.

$$(y-x)f(z) \le (y-z)f(x) + (z-x)f(y),$$

or

$$f(z) \le \frac{y-z}{y-x}f(x) + \frac{z-x}{y-x}f(y).$$

Since $z = \lambda x + (1-\lambda)y$, we have

$$\frac{y-z}{y-x} = \lambda \quad \text{and} \quad \frac{z-x}{y-x} = 1-\lambda.$$

So the above implies

$$f(z) \le \lambda f(x) + (1-\lambda)f(y),$$

and hence f is convex. Similarly (iii) $\Rightarrow$ (i). □

The geometric meaning of the first inequality is that if we let l_x be the line segment connecting $(x_0, f(x_0))$ and $(x, f(x))$ for $x > x_0$. Then the slope of l_x increases as x increases. For (2), considering now $x < x_0$, then the slope of l_x increases as x increases to x_0. Using these properties, we immediately obtain

Proposition 1.2. *Let f be convex on I. Then for $x < z < y$ in I,*

$$\frac{f(z)-f(x)}{z-x} \le \frac{f(y)-f(z)}{y-z}.$$

Proof. We have

$$\begin{aligned}\frac{f(z)-f(x)}{z-x} &\le \frac{f(y)-f(x)}{y-x}\\ &\le \frac{f(y)-f(z)}{y-z}\end{aligned}$$

after using (1.1) and then (1.2). □

Exercise: Show that the converse of the above Proposition is also true.

Theorem 1.3. *Every convex function f on the open interval I has right and left derivatives, and they satisfy*

$$f'_-(x) \le f'_+(x), \ \forall x \in I, \tag{1.3}$$

and

$$f'_+(x) \le f'_-(y), \ \forall x < y \text{ in } I. \tag{1.4}$$

In particular, f is continuous in I.

Proof. From Proposition 1.1 and Proposition 1.2 the function

$$\varphi(t) = \frac{f(t) - f(x)}{t - x}, \ t > x,$$

is increasing and is bounded below by $(f(x) - f(x_0))/(x - x_0)$, where x_0 is any fixed point in I satisfying $x_0 < x$. It follows that $\lim_{t \to x^+} \varphi(t)$ exists. In other words, $f'_+(x)$ exists. Notice that we still have

$$f'_+(x) \ge \frac{f(x) - f(x_0)}{x - x_0},$$

after passing to limit. As the quotient in the right hand side is increasing as x_0 increases to x, by (1.2), we conclude that

$$\lim_{x_0 \to x^-} \frac{f(x) - f(x_0)}{x - x_0} = f'_-(x)$$

exists and (1.3)

$$f'_+(x) \ge f'_-(x)$$

holds. After proving that the right and left derivatives of f exist everywhere in I, we let $z \to x^+$ in (1.1) to get

$$f'_+(x) \le \frac{f(y) - f(x)}{y - x};$$

and let $z \to y^-$ in (1.2) to get

$$\frac{f(y) - f(x)}{y - x} \le f'_-(y),$$

whence (1.4) follows.
□

Theorem 1.4. *Every convex function on I is differentiable except possibly at a countable set.*

Proof. Noting that every interval I can be written as the union of countably many closed and bounded intervals, it suffices to show there are at most countably many non-differentiable points in any closed and bounded interval $[a,b]$ strictly contained inside I. Fix a small $\delta > 0$ so that $[a-\delta, b+\delta] \subset I$. Since f is continuous in $[a-\delta, b+\delta]$, it is bounded in $[a-\delta, b+\delta]$. Let $M \geq |f(x)|, \forall x \in [a-\delta, b+\delta]$. By convexity

$$f'_+(b) \leq \frac{f(b+\delta) - f(b)}{(b+\delta) - b} \leq \frac{2M}{\delta},$$

and

$$f'_-(a) \geq \frac{f(a) - f(a-\delta)}{a - (a-\delta)} \geq \frac{-2M}{\delta},$$

As a result, for $x \in [a,b]$,

$$f'_-(a) \leq f'_\pm(x) \leq f'_+(b),$$

and the estimate

$$\frac{-2M}{\delta} \leq f'_\pm(x) \leq \frac{2M}{\delta}.$$

holds. Non-differentiable points in $[a,b]$ belong to the set

$$D = \{x : f'_+(x) - f'_-(x) > 0\} = \bigcup_{k=1}^{\infty} D_k,$$

where $D_k = \{x : f'_+(x) - f'_-(x) \geq \frac{1}{k}\}$. We claim that each D_k is a finite set. To see this let us pick n many points from $D_k : x_1 < x_2 < ... < x_n$. Then

$$\begin{aligned}
f'_+(x_n) - f'_-(x_1) &\geq f'_+(x_n) - f'_-(x_n) + f'_-(x_n) - f'_-(x_1) \\
&\geq \frac{1}{k} + f'_+(x_{n-1}) - f'_-(x_1) \\
&\geq \frac{2}{k} + f'_+(x_{n-2}) - f'_-(x_1) \\
&... \\
&\geq \frac{n-1}{k} + f'_+(x_1) - f'_-(x_1) \\
&\geq \frac{n}{k}.
\end{aligned}$$

It follows that

$$n \leq k\big(f'_+(x_n) - f'_-(x_1)\big) \leq \frac{4Mk}{\delta}.$$

□

When f is differentiable, Theorem 1.3 asserts that $f^{'}$ is increasing. The converse is also true.

Theorem 1.5. *Let f be differentiable in I. It is convex if and only if $f^{'}$ is increasing.*

Proof. Let $z = (1-\lambda)x + \lambda y \in [x, y]$. Applying the mean-value theorem to f there exist $c_1 \in (x, z)$ and $c_2 \in (z, y)$ such that

$$f(z) = f(x) + f^{'}(c_1)(z - x),$$

and

$$f(y) = f(z) + f^{'}(c_2)(y - z).$$

Using $f^{'}(c_1) \le f^{'}(c_2)$ we get

$$\frac{f(z) - f(x)}{z - x} \le \frac{f(y) - f(z)}{y - z},$$

which, after some computation, simplifies to yield

$$f(z) \le (1-\lambda)f(x) + \lambda f(y).$$

□

Theorem 1.6. *Let f be twice differentiable in I. It is convex if and only if $f^{''} \ge 0$.*

Proof. When f is convex, $f^{'}$ is increasing and so $f^{''} \ge 0$. On the other hand, $f^{''} \ge 0$ implies that $f^{'}$ is increasing and hence convex. □

A function is *strictly convex* on I if it is convex and

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y), \quad \forall x < y, \ \lambda \in (0, 1).$$

From the proofs of the above two theorems we readily deduce the following proposition.

Proposition 1.7. *The function f is strictly convex on I provided one of the followings hold:*

1. *f is differentiable and f' is strictly increasing; or*

2. *f is twice differentiable and $f^{''} > 0$.*

By this proposition, one can verify easily that the following functions are strictly convex.

- $e^{\alpha x}$ where $\alpha \neq 0$ on $(-\infty, \infty)$,
- x^p where $p > 1$ or $p < 0$ on $(0, \infty)$.
- $-\log x$ on $(0, \infty)$.

Two concluding remarks are in order.

First, in some books convexity is defined by a weaker condition, namely, a function f on I is convex if it satisfies

$$f(\frac{x+y}{2}) \leq \frac{1}{2}\Big(f(x) + f(y)\Big), \quad \forall x, y \in I. \tag{1.5}$$

Indeed, this implies

$$f((1-\lambda)x + \lambda y) \leq \ (1-\lambda)f(x) + \lambda f(y), \quad \forall x, y \in I,$$

provided f is continuous on I. However, this conclusion does not hold without continuity.

Second, for any convex function f on I, *Jensen's inequality* holds: Letting $x_1, x_2, \cdots, x_n \in I$ and $\lambda_1, \lambda_2, \cdots, \lambda_n \in (0,1)$ satisfying $\sum_{j=1}^n \lambda_j = 1$,

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n).$$

When f is strictly convex, equality sign in this inequality holds if and only if $x_1 = x_2 = \cdots = x_n$. Many well-known inequalities including the AM-GM inequality and Hölder inequality are special cases of Jensen's inequality. Some of them can be found in Exercise 4.