Convexity, Inequalities, and Norms

Convex Functions

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

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

For example, a quadratic function

$$\varphi(x) \;=\; ax^2 + bx + c$$

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

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

Definition: Convex and Concave Functions
Let $-\infty \leq a < b \leq \infty$, and let $\varphi\colon (a,b) \to \mathbb{R}$ be a function.

1. We say that φ is **convex** if

$$\varphi\big((1-\lambda)x + \lambda y\big) \;\leq\; (1-\lambda)\varphi(x) + \lambda\varphi(y)$$

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

2. We say that φ is **concave** if

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

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

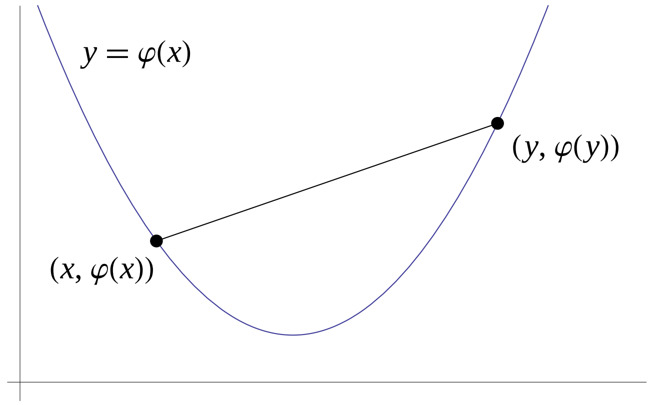


Figure 1: For a convex function, every chord lies above the graph.

Geometrically, the function

$$\lambda \mapsto \big((1-\lambda)x + \lambda y,\ (1-\lambda)\varphi(x) + \lambda\varphi(y)\big), \qquad 0 \le \lambda \le 1$$

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

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

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 Tangent Lines for Convex Functions

Let $\varphi\colon (a,b) \to \mathbb{R}$ 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:

1. *L passes through the point $\big(c, \varphi(c)\big)$.*

2. *The graph of φ lies entirely above L.*

PROOF See exercise 1. ■

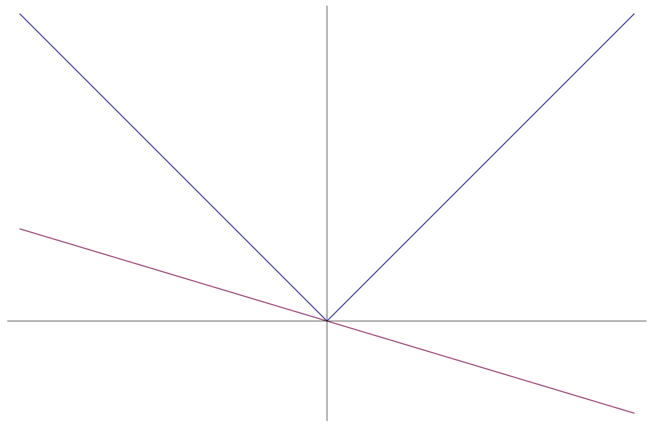


Figure 2: A tangent line to $y = |x|$ at the point $(0, 0)$.

We will refer to any line satisfying the conclusions of the above theorem as a **tangent line** for φ at c. If φ 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 φ 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 $\varphi\colon (a, b) \to \mathbb{R}$ be a convex function, and let $c \in (a, b)$. Let L be a linear function whose graph is a tangent line for φ at c, and let P be a piecewise-linear function consisting of two chords to the graph of φ 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 φ is also continuous at c. ■

We now come to one of the most important inequalities in analysis:

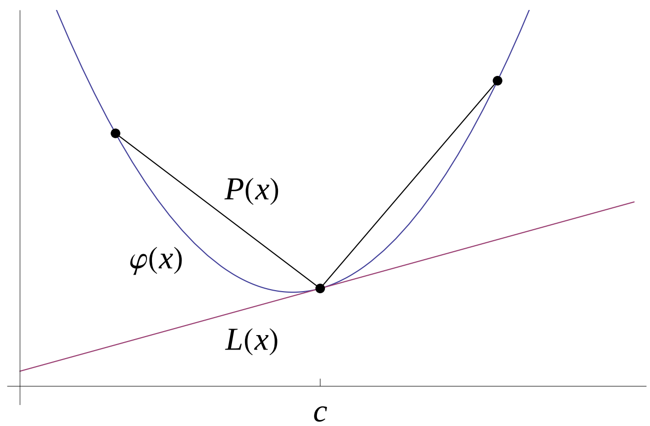


Figure 3: Using the Squeeze Theorem to prove that $\varphi(x)$ is continuous at c

Theorem 3 Jensen's Inequality (Finite Version)

Let $\varphi\colon (a,b) \to \mathbb{R}$ be a convex function, where $-\infty \leq a < b \leq \infty$, 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 φ 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

$$\begin{aligned} \varphi(c) \;=\; L(c) \;&=\; L(\lambda_1 x_1 + \cdots + \lambda_n x_n) \\ &=\; \lambda_1 L(x_1) + \cdots + \lambda_n L(x_n) \\ &\leq\; \lambda_1 \varphi(x_1) + \cdots + \lambda_n \varphi(x_n). \end{aligned}$$ ■

This statement can be generalized from finite sums to integrals. Specifically, we can replace the points $x_1, \ldots, x_n$ by a function $f\colon X \to \mathbb{R}$, and we can replace the weights $\lambda_1, \ldots, \lambda_n$ by a measure μ on X for which $\mu(X) = 1$.

Theorem 4 Jensen's Inequality (Integral Version)

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

$$\varphi\left(\int_X f\,d\mu\right) \;\le\; \int_X (\varphi\circ f)\,d\mu$$

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

$$\varphi(c) \;=\; L(c) \;=\; L\left(\int_X f\,d\mu\right) \;=\; \int_X (L\circ f)\,d\mu \;\le\; \int_X (\varphi\circ f)\,d\mu. \qquad \blacksquare$$

Means

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

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

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

$$\lambda_1x_1+\cdots+\lambda_nx_n \qquad \text{and} \qquad x_1^{\lambda_1}\cdots x_n^{\lambda_n}.$$

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

$$x_1^{\lambda_1}\cdots x_n^{\lambda_n} \;\le\; \lambda_1x_1+\cdots+\lambda_nx_n.$$

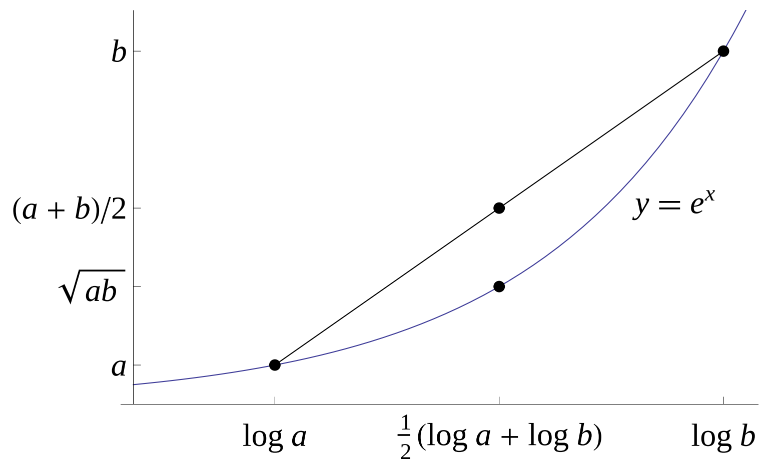


Figure 4: A visual proof that $\sqrt{ab} < (a+b)/2$.

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. ■

The following theorem generalizes this inequality to arbitrary measure spaces. The proof is essentially the same as the proof of the previous theorem.

Theorem 6 Integral AM–GM Inequality

Let (X,μ) be a measure space with $\mu(X) = 1$, and let $f\colon X \to (0,\infty)$ be a measurable function. Then

$$\exp\left(\int_X \log f\, d\mu\right) \;\leq\; \int_X f\, d\mu$$

PROOF Since the exponential function is convex, Jensen's inequality gives

$$\exp\left(\int_X \log f\, d\mu\right) \leq \int_X \exp(\log f)\, d\mu \;=\; \int_X f\, d\mu. \qquad ■$$

By the way, we can rescale to get a version of this inequality that applies whenever

$\mu(X)$ is finite and nonzero:

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

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

$$\sqrt{\frac{x_1^2 + \cdots + x_n^2}{n}}.$$

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

$$\frac{n}{1/x_1 + \cdots + 1/x_n} \;\leq\; \sqrt[n]{x_1 \cdots x_n} \;\leq\; \frac{x_1 + \cdots + x_n}{n} \;\leq\; \sqrt{\frac{x_1^2 + \cdots + x_n^2}{n}}.$$

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

$$\lim_{p\to 0}\left(\frac{x_1^p+\cdots+x_n^p}{n}\right)^{1/p} \;=\; \sqrt[n]{x_1\cdots x_n}$$

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 ∞ and $-\infty$. It turns out the the ∞-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 7 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

$$p\le q \qquad\Rightarrow\qquad \left(\lambda_1 x_1^p+\cdots+\lambda_n x_n^p\right)^{1/p} \;\le\; \left(\lambda_1 x_1^q+\cdots+\lambda_n x_n^q\right)^{1/q}$$

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

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

$$\left(\lambda_1 x_1+\cdots+\lambda_n x_n\right)^q \;\le\; \lambda_1 x_1^q+\cdots+\lambda_n x_n^q$$

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

$$\left(\lambda_1(x_1^p)+\cdots+\lambda_n(x_n^p)\right)^{q/p} \;\le\; \lambda_1(x_1^p)^{q/p}+\cdots+\lambda_n(x_n^p)^{q/p}$$

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

$$p \leq q \qquad \Rightarrow \qquad \left(\int_X f^p \, d\mu\right)^{1/p} \leq \left(\int_X f^q \, d\mu\right)^{1/q}$$

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

Norms

A **norm** is a function that measures the lengths of vectors in a vector space. The most familiar norm is the **Euclidean norm** on $\mathbb{R}^n$, which is defined by the formula

$$\|(x_1,\ldots,x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Definition: Norm on a Vector Space
Let V be a vector space over $\mathbb{R}$. A **norm** on V is a function $\|-\|\colon V \to \mathbb{R}$, denoted $\mathbf{v} \mapsto \|\mathbf{v}\|$, with the following properties:

1. $\|\mathbf{v}\| \geq 0$ for all $\mathbf{v} \in V$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
2. $\|\lambda\mathbf{v}\| = |\lambda|\,\|\mathbf{v}\|$ for all $\lambda \in \mathbb{R}$ and $\mathbf{v} \in V$.
3. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all $\mathbf{v}, \mathbf{w} \in V$.

For those familiar with topology, any norm $\|-\|$ on a vector space gives a metric d on the vector space defined by the formula

$$d(\mathbf{v},\mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|.$$

Thus any vector space with a norm can be thought of as a topological space.

The Euclidean norm on $\mathbb{R}^n$ can be generalized to the family of p-norms. For any $p \geq 1$, the ***p*-norm** on $\mathbb{R}^n$ is defined by the formula

$$\|(x_1,\ldots,x_n)\|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}$$

The usual Euclidean norm corresponds to the case where $p = 2$.

It is easy to see that the definition of the p-norm satisfies axioms (1) and (2) for a norm, but third axiom (which is known as the **triangle inequality**) is far from clear. The following theorem establishes the p-norm is in fact a norm.

Theorem 8 Minkowski's Inequality

> *If* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *and* $p \in [1, \infty)$*, then*
> $$\|\mathbf{u}+\mathbf{v}\|_p \;\leq\; \|\mathbf{u}\|_p \,+\, \|\mathbf{v}\|_p$$

PROOF Since $p \geq 1$, the function $x \mapsto |x|^p$ is convex. It follows that

$$\begin{aligned}\|(1-\lambda)\mathbf{u}+\lambda\mathbf{v}\|_p^p \;&=\; \sum_{i=1}^n |(1-\lambda)u_i + \lambda v_i|^p \\ &\leq\; \sum_{i=1}^n (1-\lambda)|u_i|^p + \lambda|v_i|^p \;=\; (1-\lambda)\,\|\mathbf{u}\|_p^p \,+\, \lambda\,\|\mathbf{v}\|_p^p\end{aligned}$$

for all $\mathbf{u}$ and $\mathbf{v}$ and $\lambda \in [0,1]$. In particular, this proves that $\|(1-\lambda)\mathbf{u}+\lambda\mathbf{v}\|_p \leq 1$ whenever $\|\mathbf{u}\|_p = \|\mathbf{v}\|_p = 1$.

From this Minkowski's inequality follows. In particular, we may assume that $\mathbf{u}$ and $\mathbf{v}$ are nonzero. Then $\mathbf{u}/\|\mathbf{u}\|_p$ and $\mathbf{v}/\|\mathbf{v}\|_p$ are unit vectors, so

$$\frac{\|\mathbf{u}+\mathbf{v}\|_p}{\|\mathbf{u}\|_p+\|\mathbf{v}\|_p} \;=\; \left\|\frac{\|\mathbf{u}\|_p}{\|\mathbf{u}\|_p+\|\mathbf{v}\|_p}\,\frac{\mathbf{u}}{\|\mathbf{u}\|_p} \;+\; \frac{\|\mathbf{v}\|_p}{\|\mathbf{u}\|_p+\|\mathbf{v}\|_p}\,\frac{\mathbf{v}}{\|\mathbf{v}\|_p}\right\|_p \;\leq\; 1. \qquad \blacksquare$$

If $\|-\|$ is a norm on a vector space V, the **unit ball** in V is the set

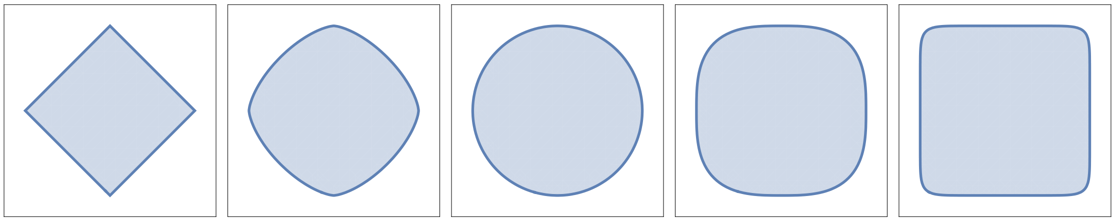
$$B_V(1) \;=\; \big\{\mathbf{v} \in V \;\big|\; \|\mathbf{v}\| \leq 1\big\}.$$

For example, the unit ball in $\mathbb{R}^2$ with respect to the Euclidean norm is a round disk of radius 2 centered at the origin.

Figure 5 shows the unit ball in $\mathbb{R}^2$ with respect to various p-norms. Their shapes are all fairly similar, with the the unit ball being a diagonal square when $p = 1$, and the unit ball approaching a horizontal square as $p \to \infty$. All of these unit balls have a certain important geometric property.

> **Definition: Convex Set**
> Let V is a vector space over $\mathbb{R}$. If $\mathbf{v}$ and $\mathbf{w}$ are points in V, the **line segment** from $\mathbf{v}$ to $\mathbf{w}$ is the set
> $$L(\mathbf{v},\mathbf{w}) \;=\; \{\lambda\mathbf{v} + (1-\lambda)\mathbf{w} \mid 0 \leq \lambda \leq 1\}.$$
> A subset $S \subseteq V$ is **convex** if $L(\mathbf{v},\mathbf{w}) \subseteq S$ for all $\mathbf{v},\mathbf{w} \in S$.

The following theorem gives a geometric interpretation of Minkowski's inequality.



$p = 1$ $p = 1.5$ $p = 2$ $p = 3$ $p = 10$

Figure 5: The shape of the unit ball in $\mathbb{R}^2$ for various p-norms.

Theorem 9 Shapes of Unit Balls

Let V be a vector space over $\mathbb{R}$, and let $\|-\| \colon V \to \mathbb{R}$ be a function satisfying conditions (1) and (2) for a norm. Then $\|-\|$ satisfies the triangle inequality if and only if the unit ball

$$\{\mathbf{v} \in V \mid \|\mathbf{v}\| \leq 1\}$$

is convex.

PROOF Suppose first that $\|-\|$ satisfies the triangle inequality. Let $\mathbf{v}$ and $\mathbf{w}$ be points in the unit ball, and let $\mathbf{p} = \lambda\mathbf{v} + (1-\lambda)\mathbf{w}$ be any point on the line segment from $\mathbf{v}$ to $\mathbf{w}$. Then

$$\begin{aligned} \|\mathbf{p}\| &= \|\lambda\mathbf{v} + (1-\lambda)\mathbf{w}\| \leq \|\lambda\mathbf{v}\| + \|(1-\lambda)\mathbf{w}\| \\ &= \lambda\|\mathbf{v}\| + (1-\lambda)\|\mathbf{w}\| \leq \lambda(1) + (1-\lambda)(1) = 1 \end{aligned}$$

so $\mathbf{p}$ lies in the unit ball as well.

For the converse, suppose that the unit ball is convex, and let $\mathbf{v}, \mathbf{w} \in V$. If $\mathbf{v}$ or $\mathbf{w}$ is $\mathbf{0}$, then clearly $\|\mathbf{v} + \mathbf{w}\| \leq \|v\| + \|\mathbf{w}\|$, so suppose they are both nonzero. Let $\hat{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|$ and $\hat{\mathbf{w}} = \mathbf{w}/\|\mathbf{w}\|$. Then

$$\|\hat{\mathbf{v}}\| = \left\|\frac{1}{\|\mathbf{v}\|}\mathbf{v}\right\| = \frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\| = 1$$

and similarly $\|\hat{\mathbf{w}}\| = 1$. Thus $\hat{\mathbf{v}}$ and $\hat{\mathbf{w}}$ both lie in the unit ball. Since

$$\frac{\|\mathbf{v}\|}{\|\mathbf{v}\| + \|\mathbf{w}\|} + \frac{\|\mathbf{w}\|}{\|\mathbf{v}\| + \|\mathbf{w}\|} = 1,$$

the point

$$\frac{\mathbf{v} + \mathbf{w}}{\|\mathbf{v}\| + \|\mathbf{w}\|} = \frac{\|\mathbf{v}\|}{\|\mathbf{v}\| + \|\mathbf{w}\|}\hat{\mathbf{v}} + \frac{\|\mathbf{w}\|}{\|\mathbf{v}\| + \|\mathbf{w}\|}\hat{\mathbf{w}}$$

must lie in the unit ball as well, and it follows that $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$. ■

Finally, we should mention that there is an integral version of Minkowski's inequality. This involves the notion of a p-norm on a measure space.

Definition: p-Norm
Let (X, μ) be a measure space, let f be a measurable function on X, and let $p \in [1, \infty)$. The ***p*-norm** of f is the quantity

$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p}$$

Note that $\|f\|_p$ is always defined, since it involves the integral of a non-negative function, but it may be infinite. We can now state Minkowski's inequality for arbitrary measure spaces.

Theorem 10 Minkowski's Inequality (Integral Version)

Let (X, μ) be a measure space, let $f, g \colon X \to \mathbb{R}$ be measurable functions, and let $p \in [1, \infty)$. Then

$$\|f + g\|_p \;\leq\; \|f\|_p + \|g\|_p.$$

PROOF Since $p \geq 1$, the function $x \mapsto |x|^p$ is convex. It follows that

$$\begin{aligned} \|(1-\lambda)f + \lambda g\|_p^p &= \int_X |(1-\lambda)f + \lambda g|^p \, d\mu \\ &\leq \int_X \big((1-\lambda)|f|^p + \lambda |g|^p\big) \, d\mu \;=\; (1-\lambda)\, \|f\|_p^p \,+\, \lambda \, \|g\|_p^p \end{aligned}$$

for any measurable functions f and g and any $\lambda \in [0, 1]$. In particular, this proves that $\|(1-\lambda)f + \lambda g\|_p \leq 1$ whenever $\|f\|_p = \|g\|_p = 1$.

From this Minkowski's inequality follows. First, observe that if $\|f\|_p = 0$, then $f = 0$ almost everywhere, so Minkowski's inequality follows in this case. A similar argument holds if $\|g\|_p = 0$, so suppose that $\|f\|_p > 0$ and $\|g\|_p > 0$. Let $\hat{f} = f/\|f\|_p$ and $\hat{g} = g/\|g\|_p$, and note that $\|\hat{f}\|_p = \|\hat{g}\|_p = 1$. Then

$$\frac{\|f+g\|_p}{\|f\|_p + \|g\|_p} \;=\; \left\| \frac{\|f\|_p}{\|f\|_p + \|g\|_p}\, \hat{f} \;+\; \frac{\|g\|_p}{\|f\|_p + \|g\|_p}\, \hat{g} \right\|_p \;\leq\; 1. \qquad \blacksquare$$

Hölder's Inequality

Recall that the Euclidean norm on R^n satisfies the **Cauchy-Schwarz Inequality**

$$|\mathbf{u}\cdot\mathbf{v}| \;\leq\; \|\mathbf{u}\|_2\,\|\mathbf{v}\|_2$$

for all $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$. Our next task is to prove a generalization of this known as Hölder's inequality.

Lemma 11 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 12 Hölder's Inequality

Let $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, and let $p,q\in(1,\infty)$ so that $1/p+1/q=1$. Then

$$|\mathbf{u}\cdot\mathbf{v}| \;\leq\; \|\mathbf{u}\|_p\,\|\mathbf{v}\|_q.$$

PROOF By Young's inequality,

$$\begin{aligned}|\mathbf{u}\cdot\mathbf{v}| \;&\leq\; |u_1v_1| + \cdots + |u_nv_n| \\ &\leq\; \frac{|u_1|^p+\cdots+|u_n|^p}{p} + \frac{|v_1|^q+\cdots+|v_n|^q}{q} \;=\; \frac{\|\mathbf{u}\|_p^p}{p} + \frac{\|\mathbf{v}\|_q^q}{q}\end{aligned}$$

In particular, if $\|\mathbf{u}\|_p = \|\mathbf{v}\|_q = 1$, then $|\mathbf{u}\cdot\mathbf{v}| \leq 1/p + 1/q = 1$, which proves Hölder's inequality in this case.

For the general case, we may assume that $\mathbf{u}$ and $\mathbf{v}$ are nonzero. Then $\mathbf{u}/\|\mathbf{u}\|_p$ and $\mathbf{v}/\|\mathbf{v}\|_q$ are unit vectors for their respective norms, and therefore

$$\left|\frac{\mathbf{u}}{\|\mathbf{u}\|_p}\cdot\frac{\mathbf{v}}{\|\mathbf{v}\|_q}\right| \;\leq\; 1.$$

Multiplying through by $\|\mathbf{u}\|_p \, \|\mathbf{v}\|_q$ gives the desired result. ■

This inequality is not hard to generalize to integrals. We begin with the following definition.

Definition: Inner Product of Functions
Let (X,μ) be a measure space, and let f and g be measurable functions on X. The $\boldsymbol{L^2}$ **inner product** of f and g is defined as follows:

$$\langle f, g\rangle \;=\; \int_X f\,g\,d\mu.$$

Note that $\langle f, g\rangle$ may be undefined if fg is not Lebesgue integrable on X. Note also that $\langle f, g\rangle = \|fg\|_1$ if f and g are nonnegative, but that in general $\langle f, g\rangle \leq \|fg\|_1$.

Theorem 13 Hölder's Inequality (Integral Version)

Let (X,μ) be a measure space, let f and g be measurable functions on X, and let $p,q \in (1,\infty)$ so that $1/p + 1/q = 1$. If $\|f\|_p < \infty$ and $\|g\|_p < \infty$, then $\langle f, g\rangle$ is defined, and

$$|\langle f, g\rangle| \;\leq\; \|f\|_p \, \|g\|_q$$

PROOF Suppose first that $\|f\|_p = \|g\|_q = 1$. By Young's inequality,

$$\|fg\|_1 \;=\; \int_X |fg|\,d\mu \;\leq\; \int_X \left(\frac{|f|^p}{p} + \frac{|g|^q}{q}\right) d\mu \;\leq\; \frac{\|f\|_p^p}{p} + \frac{\|g\|_q^q}{q} \;=\; 1.$$

Since fg is L^1, it follows that $\langle f, g\rangle$ is defined. Then $\langle f, g\rangle \leq \|fg\|_1 \leq 1$, which proves Hölder's inequality in this case.

For the general case, note first that if either $\|f\|_p = 0$ or $\|g\|_q = 0$, then either $f = 0$ almost everywhere or $g = 0$ almost everywhere, so Hölder's inequality holds in that case. Otherwise, let $\hat{f} = f/\|f\|_p$ and $\hat{g} = g/\|g\|_q$. Then $\|\hat{f}\|_p = \|\hat{g}\|_q = 1$, so

$$|\langle f, g\rangle| \;=\; \|f\|_p \, \|g\|_p \, |\langle \hat{f}, \hat{g}\rangle| \;\leq\; \|f\|_p \, \|g\|_q$$ ■

In the case where $p = q = 2$, this gives the following.

Corollary 14 Cauchy-Schwarz Inequality (Integral Version)

Let (X, μ) be a measure space, and let f and g be measurable functions on X. If $\|f\|_2 < \infty$ and $\|g\|_2 < \infty$, then $\langle f, g\rangle$ is defined, and

$$|\langle f, g\rangle| \;\leq\; \|f\|_2 \, \|g\|_2.$$

Exercises

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

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

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

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

4. If $x, y \geq 0$, prove that

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

What is $\displaystyle\lim_{p \to -\infty} \left(\frac{x^p + y^p}{2}\right)^{1/p}$?

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

$$\frac{\sqrt{x_1 y_1} + \cdots \sqrt{x_n y_n}}{n} \;\leq\; \sqrt{\frac{x_1 + \cdots + x_n}{n}}\sqrt{\frac{y_1 + \cdots + y_n}{n}}.$$

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 $f\colon [0,1] \to \mathbb{R}$ be a bounded measurable function, and define $\varphi\colon [1,\infty) \to \mathbb{R}$ by

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

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

8. Prove that

$$\bigl(1 + x^2 y + x^4 y^2\bigr)^3 \;\leq\; \bigl(1 + x^3 + x^6\bigr)^2 \bigl(1 + y^3 + y^6\bigr)$$

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

9. Let (X,μ) be a measure space, let $f, g\colon X \to [0,\infty)$ be measurable functions, and let $p, q, r \in (1,\infty)$ so that $1/p + 1/q = 1/r$. Prove that $\|fg\|_r \leq \|f\|_p \, \|g\|_q$.