Convexity Examples

CE 377K
Stephen D. Boyles
Spring 2015

1 Convex Functions

In general, it is difficult to find the optimum value with a nonlinear objective function. For instance, the function in Figure 1 has many local minima and is unbounded below as $x \to -\infty$, both of which can cause serious problems if we're trying to minimize this function. Usually, the best that a software program can do is find a local minimum. If it finds one of the local minima for this function, it may not know if there is a better one somewhere else (or if there is, how to find it). Or if it starts seeking x values which are negative, we could run into the unbounded part of this function.

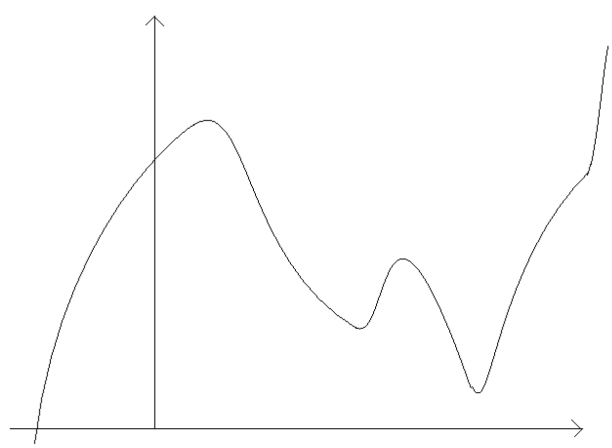

Figure 1: A function which is not convex.

On the other hand, some functions are very easy to minimize. The function in Figure 2 only has one minimum point, is not unbounded below, and as we will see in the coming weeks, there are many algorithms which can find that minimum point efficiently.

What distinguishes these is a property called *convexity*. Confusingly, the word convex can refer to both functions and sets, and it has two distinct meanings. This section focuses on convex *functions*, while the next section focuses on convex *sets*. They are similar, however, in that convex functions and convex sets are extremely desirable. If the feasible region is a convex set, and if the objective function is a convex function, then it is much easier to find the optimal solution.

Geometrically, a convex function lies below its secant lines. Remember that a secant line is the line segment joining two points on the function. As we see in Figure 3, no matter what two points we

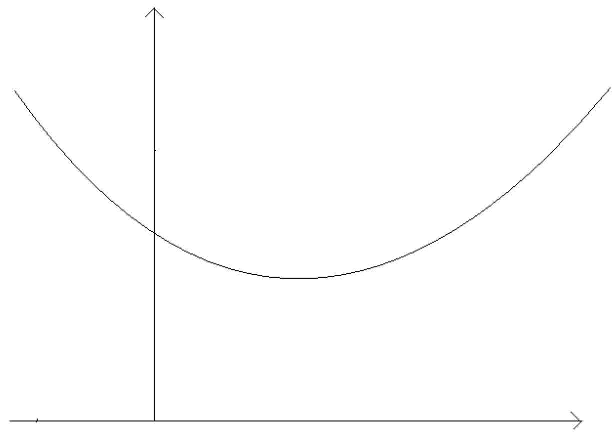

Figure 2: A convex function.

pick, the function always lies below its secant line. On the other hand, in Figure 4, not every secant line lies above the function: some lie below it, and some lie both above and below it. Even though we can draw some secant lines which are above the function, this isn't enough: *every* possible secant must lie below the function.

The following definition makes this intuitive notion formal, using mathematical language:

Definition 1. *A function* $f : X \to \mathbb{R}$ *is* convex *if, for every* $x_1, x_2 \in X$ *and every* $\lambda \in [0, 1]$,

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2) \tag{1}$$

and strictly convex *if*

$$f((1-\lambda)x_1 + \lambda x_2) < (1-\lambda)f(x_1) + \lambda f(x_2) \tag{2}$$

for all distinct $x_1, x_2 \in X, \lambda \in (0, 1)$

Essentially, x_1 and x_2 are the two endpoints for the secant line. Since this entire line segment must be above the function, we need to consider every point between x_1 and x_2. This is what λ does: as λ varies between 0 and 1, the points $\lambda x_2 + (1-\lambda)x_1$ cover every point between x_1 and x_2. You can think of $\lambda x_2 + (1-\lambda)x_1$ as a "weighted average," where λ is the weight put on x_2. For $\lambda = 0$, all the weight is on x_1. For $\lambda = 1$, all the weight is on x_2. For $\lambda = 1/2$, equal weight is put on the two points, so the weighted average is the midpoint. $\lambda = 1/3$ corresponds to the point a third of the way between x_1 and x_2.

So we need to say that, at all such intermediate x values, the value of f is lower than the y-coordinate of the secant. The value of the function at this point is simply $f((1-\lambda)x_1 + \lambda x_2)$. Because the secant is a straight line, its y-coordinate can be seen as a weighted average of the y-coordinates of its endpoints, that is, $f(x_1)$ and $f(x_2)$. This weighted average can be written as $(1-\lambda)f(x_1) + \lambda f(x_2)$, so requiring the function to lie below the secant line is exactly the same as enforcing condition (1) for all possible secant lines: that is, for all $x_1, x_2 \in X$ and all $\lambda \in [0, 1]$.

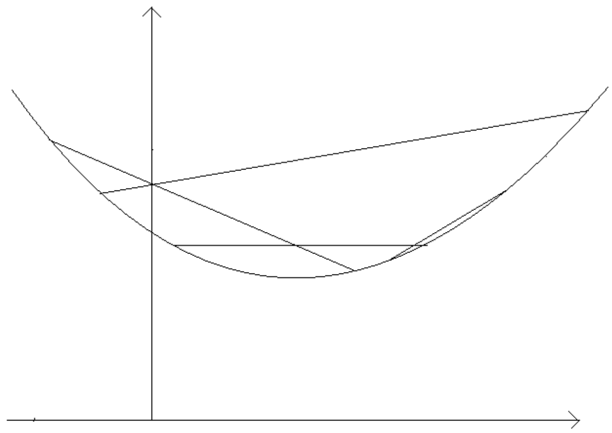

Figure 3: A convex function lies below all of its secants.

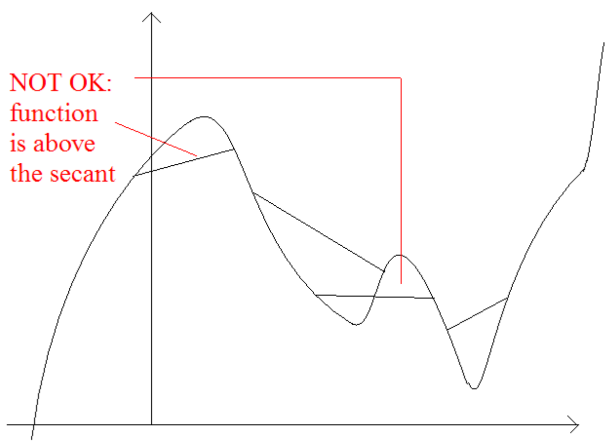

Figure 4: A nonconvex function does not lie below all of its secants.

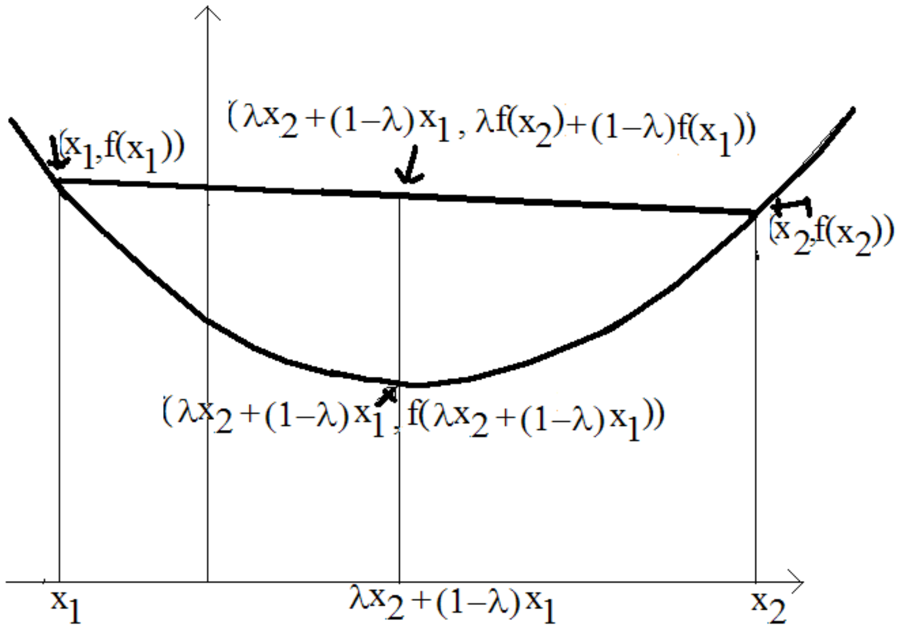

Figure 5: All of the relevant points for the definition of convexity.

Figure 5 explains this in more detail. Along the horizontal axis, the secant endpoints x_1 and x_2 are shown, along with an intermediate point $\lambda x_2 + (1-\lambda)x_1$. The y-coordinates are also shown: at the endpoints, these are $f(x_1)$ and $f(x_2)$. At the intermediate point, the y-coordinate of the function is $f(\lambda x_2 + (1-\lambda)x_1)$, while the y-coordinate of the secant is $\lambda f(x_2) + (1-\lambda)f(x_1)$. Because the function is convex, the former can be no bigger than the latter. Time spent studying this diagram is very well spent. Make sure you understand what each of the four points marked on the diagram represents, and why the given mathematical expressions correctly describe these points. Make sure you see what role λ plays: as λ increses from 0 to 1, the central vertical line moves from x_1 to x_2. (What would happen if we picked x_1 and x_2 such that $x_1 > x_2$? What if $x_1 = x_2$?)

Example 1. *Is the function $f(x) = |x|$, $x \in \mathbb{R}$ convex? Is it strictly convex?*

Solution. To see if f is convex, we need to see if (1) is true; to see if it is strictly convex, we need to check (2). Furthermore, these inequalities have to be true for *every* $x_1, x_2 \in \mathbb{R}$, and *every* $\lambda \in [0, 1]$. It is not enough to simply pick a few values randomly and check the equations. So, we have to work symbolically. In this case,

$$\begin{aligned}
f((1-\lambda)x_1 + \lambda x_2) &= |(1-\lambda)x_1 + \lambda x_2| \\
&\le |(1-\lambda)x_1| + |\lambda x_2| \quad \text{by the triangle inequality} \\
&= (1-\lambda)|x_1| + \lambda |x_2| \quad \text{because } \lambda, 1-\lambda \ge 0 \\
&= (1-\lambda)f(x_1) + \lambda f(x_2)
\end{aligned}$$

Therefore (1) is satisfied, so f is convex. To show that it is strictly convex, we would have to show that the inequality

$$|(1-\lambda)x_1 + \lambda x_2| \leq |(1-\lambda)x_1| + |\lambda x_2|$$

can be replaced by a *strict* inequality $<$. However, we can't do this: for example, if $x_1 = 1$, $x_2 = 2$, $\lambda = 0.5$, the left side of the inequality ($|1/2 + 2/2| = 3/2$) is exactly equal to the right side ($|1/2| + |2/2| = 3/2$). So f is not strictly convex. ■

Note that proving that $f(x)$ is convex requires a *general* argument, where proving that $f(x)$ was not strictly convex only required a single counterexample. This is because the definition of convexity is a "for all" or "for every" type of argument. To prove convexity, you need an argument that allows for all possible values of x_1, x_2, and λ, whereas to disprove it you only need to give one set of values where the necessary condition doesn't hold.

Example 2. *Show that every affine function $f(x) = ax + b$, $x \in \mathbb{R}$ is convex, but not strictly convex.*

Solution.

$$\begin{aligned} f((1-\lambda)x_1 + \lambda x_2) &= a((1-\lambda)x_1 + \lambda x_2) + b \\ &= a((1-\lambda)x_1 + \lambda x_2) + ((1-\lambda) + \lambda)b \\ &= (1-\lambda)(ax_1 + b) + \lambda(ax_2 + b) \\ &= (1-\lambda)f(x_1) + \lambda f(x_2) \end{aligned}$$

So we see that inequality (1) is in fact satisfed as an *equality*. That's fine, so every affine function is convex. However, this means we can't replace the inequality $\leq$ with the strict inequality $<$, so affine functions are not strictly convex. ■

Sometimes it takes a little bit more work, as in the following example:

Example 3. *Show that $f(x) = x^2$, $x \in \mathbb{R}$ is strictly convex.*

Solution. Pick x_1, x_2 so that $x_1 \neq x_2$, and pick $\lambda \in (0, 1)$.

$$\begin{aligned} f((1-\lambda)x_1 + \lambda x_2) &= ((1-\lambda)x_1 + \lambda x_2)^2 \\ &= (1-\lambda)^2 x_1^2 + \lambda^2 x_2^2 + 2(1-\lambda)\lambda x_1 x_2 \end{aligned}$$

What to do from here? Since $x_1 \neq x_2$, $(x_1 - x_2)^2 > 0$. Expanding, this means that $x_1^2 + x_2^2 > 2x_1x_2$. This means that

$$\begin{aligned} (1-\lambda)^2 x_1^2 + \lambda^2 x_2^2 + 2(1-\lambda)\lambda x_1 x_2 &< (1-\lambda)^2 x_1^2 + \lambda^2 x_2^2 + (1-\lambda)(\lambda)(x_1^2 + x_2^2) \\ &= (1 - 2\lambda - \lambda^2 + \lambda + \lambda^2)x_1^2 + (\lambda - \lambda^2 + \lambda^2)x_2^2 \\ &= (1-\lambda)x_1^2 + \lambda x_2^2 \\ &= (1-\lambda)f(x_1) + \lambda f(x_2) \end{aligned}$$

which proves strict convexity. ■

This last example shows that proving convexity can be difficult and nonintuitive, even for simple functions like x^2. The good news is that oftentimes there are simpler conditions that we can check. These conditions involve the first and second derivatives of a function.

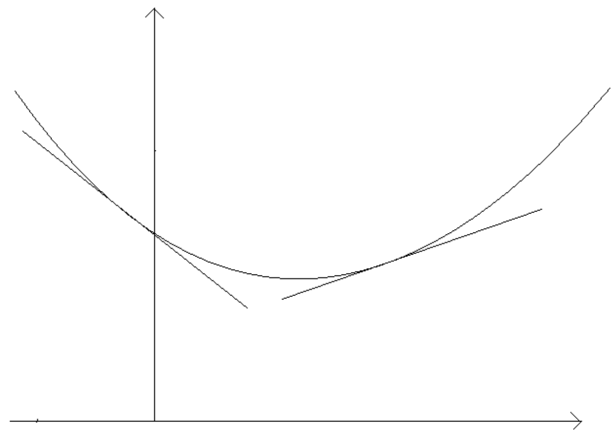

Figure 6: A convex function lies above its tangents.

Proposition 1. *Let $f : X \to \mathbb{R}$ be a function, where $X \subseteq \mathbb{R}$,* [1] *and let f be differentiable on X. Then f is convex if and only if*

$$f(x_2) \geq f(x_1) + f'(x_1)(x_2 - x_1)$$

for all $x_1, x_2 \in X$.

Proposition 2. *Let $f : X \to \mathbb{R}$ be a function, where $X \subseteq \mathbb{R}$, and let f be twice differentiable on X. Then f is convex if and only if $f''(x) \geq 0$ for all $x \in X$*

Equivalent conditions for strict convexity can be obtained in a natural way, changing $\geq$ to $>$ and requiring that x_1 and x_2 be distinct in Proposition 1. If you're interested, I can provide you with detailed proofs of these statements but in my opinion they are not particularly instructive given their length. Essentially, Proposition 1 says that f lies *above* its tangent lines (Figure 6), while Proposition 2 says that f is always "curving upward." (A convex function lies *above* its tangents, but *below* its secants.)

These conditions are usually easier to verify than that of Definition 1.

Example 4. *Show that $f(x) = x^2$ is strictly convex using Proposition 1*

Solution. Pick any $x_1, x_2 \in \mathbb{R}$ with $x_1 \neq x_2$. We have $f'(x_1) = 2x_1$, so we need to show that

$$x_2^2 > x_1^2 + 2x_1(x_2 - x_1)$$

Expanding the right-hand side and rearranging terms, we see this is equivalent to

$$x_1^2 - 2x_1x_2 + x_2^2 > 0$$

[1]This is a fancy mathematical way of saying that f is a function of one variable only

or

$$(x_1 - x_2)^2 > 0$$

which is clearly true since $x_1 \neq x_2$. Thus f is strictly convex. ■

Example 5. *Show that $f(x) = x^2$ is strictly convex using Proposition 2*

Solution. $f''(x) = 2 > 0$ for all $x \in \mathbb{R}$, so f is strictly convex. ■

Much simpler! If f is differentiable (or, better yet, twice differentiable) checking these conditions is almost always easier.

Furthermore, once we know that some functions are convex, we can use this to show that many other combinations of these functions must be convex as well.

Proposition 3. *If f and g are convex functions, and α and β are positive real numbers, then $\alpha f + \beta g$ is convex as well.*

Proposition 4. *If f and g are convex functions, then $f \circ g$ is convex as well.*

Some common convex functions are $|x|$, x^2, e^x, and $ax + b$. So, Proposition 3 tells us that $3x^2 + 4|x|$ is convex. It also tells us that any quadratic function $ax^2 + bx + c$ is convex as long as $a > 0$. Proposition 4 says that the composition of two convex functions is convex as well. For instance, e^{x^2} is convex, and $x^4 = (x^2)^2$ is convex as well.

What about functions of more than one variable? The "shortcut" conditions in Propositions 1 and 2 only apply if the domain of f is one-dimensional. It turns out that very similar conditions can be given for the multivariable case. The multi-dimensional equivalent of the first derivative is the *gradient* ∇f, which is the vector of all partial derivatives:

$$\nabla f \equiv \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]$$

and the equivalent of the second derivative is the *Hessian* $H(f)$, which is the matrix of all second partial derivatives:

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

For example, if $f(x_1, x_2) = x_1 + 3x_1^2 x_2^2$, the gradient is $\nabla f = [1 + 6x_1 x_2^2, 6x_1^2 x_2]$, and the Hessian is

$$H(f) = \begin{bmatrix} 6x_2^2 & 12x_1 x_2 \\ 12x_1 x_2 & 6x_1^2 \end{bmatrix}$$

Note that the Hessian is symmetric if all of the second partial derivatives are continuous, due to Clairaut's theorem.

The equivalent conditions on convexity are

Proposition 5. *Let $f : X \to \mathbb{R}$ be a function whose gradient exists everywhere on X. Then f is convex if and only if*

$$f(x_2) \geq f(x_1) + \nabla f(x_1)(x_2 - x_1)$$

for all $x_1, x_2 \in X$.

Proposition 6. *Let $f : X \to \mathbb{R}$ be a function whose Hessian exists everywhere on X. Then f is convex if and only if $H(f)$ is positive semidefinite*[2] *for all $x \in X$*

Unfortunately, neither of these is as easy to check as the single-dimension equivalents. In particular, it is rather tedious to check whether or not a matrix is positive semidefinite or not. For this reason, in this class I will not ask you to check the convexity of a function of more than one variable.

2 Convex Sets

The previous section discussed, in much detail, what it means for a function to be convex. We also need to talk about what it means for a *set* to be convex, in particular, the feasible set for an optimization problem. The good news is that set convexity is usually much simpler to prove than function convexity.

Let X be any subset of $\mathbb{R}^n$ (that is, X is an n-dimensional set). If X is convex, geometrically this means that line segments connecting points of X lie entirely within X. For example, the set in Figure 7 is convex, while those in Figure 8 and Figure 9 are not. Intuitively, a convex set cannot have any "holes" punched into it, or "bites" taken out of it.

Mathematically, we write this as follows:

Definition 2. *A set $X \subseteq \mathbb{R}^n$ is convex if, for all $x_1, x_2 \in X$ and all $\lambda \in [0, 1]$, the point $\lambda x_2 + (1 - \lambda)x_1 \in X$.*

(There is no such thing as "strict convexity" for sets.)

The definition of set convexity has several similarities with the definition of function convexity. Both involve a condition that has to be satisfied for all pairs of points x_1 and x_2, and for some scalar $\lambda \in [0, 1]$. This happens because both definitions involve line segments: for function convexity, the secant line segment must lie above the function. For set convexity, the line segment between any two points in the set must lie within the set. One of the most convenient ways to express the line segment between any two points is $\lambda x_2 + (1 - \lambda)x_1$, since you will cover the entire line between x_1 and x_2 as λ varies between 0 and 1, regardless of how close or far apart these two points are located.

[2] A symmetric matrix A is positive semidefinite if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$.

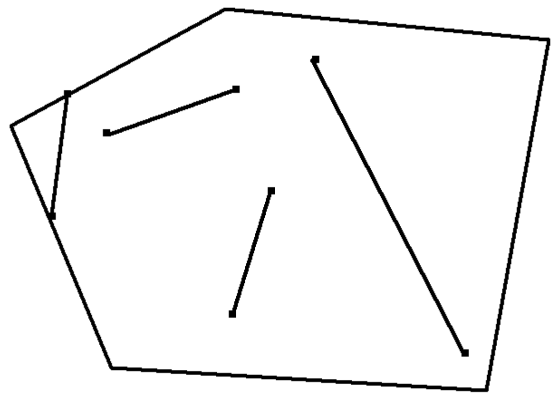

Figure 7: A convex set.

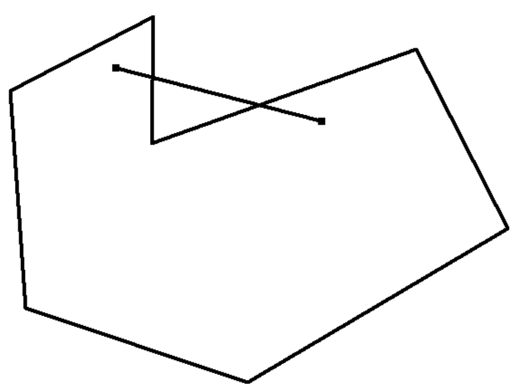

Figure 8: A nonconvex set with a “bite” taken out of it.

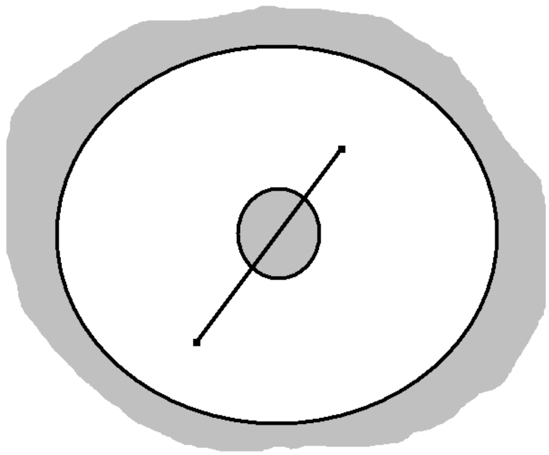

Figure 9: A nonconvex set with a "hole" in it.

Example 6. *Show that the one-dimensional set $X = \{x : x \geq 0\}$ is convex.*

Solution. Pick any $x_1, x_2 \geq 0$ and any $\lambda \in [0,1]$. Because x_1, x_2, and λ are all nonnegative, so are λx_2, $(1-\lambda)x_1$, and therefore so is $\lambda x_2 + (1-\lambda)x_1$. Therefore $\lambda x_2 + (1-\lambda)x_1$ belongs to X as well. ■

Example 7. *Show that the hyperplane $X = \{x : \sum_{i=1}^n a_i x_i - b = 0\}$ is convex.*

Solution. This set is the same as $\{x : \sum_{i=1}^n a_i x_i = b\}$ Pick any $x_1, x_2 \in X$ and any $\lambda \in [0,1]$. Then

$$\begin{aligned}\sum_{i=1}^n a_i(\lambda(x_2)_i + (1-\lambda)(x_1)_i) &= \lambda \sum_{i=1}^n a_i(x_2)_i + (1-\lambda)\sum_{i=1}^n a_i(x_1)_i \\ &= \lambda b + (1-\lambda)b \\ &= b\end{aligned}$$

so $\lambda(x_2)_i + (1-\lambda)(x_1)_i \in X$ as well. ■

Although proving set convexity is usually easier than proving function convexity, sometimes complicated arguments are still needed.

Example 8. *Show that the two-dimensional ball $B = \{[x,y] : x^2 + y^2 \leq 1\}$ is convex.*

Solution. Pick any $b_1, b_2 \in B$ and any $\lambda \in [0,1]$. The point $\lambda b_2 + (1-\lambda)b_1$ is the vector $[\lambda(b_2)_x + (1-\lambda)(b_1)_x, \lambda(b_2)_y + (1-\lambda(b_1)_y)$. To show that it is in B, we must show that the sum of the squares of these components is no greater than 1.

$$\begin{aligned}&(\lambda(b_2)_x + (1-\lambda)(b_1)_x)^2 + (\lambda(b_2)_y + (1-\lambda(b_2)_y))^2 \\ &\quad = \lambda^2((b_2)_x^2 + (b_2)_y^2) + (1-\lambda)^2((b_1)_x^2 + (b_1)_y^2) + 2\lambda(1-\lambda)((b_1)_x(b_2)_x + (b_1)_y(b_2)_y) \\ &\quad \leq \lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda)((b_1)_x(b_2)_x + (b_1)_y(b_2)_y)\end{aligned}$$

because $b_1, b_2 \in B$ (and therefore and $(b_2)_x^2 + (b_2)_y^2 \leq 1$). Notice that $(b_1)_x(b_2)_x + (b_1)_y(b_2)_y$ is simply the dot product of b_1 and b_2, which is equal to $||b_1||||b_2|| \cos\theta$, where θ is the angle between the vectors b_1 and b_2. Since $||b_1|| \leq 1$, $||b_2|| \leq 1$ (by definition of B), and since $\cos\theta \leq 1$ regardless of θ, $(b_1)_x(b_2)_x + (b_1)_y(b_2)_y \leq 1$. Therefore

$$\lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda)((b_1)_x(b_2)_x + (b_1)_y(b_2)_y) \leq \lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda) = (\lambda + (1-\lambda))^2 = 1$$

so the point $\lambda b_2 + (1-\lambda)b_1$ is in B regardless of the values of b_1, b_2, or λ. Thus B is convex. ■

Example 9. *Show that the complement of the ball $B^C = \{[x, y] : x^2 + y^2 > 1\}$ is not convex.*

Solution. Let $b_1 = [2, 0]$, $b_2 = [-2, 0]$, $\lambda = 1/2$. Then $\lambda b_2 + (1-\lambda)b_1 = [0, 0] \notin B^C$ even though $b_1, b_2 \in B^C$ and $\lambda \in [0, 1]$. Therefore B^C is not convex. ■

Again, notice that proving that a set is convex is harder and requires showing that something is true for *all* possible values of $x_1, x_2 \in X$, and $\lambda \in [0, 1]$, whereas disproving convexity only requires you to pick one set of these values where the definition fails.

Even though convex sets are different than convex functions, they are very closely related:

Proposition 7. *If $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function, then the set $X = \{x : g(x) \leq 0$ is a convex set.*

Proposition 8. *If the sets X_1 and X_2 are convex, then the set $X = X_1 \cap X_2$ is convex as well.*

These last two propositions are extremely important for letting us verify whether the feasible region for a nonlinear program is convex.

Theorem 1. *Consider a nonlinear program in standard form. If $g_1(\mathbf{x}), \ldots, g_l(\mathbf{x})$ are convex functions, and if $h_1(\mathbf{x}), \ldots, h_m(\mathbf{x})$ are hyperplanes, then the feasible region $X = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0, i \in \{1, \ldots, l\}, j \in \{1, \ldots, m\}\}$ is convex.*

Proof. Let $Y_i = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x} \leq 0\}$ represent the values of $\mathbf{x}$ which satisfy the i-th inequality constraint, and let $Z_j = \{\mathbf{x} \in \mathbb{R}^n : h_i(\mathbf{x} = 0\}$ be the values of $\mathbf{x}$ which satisfy the j-th equality constraint. From Proposition 7, all of the sets Y_i are convex. From Example 7, all of the sets Z_j are convex. The feasible region X is the set of vectors $\mathbf{x}$ which satisfy *all* of the inequality and equality constraints, that is, the intersection of all of the sets Y_i and Z_j. By Proposition 8, therefore, X is convex. □

This is a very common situation in nonlinear optimization, where the functions representing the inequality constraints are convex, and the functions representing equality constraints are hyperplanes. From this theorem, this means that the feasible region must be convex. This is one reason why the standard form is relevant! If the inequality constraints were written as $g'(\mathbf{x}) \geq 0$, just because the g are convex would not mean that the feasible region is convex.[3]

[3] You can't flip the inequality in Proposition 7, that is, just because g is convex, it is not true that $X = \{x : g(x) \geq 0$ is convex.