Convexity, Inequalities, and Norms

1. Convex Functions

You are probably familiar with the notion of concavity of functions. Given a twice-differentiable function \( \varphi : \mathbb{R} \to \mathbb{R} \),

- We say that \( \varphi \) is **concave** (or **concave down**) if \( \varphi''(x) \leq 0 \) for all \( x \in \mathbb{R} \).
- We say that \( \varphi \) is **convex** (or **concave up**) if \( \varphi''(x) \geq 0 \) for all \( x \in \mathbb{R} \).

For example, a quadratic function

\[
\varphi(x) = ax^2 + bx + c
\]

is concave if \( a \leq 0 \), and is convex if \( a \geq 0 \).

Unfortunately, the definitions above are not sufficiently general, since they require \( \varphi \) to be twice differentiable. Instead, we will use the following definitions:

---

**Definition: Concave and Convex Functions**

Let \(-\infty \leq a < b \leq \infty\), and let \( \varphi : (a, b) \to \mathbb{R} \) be a function.

1. We say that \( \varphi \) is **concave** if

\[
\varphi((1 - \lambda)x + \lambda y) \geq (1 - \lambda)\varphi(x) + \lambda\varphi(y)
\]

for all \( x, y \in (a, b) \) and \( \lambda \in [0, 1] \).

2. We say that \( \varphi \) is **convex** if

\[
\varphi((1 - \lambda)x + \lambda y) \leq (1 - \lambda)\varphi(x) + \lambda\varphi(y)
\]

for all \( x, y \in (a, b) \) and \( \lambda \in [0, 1] \).
Convexity, Inequalities, and Norms

Figure 1: For a convex function, every chord lies above the graph.

Geometrically, the function

$$
\lambda \mapsto \((1 - \lambda)x + \lambda y, (1 - \lambda)\varphi(x) + \lambda\varphi(y)\), \quad 0 \leq \lambda \leq 1
$$

is a parametrization of a line segment in $\mathbb{R}^2$. This line segment has endpoints $(x, \varphi(x))$ and $(y, \varphi(y))$, and is therefore a chord of the graph of $\varphi$ (see figure 1). Thus our definitions of concave and convex can be interpreted as follows:

- A function $\varphi$ is convex if every chord lies above the graph of $\varphi$.
- A function $\varphi$ is concave if every chord lies below the graph of $\varphi$.

Another fundamental geometric property of convex functions is that each tangent line lies entirely below the graph of the function. This statement can be made precise even for functions that are not differentiable:

**Theorem 1**  \hspace{1em} **Tangent Lines for Convex Functions**

\begin{itemize}
  \item Let $\varphi: (a, b) \to \mathbb{R}$ be a convex function. Then for every point $c \in (a, b)$, there exists a line $L$ in $\mathbb{R}^2$ with the following properties:
  \begin{enumerate}
    \item $L$ passes through the point $(c, \varphi(c))$.
    \item The graph of $\varphi$ lies entirely above $L$.
  \end{enumerate}
\end{itemize}

**Proof**  See exercise 1.  \hspace{1em} $\blacksquare$
Convexity, Inequalities, and Norms

Figure 2: A tangent line to $y = |x|$ at the point (0, 0).

We will refer to any line satisfying the conclusions of the above theorem as a **tangent line** for $\varphi$ at $c$. If $\varphi$ is not differentiable, then the slope of a tangent line may not be uniquely determined. For example, if $\varphi(x) = |x|$, then a tangent line for $\varphi$ at 0 may have any slope between $-1$ and 1 (see figure 2).

We shall use the existence of tangent lines to provide a geometric proof of the continuity of convex functions:

**Theorem 2**  **Continuity of Convex Functions**

*Every convex function is continuous.*

**PROOF**  Let $\varphi: (a, b) \to \mathbb{R}$ be a convex function, and let $c \in (a, b)$. Let $L$ be the linear function whose graph is a tangent line for $\varphi$ at $c$, and let $P$ be a piecewise-linear function consisting of two chords to the graph of $\varphi$ meeting at $c$ (see figure 3). Then $L \leq \varphi \leq P$ in a neighborhood of $c$, and $L(c) = \varphi(c) = P(c)$. Since $L$ and $P$ are continuous at $c$, it follows from the Squeeze Theorem that $\varphi$ is also continuous at $c$.

We now come to one of the most important inequalities in analysis:
Theorem 3  Jensen’s Inequality (Finite Version)

Let \( \varphi : (a, b) \rightarrow \mathbb{R} \) be a convex function, and let \( x_1, \ldots, x_n \in (a, b) \). Then

\[
\varphi(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 \varphi(x_1) + \cdots + \lambda_n \varphi(x_n)
\]

for any \( \lambda_1, \ldots, \lambda_n \in [0, 1] \) satisfying \( \lambda_1 + \cdots + \lambda_n = 1 \).

PROOF  Let \( c = \lambda_1 x_1 + \cdots + \lambda_n x_n \), and let \( L \) be a linear function whose graph is a tangent line for \( \varphi \) at \( c \). Since \( \lambda_1 + \cdots + \lambda_n = 1 \), we know that

\[
L(\lambda_1 x_1 + \cdots + \lambda_n x_n) = \lambda_1 L(x_1) + \cdots + \lambda_n L(x_n).
\]

Since \( L \leq \varphi \) and \( L(c) = \varphi(c) \), we conclude that

\[
\varphi(c) = L(c) = L(\lambda_1 x_1 + \cdots + \lambda_n x_n)
= \lambda_1 L(x_1) + \cdots + \lambda_n L(x_n)
\leq \lambda_1 \varphi(x_1) + \cdots + \lambda_n \varphi(x_n).
\]

This statement can be generalized from finite sums to integrals. Specifically, we can replace the points \( x_1, \ldots, x_n \) by a function \( f : X \rightarrow \mathbb{R} \), and we can replace the weights \( \lambda_1, \ldots, \lambda_n \) by a measure \( \mu \) on \( X \) for which \( \mu(X) = 1 \).
Theorem 4  Jensen’s Inequality (Integral Version)

Let $(X,\mu)$ be a measure space with $\mu(X) = 1$. Let $\varphi: (a,b) \to \mathbb{R}$ be a convex function, and let $f: X \to (a,b)$ be an $L^1$ function. Then

$$\varphi\left(\int_X f\right) \leq \int_X (\varphi \circ f)$$

Proof  Let $c = \int_X f$, and let $L$ be a linear function whose graph is a tangent line for $\varphi$ at $c$. Since $\mu(X) = 1$, we know that $L(\int_X f) = \int_X (L \circ f)$. Since $L(c) = \varphi(c)$ and $L \leq \varphi$, this gives

$$\varphi(c) = L(c) = L(\int_X f) = \int_X (L \circ f) \leq \int_X (\varphi \circ f).$$

2. Means

You are probably aware of the arithmetic mean and geometric mean of positive numbers:

$$\frac{x_1 + \cdots + x_n}{n} \quad \text{and} \quad \sqrt[n]{x_1 \cdots x_n}.$$  

More generally, we can define weighted versions of these means. Given positive weights $\lambda_1, \ldots, \lambda_n$ satisfying $\lambda_1 + \cdots + \lambda_n = 1$, the corresponding weighted arithmetic and geometric means are

$$\lambda_1 x_1 + \cdots + \lambda_n x_n \quad \text{and} \quad x_1^{\lambda_1} \cdots x_n^{\lambda_n}.$$  

These reduce to the unweighted means in the case where $\lambda_1 = \cdots = \lambda_n = 1/n$.

Arithmetic and geometric means satisfy a famous inequality, namely that the geometric mean is always less than or equal to the arithmetic mean. This turns out to be a simple application of Jensen’s inequality:

Theorem 5  AM–GM Inequality

Let $x_1, \ldots, x_n > 0$, and let $\lambda_1, \ldots, \lambda_n \in [0,1]$ so that $\lambda_1 + \cdots + \lambda_n = 1$. Then

$$x_1^{\lambda_1} \cdots x_n^{\lambda_n} \leq \lambda_1 x_1 + \cdots + \lambda_n x_n.$$
Figure 4: A visual proof that $\sqrt{ab} < (a + b)/2$.

**PROOF** This theorem is equivalent to the convexity of the exponential function (see figure 4). Specifically, we know that

$$e^{\lambda_1 t_1 + \cdots + \lambda_n t_n} \leq \lambda_1 e^{t_1} + \cdots + \lambda_n e^{t_n}$$

for all $t_1, \ldots, t_n \in \mathbb{R}$. Substituting $x_i = e^{t_i}$ gives the desired result. ■

Applying the same reasoning using the integral version of Jensen’s inequality gives

$$\exp \left( \int_X \log f \right) \leq \int_X f$$

for any $L^1$ function $f : X \to (0, \infty)$, where $(X, \mu)$ is a measure space with $\mu(X) = 1$.

We can rescale to get a version of this inequality that applies whenever $\mu(X)$ is finite and nonzero

$$\exp \left( \frac{1}{\mu(X)} \int_X \log f \right) \leq \frac{1}{\mu(X)} \int_X f.$$

Note that the quantity on the right is simply the average value of $f$ on $X$. The quantity on the left can be thought of as the (continuous) geometric mean of $f$.

**p-Means**

There are many important means in mathematics and science, beyond just the arithmetic and geometric means. For example, the **harmonic mean** of positive numbers $x_1, \ldots, x_n$ is

$$\frac{n}{1/x_1 + \cdots + 1/x_n}.$$
This mean is used, for example, in calculating the average resistance of resistors in parallel. For another example, the **Euclidean mean** of \(x_1, \ldots, x_n\) is
\[
\sqrt[2]{\frac{x_1^2 + \cdots + x_n^2}{n}}.
\]
This mean is used to average measurements taken for the standard deviation of a random variable.

The AM–GM inequality can be extended to cover both of these means. In particular, the inequality
\[
\frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}} \leq \sqrt[2]{x_1 \cdots x_n} \leq \frac{x_1 + \cdots + x_n}{n} \leq \sqrt[2]{\frac{x_1^2 + \cdots + x_n^2}{n}}.
\]
holds for all \(x_1, \ldots, x_n \in (0, \infty)\).

Both of these means are examples of \(p\)-means:

**Definition: \(p\)-Means**

Let \(x_1, \ldots, x_n > 0\). If \(p \in \mathbb{R} - \{0\}\), the **\(p\)-mean** of \(x_1, \ldots, x_n\) is
\[
\left(\frac{x_1^p + \cdots + x_n^p}{n}\right)^{1/p}.
\]
For example:

- The 2-mean is the same as the Euclidean mean.
- The 1-mean is the same as the arithmetic mean.
- The \((-1)\)-mean is the same as the harmonic mean.

Though it may not be obvious, the geometric mean also fits into the family of \(p\)-means. In particular, it is possible to show that
\[
\lim_{p \to 0} \left(\frac{x_1^p + \cdots + x_n^p}{n}\right)^{1/p} = \sqrt[2]{x_1 \cdots x_n}
\]
for any \(x_1, \ldots, x_n \in (0, \infty)\). Thus, we may think of the geometric mean as the 0-mean.

It is also possible to use limits to define means for \(\infty\) and \(-\infty\). It turns out the the \(\infty\)-mean of \(x_1, \ldots, x_n\) is simply \(\max(x_1, \ldots, x_n)\), while the \((-\infty)\)-mean is \(\min(x_1, \ldots, x_n)\).

As with the arithmetic and geometric means, we can also define weighted versions of \(p\)-means. Given positive weights \(\lambda_1, \ldots, \lambda_n\) satisfying \(\lambda_1 + \cdots + \lambda_n = 1\), the corresponding weighted \(p\)-mean is
\[
\left(\lambda_1 x_1^p + \cdots + \lambda_n x_n^p\right)^{1/p}.
\]
As you may have guessed, the $p$-means satisfy a generalization of the AM-GM inequality:

**Theorem 6  Generalized Mean Inequality**

Let $x_1, \ldots, x_n > 0$, and let $\lambda_1, \ldots, \lambda_n \in [0, 1]$ so that $\lambda_1 + \cdots + \lambda_n = 1$. Then

\[
\lambda_1 x_1^p + \cdots + \lambda_n x_n^p \leq \lambda_1 x_1^q + \cdots + \lambda_n x_n^q
\]

for all $p, q \in \mathbb{R} - \{0\}$.

**PROOF** If $p = 1$ and $q > 1$, this inequality takes the form

\[
(\lambda_1 x_1 + \cdots + \lambda_n x_n)^q \leq \lambda_1 x_1^q + \cdots + \lambda_n x_n^q
\]

which follows immediately from the convexity of the function $\varphi(x) = x^q$.

The case where $0 < p < q$ follows from this. Specifically, since $q/p > 1$, we have

\[
(\lambda_1 x_1^p + \cdots + \lambda_n x_n^p)^{q/p} \leq \lambda_1 x_1^q + \cdots + \lambda_n x_n^q
\]

and the desired inequality follows. Cases involving negative values of $p$ or $q$ are left as an exercise to the reader.

Applying the same reasoning using the integral version of Jensen’s inequality gives

\[
p \leq q \implies \left( \int_X f^p \right)^{1/p} \leq \left( \int_X f^q \right)^{1/q}
\]

for any $L^1$ function $f: X \to (0, \infty)$, where $(X, \mu)$ is a measure space with a total measure of one.

### 3. Norms

As a final application of convexity, we shall prove some of the fundamental inequalities for $p$-norms. First, recall that the $p$-norm on $\mathbb{R}^n$ is defined by the formula

\[
\| (x_1, \ldots, x_n) \|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}
\]

We have never proved that this is in fact a norm, and it is hardly obvious that it satisfies the triangle inequality. This is our first task.
Theorem 7  Minkowski’s Inequality

If \( u, v \in \mathbb{R}^n \) and \( p \in [1, \infty) \), then

\[
\|u + v\|_p \leq \|u\|_p + \|v\|_p
\]

Proof  Since \( p \geq 1 \), the function \( x \mapsto |x|^p \) is convex. It follows that

\[
\|(1 - \lambda)u + \lambda v\|_p^p = \sum_{i=1}^{n} |(1 - \lambda)u_i + \lambda v_i|^p
\]

\[
\leq \sum_{i=1}^{n} (1 - \lambda)|u_i|^p + \lambda|v_i|^p = (1 - \lambda)\|u\|_p^p + \lambda\|v\|_p^p
\]

for all \( u \) and \( v \). In particular, this proves that \( \|(1 - \lambda)u + \lambda v\|_p \leq 1 \) whenever \( \|u\| = \|v\| = 1 \). (That is, the unit ball with respect to \( \| \cdot \|_p \) is convex.)

From this the triangle inequality follows. In particular, we may assume that \( u \) and \( v \) are nonzero. Then \( u/\|u\|_p \) and \( v/\|v\|_p \) are unit vectors, so

\[
\frac{\|u + v\|_p}{\|u\|_p + \|v\|_p} = \left\| \frac{\|u\|_p u}{\|u\|_p + \|v\|_p} + \frac{\|v\|_p v}{\|u\|_p + \|v\|_p} \right\| \leq 1.
\]

The 2-norm is the usual Euclidean norm on \( \mathbb{R}^n \), defined by the dot product of vectors:

\[
\|v\| = \sqrt{v \cdot v}.
\]

This satisfies the Cauchy-Schwarz Inequality

\[
|u \cdot v| \leq \|u\|_2 \|v\|_2
\]

for all \( u, v \in \mathbb{R}^n \). Our next task is to prove a generalization of this known as Hölder’s Inequality.

Lemma 8  Young’s Inequality

If \( x, y \in [0, \infty) \) and \( p, q \in (1, \infty) \) so that \( 1/p + 1/q = 1 \), then

\[
xy \leq \frac{x^p}{p} + \frac{y^q}{q}
\]
PROOF This can be written
\[
(x^p)^{1/p}(y^q)^{1/q} \leq \frac{1}{p} x^p + \frac{1}{q} y^q
\]
which is an instance of the weighted AM–GM inequality. ■

**Theorem 9** Hölder’s Inequality

Let \( u, v \in \mathbb{R}^n \), and let \( p, q \in (1, \infty) \) so that \( 1/p + 1/q = 1 \). Then
\[
|u \cdot v| \leq \|u\|_p \|v\|_q.
\]

**PROOF** By Young’s inequality,
\[
|u \cdot v| \leq |u_1v_1| + \cdots + |u_nv_n|
\leq \frac{|u_1|^p + \cdots + |u_n|^p}{p} + \frac{|v_1|^q + \cdots + |v_n|^q}{q}
\leq \frac{\|u\|_p^p}{p} + \frac{\|v\|_q^q}{q}.
\]
In particular, if \( \|u\|_p = \|v\|_q = 1 \), then \( |u \cdot v| \leq 1/p + 1/q = 1 \), which proves Hölder’s Inequality in this case.

For the general case, we may assume that \( u \) and \( v \) are nonzero. Then \( u/\|u\|_p \) and \( v/\|v\|_q \) are unit vectors for their respective norms, and therefore
\[
\frac{\|u\|_p}{\|u\|_p} \cdot \frac{\|v\|_q}{\|v\|_q} \leq 1.
\]
Multiplying through by \( \|u\|_p \|v\|_q \) gives the desired result. ■

Both of these inequalities generalize easily to integrals. We leave the proofs to the reader:

**Theorem 10** Minkowski’s Inequality (Integral Version)

Let \((X, \mu)\) be a measure space, let \( f, g: X \to \mathbb{R} \) be measurable functions, and let \( p \in [1, \infty) \). Then
\[
\left( \int_X |f + g|^p \right)^{1/p} \leq \left( \int_X |f|^p \right)^{1/p} + \left( \int_X |g|^p \right)^{1/p}.
\]
**Theorem 11**  Hölder’s Inequality (Integral Version)

Let $(X, \mu)$ be a measure space, let $f, g: X \to \mathbb{R}$ be measurable functions, and let $p, q \in (1, \infty)$ so that $1/p + 1/q = 1$. Then

$$
\left| \int_X fg \right| \leq \left( \int_X |f|^p \right)^{1/p} \left( \int_X |g|^q \right)^{1/q}
$$

**Exercises**

1. a) Prove that a function $\varphi: \mathbb{R} \to \mathbb{R}$ is convex if and only if

$$
\frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(z) - \varphi(x)}{z - x} \leq \frac{\varphi(z) - \varphi(y)}{z - y}
$$

for all $x, y, z \in \mathbb{R}$ with $x < y < z$.

b) Use this characterization of convex functions to prove Theorem 1 on the existence of tangent lines.

2. Let $\varphi: (a, b) \to \mathbb{R}$ be a differentiable function. Prove that $\varphi$ is convex if and only if $\varphi'$ is non-decreasing.

3. Let $\varphi: \mathbb{R} \to \mathbb{R}$ be a convex function. Prove that

$$
\varphi((1 - \lambda)x + \lambda y) \geq (1 - \lambda)\varphi(x) + \lambda \varphi(y)
$$

for $\lambda \in \mathbb{R} - [0, 1]$. Use this to provide an alternative proof that $\varphi$ is continuous.

4. Prove that $\lim_{p \to 0} \left( \frac{x^p + y^p}{2} \right)^{1/p} = \sqrt{xy}$ and $\lim_{p \to \infty} \left( \frac{x^p + y^p}{2} \right)^{1/p} = \max(x, y)$.

5. a) If $x_1, \ldots, x_n, y_1, \ldots, y_n \in (0, \infty)$, prove that

$$
\sqrt[2n]{x_1 y_1 \cdots x_n y_n} \leq \sqrt{\frac{x_1 + \cdots + x_n}{n}} \sqrt{\frac{y_1 + \cdots + y_n}{n}}.
$$

That is, the arithmetic mean of geometric means is less than or equal to the corresponding geometric mean of arithmetic means.

b) If $\lambda, \mu \in [0, 1]$ and $\lambda + \mu = 1$, prove that

$$
\frac{x_1^\lambda y_1^\mu + \cdots + x_n^\lambda y_n^\mu}{n} \leq \left( \frac{x_1 + \cdots + x_n}{n} \right)^\lambda \left( \frac{y_1 + \cdots + y_n}{n} \right)^\mu.
$$
6. Prove the Generalized Mean Inequality (Theorem 6) in the case where \( p \) or \( q \) is negative.

7. Let \( f : [0, 1] \to \mathbb{R} \) be a bounded measurable function, and define \( \varphi : [1, \infty) \to \mathbb{R} \) by

\[
\varphi(p) = \int_{[0,1]} f^p. 
\]

Prove that \( \log \varphi \) is convex on \([1, \infty)\).

8. Prove that

\[
(1 + x^2 y + x^4 y^2)^3 \leq (1 + x^3 + x^6)^2 (1 + y^3 + y^6)
\]

for all \( x, y \in (0, \infty) \).

9. Let \((X, \mu)\) be a measure space, let \( f, g : X \to [0, \infty) \) be measurable functions, and let \( p, q, r \in (1, \infty) \) so that \( 1/p + 1/q = 1/r \). Prove that

\[
\left( \int_X (fg)^r \right)^{1/r} \leq \left( \int_X f^p \right)^{1/p} \left( \int_X g^q \right)^{1/q}
\]