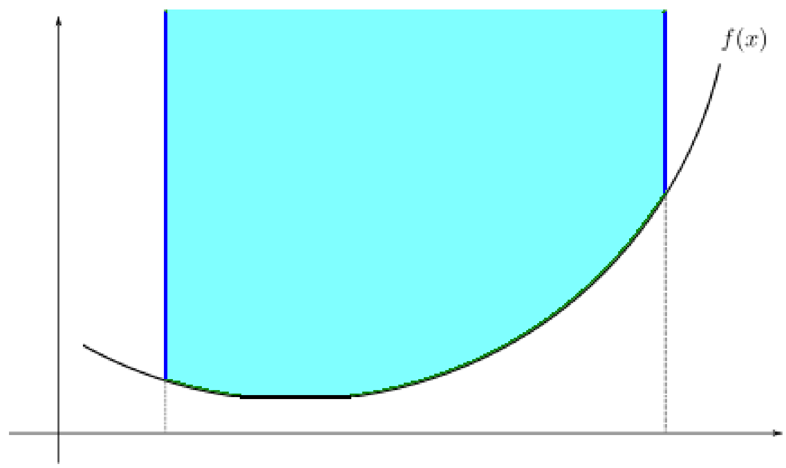
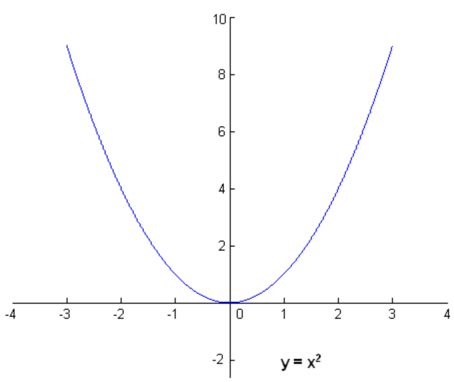
Convexity and differentiable functions

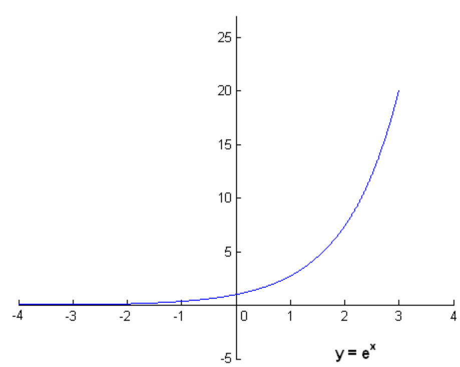
We know that half – planes in $\mathbb{R}^2$ and half – spaces in $\mathbb{R}^3$ are fundamental examples of convex sets. Many of these examples are defined by inequalities of the form $y \geq f(x_1, x_2, \dots, x_k)$ where f is a first degree polynomial in the coordinates x_j and $k = 1$ or 2 depending upon whether we are looking at $\mathbb{R}^2$ or $\mathbb{R}^3$. Our objective here is to derive a simple criterion for recognizing other convex sets defined by inequalities of the form $y \geq f(x_1, x_2, \dots, x_k)$, where f is a function with continuous second partial derivatives; the set defined in this manner is sometimes called the ***epigraph*** of the real valued function f.



(***Source:*** **http://en.wikipedia.org/wiki/File:Epigraph_convex.svg**)

Many functions in elementary calculus have convex epigraphs. In particular, if we take $f(x) = x^2$ or e^x, then inspection of the graphs strongly suggests that their epigraphs are convex; our recognition criterion will prove these statements (and also yield many others of the same type).





(***Source:*** **http://algebra.freehomeworkmathhelp.com/Relations_and_Functions/Graphs/Graphs_of_Algebra_Functions/graphs_of_algebra_functions.html**)

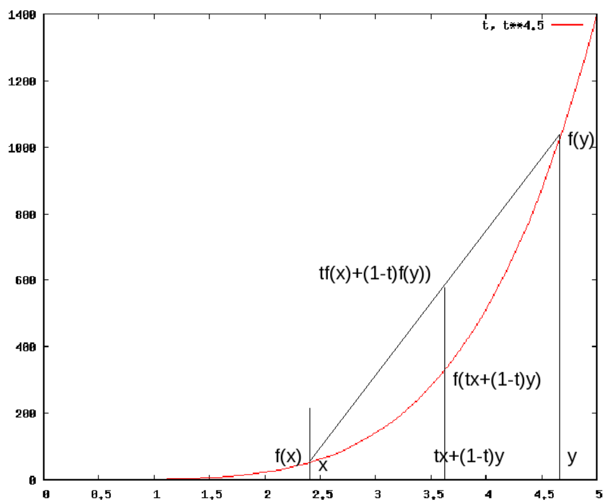
Nearly all of our discussion generalizes to a class of examples known as convex functions, but our discussion will be limited because we mainly interested in finding examples. Here are some references for the more general results about convex functions:

We shall begin by considering convex functions of one variable, and afterwards we shall explain how everything can be extended to functions of two (or more) variables.

Definition. If $\mathbf{K}$ is a convex subset of $\mathbb{R}^n$ and f is a real valued function on $\mathbf{K}$, then f is said to be ***convex*** if for each $\mathbf{x}, \mathbf{y}$ in $\mathbf{K}$ and each t in the open interval $(0, 1)$ we have

$$f\big(t\mathbf{x} + (1-t)\mathbf{y}\big) \leq t f(\mathbf{x}) + (1-t) f(\mathbf{y}).$$

In calculus textbooks (*but practically nowhere else*!) such functions are often said to be ***concave upward.***



(***Source:*** http://withfriendship.com/user/levis/convex-function.php)

Convex functions have been studied extensively in both theoretical and applied mathematics. Further information can be found in the following online article:

http://en.wikipedia.org/wiki/Convex_function

Our first result gives an alternate characterization of convexity for functions.

Theorem 1. *A real valued function on a convex set* $\mathbf{K}$ *in* $\mathbb{R}^n$ *is a convex function if and only it its epigraph in* $\mathbf{K} \times \mathbb{R}$ *is a convex set* (*we view the latter as a subspace of* $\mathbb{R}^{n+1}$ *in the usual way*).

Proof of Theorem 1. $(\Leftarrow)$ Suppose that $\mathbf{x}, \mathbf{y} \in K$ and $t \in (0,1)$. Since the epigraph E of f is convex and it contains the graph of f, it follows that

$$(t\,\mathbf{x} \;+\; (1-t)\,\mathbf{y},\; t\,f(\mathbf{x}) \;+\; (1-t)\,f(\mathbf{y})\,) \quad = \quad t\;(\mathbf{x}, f(\mathbf{x})\,) \;\;+\;\; (1-t)\;(\mathbf{y}, f(\mathbf{y})\,)$$

also lies in E. By definition of the latter, this means that

$$t\,f(\mathbf{x}) \;+\; (1-t)\,f(\mathbf{y}) \quad \geq \quad f\,(t\,\mathbf{x} \;+\; (1-t)\,\mathbf{y}\,)$$

and therefore f is a convex function.

$(\Rightarrow)$ Suppose that $(\mathbf{x}, u)$ and $(\mathbf{y}, v)$ lie in the epigraph, so that $u \geq f(\mathbf{x})$ and $v \geq f(\mathbf{y})$; we need to prove that

$$t\,u \;+\; (1-t)\,v \quad \geq \quad f\,(t\,\mathbf{x} \;+\; (1-t)\,\mathbf{y}\,)$$

for all $t \in (0,1)$. The hypotheses imply that the left hand side satisfies

$$t\,u \;+\; (1-t)\,v \quad \geq \quad t\,f(\mathbf{x}) \;+\; (1-t)\,f(\mathbf{y})$$

and the convexity of f shows that the right hand side is greater than or equal to $f\,(t\,\mathbf{x} \;+\; (1-t)\,\mathbf{y}\,)$. Combining these, we obtain the inequality in first sentence of the paragraph.■

The Second Derivative Test for Convexity

We shall now state the main result; versions of it are implicit in the discussions of curve sketching that appear in standard calculus texts.

Theorem 2. *Let $K \subset \mathbb{R}$ be an interval, and let f be a real valued function on K with a continuous second derivative. If f'' is nonnegative everywhere, then f is convex on K.*

The next result contains the main steps in the proof of Theorem 2.

Lemma 3. *Let f be a real valued function on the closed interval $[a,b]$ with a second continuous derivative. Suppose further that f'' is nonnegative on $[a,b]$ let g be the linear function with $f(a) = g(a)$ and $f(b) = g(b)$. Then $f(x) \leq g(x)$ for all $x \in [a,b]$.*

Proof of Lemma 3. Define a new function h on $[a,b]$ by $h = g - f$. Then by construction we have $h(a) = h(b) = 0$ and $h''(x) = g''(x) - f''(x) = 0 \;-\; f''(x)$ because the second derivative of the linear function g is zero; since $f'' \geq 0$ it follows that $h'' \leq 0$. The preceding observations then yield the following properties of the function h:

(i) For some $C \in (a,b)$ we have $h'(C) = 0$.

(ii) The derivative h' is nondecreasing.

The first of these is a consequence of Rolle's Theorem, and the second follows because the derivative h'' of h' is nonpositive. If we combine (i) and (ii) we see that $h'(x) \geq 0$ for $x \leq C$ and $h'(x) \leq 0$ for $x \geq C$.

In order to prove Lemma 3, we must show that $h(x) \geq 0$ everywhere. Since $h(a) = 0$ and $h' \geq 0$ for $x \leq C$, it follows that $h(x) \geq 0$ for $x \leq C$. We need to prove the same conclusion for $x \geq C$; assume this is false, and assume specifically that $h(D) < 0$ for some $D \in (C,b)$. Since $h' \leq 0$ for $x \geq D$, we must also have $h(x) < 0$ for all $x \in (D,b]$. In particular, this implies that $h(b) < 0$, which is a contradiction because we know that $h(b) = 0$. The source of this contradiction is our supposition that $h(D) < 0$ for some D, and thus we must have $h \geq 0$ everywhere.■

Proof of Theorem 2. Let $a, b \in K$; our objective is to prove that if $t \in (0,1)$ then

$$f\,((1-t)\,a \;+\; t\,b) \quad \leq \quad (1-t)\,f(a) \;+\; t\,f(b)\ .$$

We claim that, without loss of generality, we may assume a is less than b; this is true because for each $t \in (0,1)$ we may rewrite $(1-t)\,a \;+\; t\,b$ as $(1-s)\,a \;+\; s\,b$ where $s = 1-t$ also lies in $(0,1)$.

Assuming $a < b$. take g to be the linear function defined in Lemma 3, let $t \in (0,1)$ and set x equal to $(1-t)\,a \;+\; t\,b$; since $x \in (a,b)$ there is a unique solution t to this equation given by

$$t \quad = \quad \frac{x-a}{b-a}$$

and this solution lies in $(0,1)$. We can now apply Lemma 3 to conclude that $f(x) \leq g(x)$. Since g is a linear function we have

$$g(x) \quad = \quad g\,(\,(1-t)\,a \;+\; t\,b) \quad = \quad (1-t)\,g(a) \;+\; t\,g(b) \quad = \quad (1-t)\,f(a) \;+\; t\,f(b)$$

and by the preceding sentence we know this is greater than or equal to $f\,(\,(1-t)\,a \;+\; t\,b)$. Therefore f is a convex function.■

EXAMPLES. Theorem 2 implies that both $f(x) = x^2$ and $f(x) = e^x$ are convex because their second derivatives are the positive valued functions 2 (the constant function) and e^x respectively. Similarly, $f(x) = 1/x$ is convex on the open half-line defined by $x > 0$ because $f''(x) = 2/x^3$ is positive for $x > 0$.

Generalization to higher dimensions

Although a few complications arise, we can prove a corresponding Second Derivative Test for recognizing convex functions of finitely many (say n) variables. The first of these is standard in multivariable differential calculus; namely, we must restrict attention to **open** convex subsets of $\mathbb{R}^n$ if $n \geq 2$. Likewise, in analogy with the second derivative tests for relative maxima and minima, we need to consider certain algebraic properties of the Hessian matrix

$$H(f) \quad = \quad \left(\frac{\partial^2 f}{\partial x_j\,\partial x_j}\right)$$

which is symmetric (mixed partials do not depend upon the order in which the partial derivatives are taken).

Algebraic digression. If $A = (a_{i,j})$ is a symmetric $n \times n$ matrix, then A is said to be **positive definite** if for each nonzero vector $\mathbf{v} = (v_1,\ \cdots\ ,v_n)$ we have

$$\sum_{i,j}\ a_{i,j}\,v_i\,v_j \quad > \quad 0\ .$$

The standard test for recognizing such matrices is the **principal minors test**:

> *Given a symmetric matrix A as above, let A_k be the $k \times k$ submatrix generated by the first k rows and columns of A. Then A is positive definite if and only if* $\det A_k > 0$ *for* $k = 1,\ \ldots\ ,n$.

See pages 84 and 88–90 of the document

`http://math.ucr.edu/≃res/math132/linalgnotes.pdf`

for further information, including a proof of this fact.

The relevance of positive definite matrices arises from the following observation.

Lemma 4. *Let U be a convex open subset of $\mathbb{R}^n$, let f be a real valued function with continuous second partial derivatives, let $\mathbf{x}$ and $\mathbf{y}$ be distinct points of U, and write $\mathbf{v} = \mathbf{y} - \mathbf{x}$ (hence $\mathbf{v}$ is nonzero. If $\varphi(t) = f(x + t\,\mathbf{v})$ for t in some open interval containing $[0, 1]$, then*

$$\varphi''(t) \quad = \quad \sum_{i,j} \frac{\partial^2 f}{\partial x_j\,\partial x_j}(x + t\,\mathbf{v})\,v_i\,v_j\ .$$

NOTE. The convexity and openness of U imply that $\varphi(t)$ can be always be defined for all t in some open interval containing $[0, 1]$.

Proof of Lemma 4. This follows immediately from successive applications of the Chain Rule to $\varphi(t)$ and $\varphi'(t)$.■

Theorem 5. (Multivariable Second Derivative Test for Convexity) *Let $K \subset \mathbb{R}^n$ be an open convex set, and let f be a real valued function on K with continuous second partial derivatives. If the Hessian of f is positive definite everywhere, then f is convex on K.*

Proof. Let $\mathbf{x}$ and $\mathbf{y}$ be distinct points of K, let $t \in (0, 1)$, and let $\varphi(u)$ be defined as in Lemma 4. Since the Hessian of f is positive definite everywhere, Lemma 4 implies that $\varphi''(u) > 0$ for all u and hence Theorem 2 shows that φ is a convex function on the open interval containing $[0, 1]$. Using the identity $t = t \cdot 1 + (1 - t) \cdot 0$, we then have

$$f\,(t\,\mathbf{y}\ +\ (1-t)\,\mathbf{x}\,) \quad = \quad \varphi(t) \quad = \quad \varphi\,(t\cdot 1 + (1-t)\cdot 0) \quad \leq \quad t\varphi(1)\ +\ (1-t)\varphi(0)\ .$$

By construction $\varphi(0) = f(\mathbf{x})$ and $\varphi(1) = f(\mathbf{y})$, so the convexity of f follows from substitution of these values into the right hand side of the display above.■

If we specialize to the case $n = 2$ the Principal Minors Test for the Hessian of f reduces to the pair of inequalities

$$\frac{\partial^2 f}{\partial x_1^2} \quad > \quad 0, \qquad \det\left(\frac{\partial^2 f}{\partial x_j\,\partial x_j}\right) \quad > \quad 0$$

and computations for specific examples are often very easy.

EXAMPLES. If $K = \mathbb{R}^2$, then the functions $f(x, y) = x^2 + y^2$ and $f(x, y) = e^x + e^y$ are convex on K by Theorem 5 and the preceding simplification of the Principal Minors Test. Similar considerations show that if K is the open first quadrant defined by $x > 0$ and $y > 0$, then $f(x, y) = 1/xy$ is convex on K.

Exercise

Show that if K is an open convex set and f is a convex function on K then the *open epigraph* consisting of all $(\mathbf{x}, u) \in K \times \mathbb{R}$ such that $u > f(\mathbf{x})$ (*i.e.*, we have strict inequality) is also a convex set. [*Hint:* Imitate the relevant portion of the proof for Theorem 1.]