1 Convex functions and Jensen’s inequality

In the sequel $I$ will denote an interval on the real axis $\mathbb{R}$. The interval $I$ could be of the form $[a, b], (a, b], [a, b), (-\infty, b], [a, \infty)$ etc.

**Definition 1.1.** (a) A function $f : I \to \mathbb{R}$ is called **convex** (resp. **concave**) if for any two points $x_0, x_1 \in I$ the portion of the graph of $f$ over $[x_0, x_1]$ lies below (resp. above) the line connecting the endpoints of this portion of the graph.

![Convex and Concave Functions](Image)

**Figure 1:** A convex and a concave function
**Theorem 1.2.** Suppose \( f : I \to \mathbb{R} \) is a twice differentiable function. Then the following statements are equivalent.

(i) \( f \) is convex.
(ii) For any \( x_1, x_2 \in I \) and any \( w_1, w_2 \in [0, 1] \) such that \( w_1 + w_2 = 1 \) we have

\[
f(w_1 x_1 + w_2 x_2) \leq w_1 f(x_1) + w_2 f(x_2).
\]

(iii) The derivative \( f'(x) \) of \( f \) is increasing.
(iv) \( f''(x) \geq 0 \), for any \( x \in I \).

Since a function is convex iff \(-f\) is concave we deduce the following result.

**Corollary 1.3.** Suppose \( f : I \to \mathbb{R} \) is a twice differentiable function. Then the following statements are equivalent.

(i) \( f \) is concave.
(ii) For any \( x_1, x_2 \in I \) and any \( w_1, w_2 \in [0, 1] \) such that \( w_1 + w_2 = 1 \) we have

\[
f(w_1 x_1 + w_2 x_2) \geq w_1 f(x_1) + w_2 f(x_2).
\]

(iii) The derivative \( f'(x) \) of \( f \) is decreasing.
(iv) \( f''(x) \leq 0 \), for any \( x \in I \).

**Exercise 1.1.** Prove Theorem 1.2. \( \square \)

**Example 1.4.** Using the above results we can produce easily many examples of convex/concave functions. For example the function

\[
(0, \infty) \to \mathbb{R}, \quad x \mapsto x^\alpha
\]

is convex for \( \alpha > 1 \) or \( \alpha < 0 \) and concave if \( \alpha \in (0, 1) \). In particular, the functions \( x \mapsto x^2 \) and \( x \mapsto \frac{1}{x} \) are convex while the function \( x \mapsto \sqrt{x} \) is concave.

The function \( \mathbb{R} \to \mathbb{R}, \ x \mapsto e^x \) is convex while the function \( (0, \infty) \to \mathbb{R}, \ x \mapsto \log x \) is concave.

**Exercise 1.2.** Prove that if \( f, g : \mathbb{R} \to \mathbb{R} \) are twice differentiable functions such that \( g(x) \) is convex and \( f \) is convex and increasing then the composition \( x \mapsto f(g(x)) \) is also convex. \( \square \)

**Theorem 1.5 (Jensen’s Inequality).** Suppose \( f : I \to \mathbb{R} \) is a twice differentiable function satisfying

\[
f''(x) \geq 0, \ \forall x \in I.
\]

Then for any integer \( n \geq 2 \), any \( x_1, x_2, \ldots, x_n \in I \) and any \( w_1, \ldots, w_n \in [0, \infty) \) such that \( w_1 + \cdots + w_n = 1 \) we have

\[
f(w_1 x_1 + \cdots + w_n x_n) \leq w_1 f(x_1) + \cdots + w_n f(x_n).
\]
Remark 1.6. (a) If the function in Theorem 1.5 satisfies the more stringent condition
\[ f''(x) > 0, \forall x \in I. \] (>)
then in the inequality (1.1) becomes an equality if and only if \( x_1 = \cdots = x_n \).
(b) If \( f : I \to \mathbb{R} \) is a twice differentiable function satisfying
\[ f''(x) \leq 0, \forall x \in I. \] (≥)
then for any integer \( n \geq 2 \), any \( x_1, x_2, \ldots, x_n \in I \) and any \( w_1, \ldots, w_n \in [0, \infty) \) such that \( w_1 + \cdots + w_n = 1 \) we have
\[ f\left(w_1 x_1 + \cdots + w_n x_n\right) \geq w_1 f(x_1) + \cdots + w_n f(x_n). \] (1.2)

Exercise 1.3. Prove Theorem 1.5 by induction using Theorem 1.2.

2 Some classical applications of Jensen’s inequality

We can get many interesting and nontrivial results by looking at concrete choices of \( f, x_i \) and \( w_i \) in Jensen’s inequality. Suppose we choose \( w_1 = \cdots = w_n = \frac{1}{n} \). Then for any twice differentiable function such that \( f'' \geq 0 \) on \( I \) we have
\[ f\left(\frac{x_1 + \cdots + x_n}{n}\right) \leq \frac{f(x_1) + \cdots + f(x_n)}{n}, \forall x_1, \ldots, x_n \in I. \] (2.1)
If in the above inequality we choose \( f(x) = x^\alpha, \alpha > 1 \) we obtain
\[ \left(\frac{x_1 + \cdots + x_n}{n}\right)^\alpha \leq \frac{x_1^\alpha + \cdots + x_n^\alpha}{n}, \forall x_1, \ldots, x_n > 0, \forall \alpha > 1. \] (2.2)
For example, if \( \alpha = 2 \) we get
\[ \left(\frac{x_1 + \cdots + x_n}{n}\right)^2 \leq \frac{x_1^2 + \cdots + x_n^2}{n}, \forall x_1, \ldots, x_n > 0. \] (2.3)
Let us specialize even more in (2.3) assume \( n = 2 \) and \( x_1 = x, x_2 = y \). We get
\[ \left(\frac{x + y}{2}\right)^2 \leq \frac{x^2 + y^2}{2}, \forall x, y \geq 0. \] (2.4)

Exercise 2.1. (a) Can you prove (2.4) directly, using only elementary algebra?
(b) Using only elementary algebra prove the inequality (2.4) in the special case \( n = 3 \).

Exercise 2.2. Prove the following inequality
\[ n^2 \leq \left(\frac{1}{x_1} + \cdots + \frac{1}{x_n}\right)(x_1 + \cdots + x_n), \forall x_1, \ldots, x_n > 0 \] (2.5)
Example 2.1. Here is a simple geometric application of (2.3) Suppose we are given \(n\) nonoverlapping disks \(D_1, \cdots, D_n\) inside a plane region \(R\) of area \(A < \infty\). Denote by \(r_i\) the radius of \(D_i\). We will prove that
\[
r_1 + \cdots + r_n \leq \sqrt{\frac{nA}{\pi}}.
\]
Note that
\[
\sum_{i=1}^{n} \text{area } (D_i) \leq \text{area } (R) = A
\]
so that
\[
\pi \sum_{i=1}^{n} r_i^2 \leq A \iff \sum_{i=1}^{n} r_i^2 \leq \frac{A}{\pi}.
\]
Using (2.3) we deduce
\[
\frac{1}{n^2} \left( \sum_{i=1}^{n} r_i \right)^2 \leq \frac{1}{n} \sum_{i=1}^{n} r_i^2 \leq \frac{A}{n\pi}.
\]
Multiplying the last inequality by \(n^2\) we obtain
\[
\left( \sum_{i=1}^{n} r_i \right)^2 \leq \frac{nA}{\pi}
\]
which is the inequality we sought. \(\square\)

Suppose we have a function \(f : (0, \infty) \to \mathbb{R}\) such that \(f''(x) < 0\) for any \(x > 0\). If we choose \(w_1 = \cdots = w_n = \frac{1}{n}\) in (1.2) we deduce
\[
f\left(\frac{x_1 + \cdots + x_n}{n}\right) \geq \frac{f(x_1) + \cdots + f(x_n)}{n}, \ \forall x_1, \cdots, x_n > 0. \quad (2.6)
\]
Let us further analyze (2.6) for some very special choices of \(f\). If we choose \(f(x) = \sqrt{x}\) we deduce
\[
\sqrt{\frac{x_1 + \cdots + x_n}{n}} \geq \frac{\sqrt{x_1} + \cdots + \sqrt{x_n}}{n}, \ \forall x_1, \cdots, x_n > 0 \quad (2.7)
\]
If we choose \(f(x) = \log x\) then we deduce
\[
\log \frac{x_1 + \cdots + x_n}{n} \geq \frac{1}{n} \left( \log x_1 + \cdots + \log x_n \right) = \frac{1}{n} \log (x_1 \cdots x_n) = \log \sqrt[n]{x_1 \cdots x_n}.
\]
We deduce
\[
\frac{x_1 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n}, \ \forall x_1, \cdots, x_n > 0. \quad (2.8)
\]
The expression in the left-hand-side is called the arithmetic mean of the numbers \(x_1, \cdots, x_n\) while the expression in the right-hand-side is called the geometric mean. Thus we can rephrase (2.8) as
\[
\text{Arithmetic Mean} \geq \text{Geometric Mean}.
\]
Consider now the function \( f(x) = x^p, \ p > 1, \ x > 0 \). It satisfies Jensen’s inequality (1.1).

Let \( x_1, \ldots, x_n > 0 \) and \( u_1, \ldots, u_n \geq 0 \) such that \( U = u_1 + \cdots + u_n > 0 \). If we set

\[
    w_i = \frac{u_i}{U}
\]

then

\[
w_1 + \cdots + w_n = 1
\]

and using Jensen’s inequality we have

\[
\left( \frac{u_1 x_1 + \cdots + u_n x_n}{U} \right)^p \leq \frac{u_1 x_1^p + \cdots + u_n x_n^p}{U}.
\]

Multiplying both sides of the above inequality by \( U^p \) we deduce

\[
\left( u_1 x_1 + \cdots + u_n x_n \right)^p \leq (u_1 x_1^p + \cdots + u_n x_n^p)^{p-1}, \ \forall u_i, x_j > 0. \tag{2.9}
\]

Let us specialize \( p = 2 \) in the above inequality.

Suppose \( a_i, b_j \) are nonzero real numbers, \( \forall i, j = 1, \ldots, n \). Choose positive numbers \( u_i, x_j \) such that

\[
u_i x_i^2 := a_i^2, \quad u_i = b_i^2 \quad \iff \quad x_i = \frac{|a_i|}{|b_i|}, \ u_i = b_i^2.
\]

Observe that \( u_i x_i = |a_i b_i| \). Using these numbers in (2.9) we deduce that \( \forall a_i, b_j \in \mathbb{R} \).

\[
\left( a_1 b_1 + \cdots + a_n b_n \right)^2 \leq \left( |a_1 b_1| + \cdots + |a_n b_n| \right)^2 \leq \left( a_1^2 + \cdots + a_n^2 \right) \cdot \left( b_1^2 + \cdots + b_n^2 \right)^2. \tag{2.10}
\]

The last inequality is usually known as the Cauchy-Schwartz inequality.

**Exercise 2.3.** Let \( p > 1 \) and \( q = \frac{p}{p-1} \), i.e. \( 1 = \frac{1}{p} + \frac{1}{q} \). Use the inequality (2.9) to prove that for any real numbers \( a_i, b_i \) we have the Hölder inequality

\[
\left( |a_1 b_1| + \cdots + |a_n b_n| \right) \leq \left( |a_1|^p + \cdots + |a_n|^p \right)^{\frac{1}{p}} \left( |b_1|^q + \cdots + |b_n|^q \right)^{\frac{1}{q}}
\]

(Observe that when \( p = 2 \) so that \( q = 2 \) the Hölder inequality becomes the Cauchy-Schwartz inequality.)

For a rich presentation of this subject we refer to the classical source, [1]

**References**