Notes 1. CONVEX FUNCTIONS

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

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

Observe that $z = (1 - \lambda)x + \lambda y$ is a point on the line segment connecting $x$ and $y$. As $\lambda$ 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

$$l(z) = \left( \frac{f(y) - f(x)}{y - x} \right) (z - x) + f(x)$$

$$= \left( \frac{f(x) - f(y)}{x - y} \right) (z - y) + f(y).$$

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

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

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 $y$.

A function is called **concave** if its negative is convex. Apparently every result for convex functions has a corresponding one for concave functions. In some situations the use of concavity is more appropriate than convexity.

**Proposition 1.1.** Let $f$ be defined on the interval $I$. For $x, y, z \in I, x < z < y$, $f$ is convex if and only if either one of the following inequalities holds

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x}, \quad (1.1)$$

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(y) - f(z)}{y - z}. \quad (1.2)$$

**Proof.** Let $x < y$ be in $I$. Now $f$ is convex if and only if for $z \in [x, y]$, $f(z) \leq l(z)$, that is,

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

Move $f(x)$ to the left hand side and then divide both sides by $z - x$ we get (1.1). Similarly, using the second form of $l(z)$ we have

$$f(z) \leq \frac{f(x) - f(y)}{x - y} (z - y) + f(y).$$
so (1.2) follows by first moving \( f(y) \) to left and then dividing by \( z - y \). \( \square \)

Geometrically this is evident. We fix \( x \) first and consider the point \( z \) moving from \( x \) to \( y \), (1.1) tells us that the slope keeps increasing. On the other hand, we fix \( y \) and consider the point \( z \) moving from \( x \) to \( y \), (1.2) tells us that again the slope increases.

**Proposition 1.2.** Let \( f \) be defined on \( I \). Then \( f \) is convex if and only if for \( x < z < y \) in \( I \),

\[
\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(z)}{y - z}.
\]

**Proof.** This inequality can be rewritten as

\[
f(z)(y - z) - f(x)(y - z) \leq f(y)(z - x) - f(z)(z - x),
\]

which is the same as

\[
f(z)(y - x) \leq f(y)(z - x) + f(x)(y - z) = (f(y) - f(x))(z - x) + f(x)(y - x).
\]

Now (1.1) follows by dividing both sides by \( y - x \). By Proposition 1 \( f \) is convex. We can reverse the reasoning to get the converse. \( \square \)

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

\[
f'_-(x) \leq f'_+(x), \ \forall x \in I, \tag{1.3}
\]

and

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

In particular, \( f \) is continuous in \( I \).

We note that \( f \) is right continuous at \( x \) if \( f^+(x) \) exists and is left continuous at \( x \) if \( f^-(x) \) exists, see the Lemma 1.5 below. Hence it is continuous at \( x \) if both one-sided derivatives exist at \( x \). We point out that this theorem does not necessarily hold on a closed interval. For instance, let \( f \) be a continuous convex function on \([a,b]\) and define another function \( g \) which is equal to \( f \) on \((a,b)\), but assign its values at the endpoints so that \( g(a) > f(a) \) and \( g(b) > f(a) \). Then \( g \) is convex on \([a,b]\) but not continuous at \( a, b \).

**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. (If you are not sure why this is true, see the Lemma 1.4.) In other words, \(f'_+(x)\) exists. Notice that we still have

\[f'_+(x) \geq \frac{f(x) - f(x_0)}{x - x_0},\]

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)

\[f'_+(x) \geq f'_-(x)\]

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

\[f'_+(x) \leq \frac{f(y) - f(x)}{y - x};\]

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

\[\frac{f(y) - f(x)}{y - x} \leq f'_-(y),\]

whence (1.4) follows.

\[\square\]

**Lemma 1.4.** Let \(h\) be an increasing function on \((a,b)\). Suppose that \(h(t) \geq -M, \forall t \in (a,b)\), for some constant \(M\). Then \(\lim_{t \to a^+} h(t)\) exists.

**Proof.** We fix a sequence \(\{t_n\}\) in \((a,b)\) satisfying \(t_n \to a^+\). Since \(h\) is increasing and \(h \geq -M\), \(\{h(t_n)\}\) is a decreasing sequence bounded from below, so \(A = \lim_{n \to \infty} h(t_n)\) must exist. For each \(\varepsilon > 0\), there is some \(n_0\) such that \(0 \leq h(t_n) - A < \varepsilon\) for all \(n \geq n_0\). Therefore, for all \(t < t_{n_0}\), \(h(t) - A \leq h(t_{n_0}) - A < \varepsilon\). On the other hand, since \(t_n \to a^+\), we can find some \(n_1\) such that \(h(t_{n_1}) \leq h(t)\). Thus, \(0 \leq h(t_{n_1}) - A \leq h(t) - A\). By taking \(\delta = t_{n_0} - a\), we have \(0 \leq h(t) - A < \varepsilon\) for all \(t \in (a,a + \delta)\).

\[\square\]

**Lemma 1.5.** Let \(f\) be a function on \((a,b)\) and \(c \in (a,b)\). Then \(f\) is right continuous (resp. left continuous) at \(c\) if \(f'_+(c)\) (resp. \(f'_-(c)\)) exists. Hence conclude that \(f\) is continuous at \(c\) if both one-sided derivatives exist at \(c\).
Proof. Assume $f_+(c)$ exists. Taking $\varepsilon = 1$, there exists some $\delta$ such that

$$\left| \frac{f(x) - f(c)}{x-c} - f_+(c) \right| < 1, \quad \forall x \in (c, c+\delta).$$

It follows that

$$(f_+(c) - 1)(x-c) < f(x) - f(c) < (f_+(c) + 1)(x-c), \forall x \in (c, c+\delta).$$

Hence

$$\lim_{x \to c^+} (f_+(c) + 1)(x-c) \leq \lim_{x \to c^+} (f(x) - f(c)) \leq \lim_{x \to c^+} (f_+(c) + 1)(x-c),$$

which forces that

$$\lim_{x \to c^+} (f(x) - f(c)) = 0.$$

The other case can be treated similarly. \qed

The following far-reaching theorem is for optional reading.

**Theorem 1.6.** *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 $x \in [a, b]$,

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

and the estimate

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

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 \( D_k : x_1 < x_2 < \ldots < x_n \). Then

\[
\begin{align*}
  f'_+(x_n) - f'_-(x_1) &= (f'_+(x_n) - f'_-(x_n)) + (f'_+(x_n) - f'_-(x_{n-1})) + \cdots + (f'_-(x_2) - f'_-(x_1)) \\
  &\geq (f'_+(x_n) - f'_-(x_n)) + (f'_+(x_{n-1}) - f'_-(x_{n-1})) + (f'_+(x_{n-2}) - f'_-(x_{n-2})) + \cdots + (f'_+(x_1) - f'_-(x_1)) \\
  &\geq \frac{n}{k},
\end{align*}
\]

which imposes a bound on \( n \): \( n \leq 4kM/\delta \). \( \square \)

When \( f \) is differentiable, Theorem 1.3 asserts that \( f' \) is increasing. The converse is also true.

**Theorem 1.7.** Let \( f \) be differentiable in \( I \). It is convex if and only if \( f' \) is increasing.

**Proof.** Theorem 1.3 asserts that \( f' \) is increasing if \( f \) is convex and differentiable. To show that converse, 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 \( f'(c_1) \leq f'(c_2) \) we get

\[
\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(z)}{y - z},
\]

which, by Proposition 1.2, implies that \( f \) is convex. \( \square \)

**Theorem 1.8.** Let \( f \) be twice differentiable in \( I \). It is convex if and only if \( f'' \geq 0 \).

**Proof.** When \( f \) is convex, \( f' \) is increasing and so \( f'' \geq 0 \). On the other hand, \( f'' \geq 0 \) implies that \( f' \) is increasing and hence convex. \( \square \)

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. Likewise, a function is \textbf{strictly concave} if its negative is strictly convex.

**Proposition 1.9.** The function $f$ is strictly convex on $I$ provided one of the followings hold:

(a) $f$ is differentiable and $f'$ is strictly increasing; or

(b) $f$ 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 $(0, \infty)$.

Convexity is a breeding ground for inequalities. We establish a fundamental one here.

**Theorem 1.10 (Jensen’s Inequality).** For a convex function $f$ on the interval $I$, let $x_1, x_2, \cdots, x_n \in I$ and $\lambda_1, \lambda_2, \cdots, \lambda_n \in (0, 1)$ satisfying $\sum_{j=1}^{n} \lambda_j = 1$. Then

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

When $f$ is strictly convex, equality sign in this inequality holds if and only if $x_1 = x_2 = \cdots = x_n$.

Perhaps we need to explain why the linear combination is still contained in the same interval. WLOG let $x_1 \leq x_2 \leq \cdots \leq x_n$. Then

$$\sum_j \lambda_j x_j \leq \sum_{j} \lambda_j x_n = x_n,$$

$$\sum_j \lambda_j x_j \geq \sum_{j} \lambda_j x_1 = x_1,$$

together imply that $\sum_j \lambda_j x_j$ is bounded between $x_1$ and $x_n$ and hence belongs to $I$.

Many well-known inequalities including the AM-GM inequality and H"older inequality are special cases of Jensen’s inequality. Some of them are found in the exercise.

**Proof.** We prove Jensen’s inequality by an inductive argument on the number of points. When $n = 2$, the inequality follows from the definition of convexity.
Assuming that it is true for \( n - 1 \) many points, we show its validity for \( n \) many points. Let \( \lambda_1, \cdots, \lambda_n \in (0, 1), \sum_j \lambda_j = 1 \) and let

\[
y = \sum_{j=1}^{n-1} \frac{\lambda_j}{1 - \lambda_n} x_j.
\]

Using first the definition of convexity and then the induction hypothesis,

\[
f(\lambda_1 x_1 + \cdots + \lambda_n x_n) = f((1 - \lambda_n)y + \lambda_n x_n) \\
\leq (1 - \lambda_n)f(y) + \lambda_n f(x_n) \\
= (1 - \lambda_n)f\left(\sum_{j=1}^{n-1} \frac{\lambda_j}{1 - \lambda_n} x_j\right) + \lambda_n f(x_n) \\
\leq (1 - \lambda_n)\sum_{j=1}^{n-1} \frac{\lambda_j}{1 - \lambda_n} f(x_j) + \lambda_n f(x_n) \\
= \sum_{j=1}^n \lambda_j f(x_j).
\]

When \( f \) is strictly convex, it follows straightly from definition that the strict inequality sign in Jensen’s inequality holds when \( n = 2, x_1 \neq x_2 \). In general, let us assume that the strictly inequality sign holds when \( x_1, \cdots, x_{n-1} \) are distinct and prove it when \( x_1, \cdots, x_n \) are not all equal. For, when all \( x_1, \cdots, x_n \) are distinct, the second \( \leq \) in the above inequalities becomes \(<\) due to the induction hypothesis and hence the strict inequality holds for \( n \). When some \( x_j \)'s are equal, we can group the expression \( \sum_{j=1}^n \lambda_j x_j \) into \( \sum_{j=1}^m \mu_j y_j \) where all \( y_j \)'s are distinct and \( m \) is less than \( n \). In this case the desired result comes from the induction hypothesis.

When \( \lambda_j \in [0, 1] \), let \( I_1 = \{ j : \lambda_j \in (0, 1) \} \) and \( I_2 = \{ j : \lambda_j = 0 \} \). Then in the strictly convex case, equality sign holds if and only if \( x_j = x_k \) for \( j, k \in I_1 \). The proof is essentially the same after observing that \( \lambda_j x_j = 0 \) and \( \lambda_j f(x_j) = 0 \) for \( j \in I_2 \) as well as \( \sum_{j \in I_1} \lambda_j = 1 \).

Jensen’s inequality is applied to the strictly convex function \( e^x \) to yield

\[
e^{\sum_{j=1}^n \lambda_j x_j} \leq \sum_{j=1}^n \lambda_j e^{x_j}.
\]
It can be rewritten as the generalized Young’s inequality
\[ a_1 a_2 \cdots a_n \leq \frac{a_1^{p_1}}{p_1} + \frac{a_2^{p_2}}{p_2} + \cdots + \frac{a_n^{p_n}}{p_n} \]
where
\[ a_j > 0, \sum_j \frac{1}{p_j} = 1, \quad p_j > 1, \quad j = 1, \cdots, n. \]
Moreover, the equality sign in this inequality holds if and only if all \( a_j^{p_j}, j = 1, \cdots, n, \) are equal. Taking and \( x_j = a_j^{p_j} \) and \( p_j = n \) for all \( j \) in the general Young’s Inequality, we recover the AM-GM Inequality
\[ (x_1 x_2 \cdots x_n)^{1/n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}, \quad x_j > 0, j = 1, \cdots, n, \]
with equality holds if and only if all \( x_j \)'s are equal. You may use the function \(- \log x\) instead of \( e^x\) to obtain the same results. In the exercises other inequalities following from Jensen’s are present.

Finally, we remark that in some books convexity is defined by a weaker condition, namely, a function \( f \) on \( I \) is convex if it satisfies
\[ f \left( \frac{x + y}{2} \right) \leq \frac{1}{2} \left( f(x) + f(y) \right), \quad \forall x, y \in I. \] (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 \( I \). I will leave it as an exercise. However, this conclusion does not hold without continuity. You may google under “weakly convex and continuity” for further information.