Chapter 5

Convex Functions and Optimization

5.1 Convex Functions

Our next topic is that of convex functions. Again, we will concentrate on the context of a map \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) although the situation can be generalized almost without change by replacing \( \mathbb{R}^n \) with any real vector space \( V \). We will also find it useful, and in fact modern algorithms reflect this usefulness, to consider functions \( f : \mathbb{R}^n \rightarrow \mathbb{R}^* \) where \( \mathbb{R}^* \) is the set of extended real numbers introduced earlier\(^1\). Before beginning with the main part of the discussion, we want to keep some examples in mind.

A simple example of a convex function is \( x \mapsto x^2, \) \( x \in \mathbb{R} \). Indeed, in the smooth case it is arguably the most basic. As we learn in elementary calculus, this function is infinitely often differentiable and has a single critical point at which the function in fact takes on, not just a relative minimum, but an absolute minimum.

A critical point is, by definition, the solution of the equation \( \frac{d}{dx} x^2 = 2x \) or \( 2x = 0 \). We can apply the second derivative test at the point \( x = 0 \) to determine the nature of the critical point and we find that, since \( \frac{d^2}{dx^2} (x^2) = 2 > 0 \), the function is “concave up” and the critical point is indeed a point of relative minimum. That this point gives an absolute minimum to the function, we need only remark that the function values are bounded below by zero since \( x^2 > 0 \) for all \( x \neq 0 \).

We can give a similar example in \( \mathbb{R}^2 \).

**Example 5.1.1** We consider the function

\[
(x, y) \mapsto \frac{1}{2} x^2 + \frac{1}{3} y^2 := z,
\]

\(^1\)See the discussion in Appendix B.
The graph of this function is an elliptic paraboloid. In this case we expect that, once again, the minimum will occur at the origin of coordinates and, setting $f(x, y) = z$, we can compute

$$\nabla f(x, y) = \begin{pmatrix} x \\ 2/3 \end{pmatrix}, \quad \text{and} \quad H(f(x, y)) = \begin{pmatrix} 1 & 0 \\ 0 & 2/3 \end{pmatrix}. $$

Notice that the Hessian matrix, $H(f)$, is positive definite at all points $(x, y) \in \mathbb{R}^2$. Here the critical points are exactly those for which $\nabla f(x, y) = 0$ whose only solution is $x = 0, y = 0$. The second derivative test for problems of this type is just that

$$\det H(f(x, y)) = \left( \frac{\partial^2 f}{\partial x^2} \right) \left( \frac{\partial^2 f}{\partial y^2} \right) - \frac{\partial^2 f}{\partial x \partial y} > 0$$

which is clearly satisfied in the present example. It is simply a condition that guarantees that the Hessian is a positive definite matrix. Again, since for all $(x, y) \neq (0, 0)$, $z > 0$, the origin is a point where $f$ has an absolute minimum.

As the idea of a convex set lies at the foundation of our analysis, it is natural to describe the notion of convex functions in terms of such sets. We recall that, if $A$ and $B$ are two non-empty sets, then the Cartesian product of these two sets $A \times B$ is defined as the set of ordered pairs $\{(a, b) : a \in A, b \in B\}$. We recall that, if the two sets are convex then their Cartesian product is as well.

Previously, we introduced the idea of the epigraph of a function $f : X \to \mathbb{R}$ where $X \subset \mathbb{R}^n$ (see Appendix B section 2.2). For convenience, we repeat the definition here.

**Definition 5.1.2** Let $X \subset \mathbb{R}^n$ be a non-empty set. If $f : X \to \mathbb{R}$ then $\text{epi}(f)$ is defined by

$$\text{epi}(f) := \{(x, z) \in X \times \mathbb{R} | z \geq f(x)\}.$$ 

Convex functions are now defined in terms of their epigraphs:

**Definition 5.1.3** Let $C \subset \mathbb{R}^n$ be convex and $f : C \to \mathbb{R}^*$. Then the function $f$ is called a convex function provided $\text{epi}(f) \subset \mathbb{R} \times \mathbb{R}^n$ is a convex set.
Since we admit “infinite” values we have to be careful when doing computations. As a preliminary precaution we will assume that all the convex functions that we treat here, unless specifically stated to the contrary, are proper convex functions which means that the epigraph of $f$ is non-empty, and does not contain a line. Hence we insist that $f(x) > -\infty$ for all $x \in C$ and that $(\bar{x}) < \infty$ for at least one $x \in C$.

We emphasize that Definition 5.1.3 has the advantage of directly relating the theory of convex sets to the theory of convex functions. A more traditional definition is that a function is convex provided that, for any $x, y \in C$ and any $\lambda \in [0, 1]$

$$f \left( (1 - \lambda) x + \lambda y \right) \leq (1 - \lambda) f(x) + \lambda f(y),$$

which is sometimes referred to as Jensen’s inequality.

In fact, these definitions turn out to be equivalent. Indeed, we have the following result which involves a more general form of Jensen’s Inequality.

**Theorem 5.1.4** Let $C \subset \mathbb{R}^n$ be convex and $f : C \rightarrow \mathbb{R}^*$. Then the following are equivalent:

(a) $\text{epi}(f)$ is convex.

(b) For all $\lambda_1, \lambda_2, \ldots, \lambda_n$ with $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$, and points $x^{(i)} \in C, i = 1, 2, \ldots, n$, we have

$$f \left( \sum_{i=1}^{n} \lambda_i x^{(i)} \right) \leq \sum_{i=1}^{n} \lambda_i f(x^{(i)}).$$

(c) For any $x, y \in C$ and $\lambda \in [0, 1],$

$$f \left( (1 - \lambda) x + \lambda y \right) \leq (1 - \lambda) f(x) + \lambda f(y).$$

**Proof:** To see that (a) implies (b) we note that, if for all $i = 1, 2, \ldots, n,$ $(x^{(i)}, f(x^{(i)})) \in \text{epi}(f)$, then since this latter set is convex, we have

$$\sum_{i=1}^{n} \lambda_i (x^{(i)}, f(x^{(i)})) = \left( \sum_{i=1}^{n} \lambda_i x^{(i)}, \sum_{i=1}^{n} \lambda_i f(x^{(i)}) \right) \in \text{epi}(f),$$

which, in turn, implies that
\[ f \left( \sum_{i=1}^{n} \lambda_i x^{(i)} \right) \leq \sum_{i=1}^{n} \lambda_i f(x^{(i)}). \]

This establishes (b). It is obvious that (b) implies (c). So it remains only to show that (c) implies (a) in order to establish the equivalence.

To this end, suppose that \((x^{(1)}, z_1), (x^{(2)}, z_2) \in \text{epi}(f)\) and take \(0 \leq \lambda \leq 1\). Then

\[
(1 - \lambda) (x^{(1)}, z_1) + \lambda (x^{(2)}, z_2) = ((1 - \lambda) x^{(1)} + \lambda x^{(2)}, (1 - \lambda) z_1 + \lambda z_2),
\]

and since \(f(x^{(1)}) \leq z_1\) and \(f(x^{(2)}) \leq z_2\) we have, since \((1 - \lambda) > 0\), and \(\lambda > 0\), that

\[
(1 - \lambda) f(x^{(1)}) + \lambda f(x^{(2)}) \leq (1 - \lambda) z_1 + \lambda z_2.
\]

Hence, by the assumption (c), \(f \left( (1 - \lambda) x^{(1)} + \lambda x^{(2)} \right) \leq (1 - \lambda) z_1 + \lambda z_2\), which shows the point \( (1 - \lambda) x^{(1)} + \lambda x^{(2)}, (1 - \lambda) z_1 + \lambda z_2\) is in \(\text{epi}(f)\).

Convex functions are fundamental to minimization problems and we shall concentrate on them in the following sections. But since many applications to Economics involve maximization rather than minimization problems, many discussions in this area involve concave functions rather than convex ones. These latter functions are simply related to convex functions. Indeed a function \(f\) is concave if and only if the function \(-f\) is convex\(^2\).

In any systematic presentation, it is most economical to choose one type of optimization problem on which to concentrate; they are completely interchangeable in that minimizing a function \(f\) is the same problem as maximizing the function \(-f\). Here, we concentrate on convex functions.

We can see another connection between convex sets and convex functions if we introduce the indicator function, \(\psi_K\) of a set \(K \subset \mathbb{R}^n\). Indeed, \(\psi_K : \mathbb{R}^n \to \mathbb{R}^*\) is defined by

\[
\psi_K(x) = \begin{cases} 
0 & \text{if } x \in K, \\
+\infty & \text{if } x \not\in K.
\end{cases}
\]

**Proposition 5.1.5** A non-empty subset \(D \subset \mathbb{R}^n\) is convex if and only if its indicator function is convex.

\(^2\)This will also be true of quasi-convex and quasi-concave functions which we will define below.
5.1. CONVEX FUNCTIONS

Proof: The result follows immediately from the fact that \( \text{epi}(\psi_D) = D \times \mathbb{R}_{\geq 0} \).

Certain simple properties follow immediately from the analytic form of the definition (part (c) of the equivalence theorem above). Indeed, it is easy to see, and we leave it as an exercise for the reader, that if \( f \) and \( g \) are convex functions defined on a convex set \( C \), then \( f + g \) is likewise convex on \( C \) provided there is no point for which \( f(x) = \infty \) and \( g(x) = -\infty \). The same is true if \( \beta \in \mathbb{R}, \beta > 0 \) and we consider \( \beta f \).

Moreover, we have the following result which is extremely useful in broadening our class of convex functions.

Proposition 5.1.6 Let \( f : \mathbb{R}^n \to \mathbb{R} \) be given, \( x^{(1)}, x^{(2)} \in \mathbb{R}^n \) be fixed and define a function \( \varphi : [0, 1] \to \mathbb{R} \) by \( \varphi(\lambda) := f((1 - \lambda)x^{(1)} + \lambda x^{(2)}) \). Then the function \( f \) is convex on \( \mathbb{R}^n \) if and only if the function \( \varphi \) is convex on \([0, 1]\).

Proof: Suppose, first, that \( f \) is convex on \( \mathbb{R}^n \). Then it is sufficient to show that \( \text{epi}(\varphi) \) is a convex subset of \( \mathbb{R}^2 \). To see this, let \((\lambda_1, z_1), (\lambda_2, z_2) \in \text{epi}(\varphi)\) and let

\[
\begin{align*}
y^1 &= \lambda_1 x^1 + (1 - \lambda_1) x^2, \\
y^2 &= \lambda_2 x^1 + (1 - \lambda_2) x^2.
\end{align*}
\]

Then

\[
f(y^1) = \varphi(\lambda_1) \leq z_1 \quad \text{and} \quad f(y^2) = \varphi(\lambda_2) \leq z_2.
\]

Hence \((y^1, z_1) \in \text{epi}(f)\) and \((y^2, z_2) \in \text{epi}(f)\). Since \( \text{epi}(f) \) is a convex set, we also have \((\mu y^1 + (1 - \mu) y^2, \mu z_1 + (1 - \mu) z_2) \in \text{epi}(f)\) for every \( \mu \in [0, 1] \). It follows that \( f(\mu y^1 + (1 - \mu) y^2) \leq \mu z_1 + (1 - \mu) z_2 \).

Now

\[
\mu y^1 + (1 - \mu) y^2 = \mu(\lambda_1 x^1 + (1 - \lambda_1) x^2) + (1 - \mu)(\lambda_2 x^1 + (1 - \lambda_2) x^2)
\]

\[
= (\mu \lambda_1 + (1 - \mu) \lambda_2)x^1 + \mu (1 - \lambda_1) + (1 - \mu)(1 - \lambda_2)x^2,
\]

and since
1 − [µλ₁ + (1 − µ)λ₂] = [µ + (1 − µ)] − [µλ₁ + (1 − µ)λ₂] = µ(1 − λ₁) + (1 − µ)(1 − λ₂),

we have from the definition of φ that \( f(µy¹ + (1 − µ)y²) = ϕ(µλ₁ + (1 − µ)λ₂) \) and so \( (µλ₁ + (1 − µ)λ₂, µz₁ + (1 − µ)z₂) ∈ epi(ϕ) \) i.e., φ is convex.

We leave the proof of the converse statement as an exercise. ■

It should be clear that if \( f : \mathbb{R}^n → \mathbb{R} \) is a linear or affine, then \( f \) is convex. Indeed, suppose that for a vector \( a ∈ \mathbb{R}^n \) and a real number \( b \), the affine function \( f \) is given by \( f(x) = <a, x> + b \). Then we have, for any \( λ ∈ [0, 1] \),

\[
\begin{align*}
  f((1 − λ)x + λy) &= <a, (1 − λ)x + λy> \\
                      &= (1 − λ)<a, x> + λ<a, y> + (1 − λ)b + λb \\
                      &= (1 − λ)(<a, x> + b) + λ(<a, y> + b) = (1 − λ)f(x) + λf(y),
\end{align*}
\]

and so \( f \) is convex, the weak inequality being an equality in this case.

In the case that \( f \) is linear, that is \( f(x) = <a, x> \) for some \( a ∈ \mathbb{R}^n \) then it is easy to see that the map \( ϕ : x → [f(x)]^2 \) is also convex. Indeed, if \( x, y ∈ \mathbb{R}^n \) then, setting \( α = f(x) \) and \( β = f(y) \), and taking \( 0 < λ < 1 \) we have

\[
\begin{align*}
  (1 − λ)ϕ(x) + λϕ(y) − ϕ((1 − λ)x) + λy) &= (1 − λ)α² + λβ² − ((1 − λ)α + λβ)² \\
                                               &= (1 − λ)λ(α − β)² ≥ 0.
\end{align*}
\]

Note, that in particular for the function \( f : \mathbb{R} → \mathbb{R} \) given by \( f(x) = x \) is linear and that \( [f(x)]^2 = x² \) so that we have a proof that the function that we usually write \( y = x² \) is a convex function.

The next result expands our repertoire of convex functions.

**Proposition 5.1.7** (a) If \( A : \mathbb{R}^m → \mathbb{R}^n \) is linear and \( f : \mathbb{R}^n → \mathbb{R}^* \) is convex, then \( f ∘ A \) is convex as a map from \( \mathbb{R}^m \) to \( \mathbb{R} \).
(b) If $f$ is as in part (a) and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is convex and non-decreasing, then $\varphi \circ f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex.

(c) Let $\{ f_\alpha \}_{\alpha \in \mathcal{A}}$ be a family of functions $f_\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^*$ then its upper envelope $\sup_{\alpha \in \mathcal{A}} f_\alpha$ is convex.

**Proof:** To prove (a) we use Jensen’s inequality: Given any $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$ we have

$$(f \circ A)((1 - \lambda)x + \lambda y) = f((1 - \lambda)(Ax) + \lambda(Ay)) \leq (1 - \lambda)f(Ax) + \lambda f(Ay) = (1 - \lambda)(f \circ A)(x) + \lambda(f \circ A)(y).$$

For part (b), again we take $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$. Then

$$(\varphi \circ f)[(1 - \lambda)x + \lambda y] \leq \varphi[(1 - \lambda)f(x) + \lambda f(y)]$$
$$\leq (1 - \lambda)\varphi(f(x)) + \lambda\varphi(f(y)) = (1 - \lambda)(\varphi \circ f)(x) + \lambda(\varphi \circ f)(y),$$

where the first inequality comes from the convexity of $f$ and the monotonicity of $\varphi$ and the second from the convexity of this later function. This proves part (b).

To establish part (c) we note that, since the arbitrary intersection of convex sets is convex, it suffices to show that

$$\text{epi} \left( \sup_{\alpha \in \mathcal{A}} f_\alpha \right) = \bigcup_{\alpha \in \mathcal{A}} \text{epi} (f_\alpha).$$

To check the equality of these two sets, start with a point

$$(x, z) \in \text{epi} \left( \sup_{\alpha \in \mathcal{A}} f_\alpha \right).$$

Then $z \geq \sup_{\alpha \in \mathcal{A}} f_\alpha(x)$ and so, for all $\beta \in \mathcal{A}$, $z \geq f_\beta(x)$. Hence, by definition, $(x, z) \in \text{epi} f_\beta$ for all $\beta$ from which it follows that

$$(x, z) \in \bigcap_{\alpha \in \mathcal{A}} \text{epi} (f_\alpha).$$
Conversely, suppose \((x, z) \in \text{epi}(f_\alpha)\) for all \(\alpha \in \mathcal{A}\). Then \(z \geq f_\alpha(x)\) for all \(\alpha \in \mathcal{A}\) and hence \(z \geq \sup_{\alpha \in \mathcal{A}} f_\alpha\). But this, by definition, implies \((x, z) \in \text{epi}(\sup_{\alpha \in \mathcal{A}} f_\alpha)\). This completes the proof of part (c) and the proposition is proved. 

Next, we introduce the definition:

**Definition 5.1.8** Let \(f : \mathbb{R}^n \to \mathbb{R}^*\), and \(\alpha \in \mathbb{R}\). Then the sets

\[
S(f, \alpha) := \{x \in \mathbb{R}^n \mid f(x) < \alpha\} \quad \text{and} \quad \overline{S}(f, \alpha) := \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\},
\]

are called lower sections of the function \(f\).

**Proposition 5.1.9** If \(f : \mathbb{R} \to \mathbb{R}^*\) is convex, then its lower sections are likewise convex.

The proof of this result is trivial and we omit it.

The converse of this last proposition is false as can be easily seen from the function \(x \mapsto x^{\frac{1}{2}}\) from \(\mathbb{R}_+\) to \(\mathbb{R}\). However, the class of functions whose lower level sets \(\overline{S}(f, \alpha)\) (or equivalently the sets \(S(f, \alpha)\)) are all convex is likewise an important class of functions and are called quasi-convex. These functions appear in game theory nonlinear programming (optimization) problems and mathematical economics. For example, quasi-convex utility functions imply that consumers have convex preferences. They are obviously generalizations of convex functions since every convex function is clearly quasi-convex. However they are not as easy to work with. In particular, while the sum of two convex functions is convex, the same is not true of quasi-convex functions as the following example shows.

**Example 5.1.10** Define

\[
f(x) = \begin{cases} 
0 & x \leq -2 \\
-(x + 2) & -2 < x \leq -1 \\
x & -1 < x \leq 0 \\
0 & x > 0
\end{cases}
\quad \text{and} \quad g(x) = \begin{cases} 
0 & x \leq 0 \\
x & 0 < x \leq 1 \\
x - 2 & 1 < x \leq 2 \\
0 & x > 2
\end{cases}.
\]

Here, the functions are each concave, the level sections are convex for each function so that each is quasi-convex, and yet the level section corresponding to \(\alpha = -1/2\) for the sum \(f + g\) is not convex. Hence the sum is not quasi-convex.
It is useful for applications to have an analytic criterion for quasi-convexity. This is the content of the next result.

**Proposition 5.1.11** A function \( f : \mathbb{R}^n \rightarrow \mathbb{R}^* \) is quasi-convex if and only if, for any \( x, y \in \mathbb{R}^n \) and any \( \lambda \in [0, 1] \) we have

\[
f((1 - \lambda) x + \lambda y) \leq \max\{f(x), f(y)\}.
\]

**Proof:** Suppose that the sets \( \overline{S}(f, \alpha) \) are convex for every \( \alpha \). Let \( x, y \in \mathbb{R}^n \) and let \( \bar{\alpha} := \max\{f(x), f(y)\} \). Then \( \overline{S}(f, \bar{\alpha}) \) is convex and, since both \( f(x) \leq \bar{\alpha} \) and \( f(y) \leq \bar{\alpha} \), we have that both \( x \) and \( y \) belong to \( \overline{S}(f, \bar{\alpha}) \). Since this latter set is convex, we have

\[
(1 - \lambda) x + \lambda y \in \overline{S}(f, \bar{\alpha}) \text{ or } f((1 - \lambda) x + \lambda y) \leq \bar{\alpha} = \max\{f(x), f(y)\}.
\]

As we have seen above, the sum of two quasi-convex functions may well not be quasi-convex. With this analytic test for quasi-convexity, we can check that there are certain operations which preserve quasi-convexity. We leave the proof of the following result to the reader.

**Proposition 5.1.12**

(a) If the functions \( f_1, \ldots, f_k \) are quasi-convex and \( \alpha_1, \ldots, \alpha_k \) are non-negative real numbers, then the function \( f := \max\{\alpha_1 f_1, \ldots, \alpha_k f_k\} \) is quasi-convex.

(b) If \( \varphi : \mathbb{R} \rightarrow \mathbb{R} \) is a non-decreasing function and \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is quasi-convex, then the composition \( \varphi \circ f \) is a quasi-convex function.

We now return to the study of convex functions.

A simple sketch of the parabola \( y = x^2 \) and any horizontal cord (which necessarily lies above the graph) will convince the reader that all points in the domain corresponding to the values of the function which lie below that horizontal line, form a convex set in the domain. Indeed, this is a property of convex functions which is often useful.

**Proposition 5.1.13** If \( C \subset \mathbb{R}^n \) is a convex set and \( f : C \rightarrow \mathbb{R} \) is a convex function, then the level sets \( \{x \in C \mid f(x) \leq \alpha\} \) and \( \{x \in C \mid f(x) < \alpha\} \) are convex for all scalars \( \alpha \).
CHAPTER 5. CONVEX FUNCTIONS

Proof: We leave this proof as an exercise.

Notice that, since the intersection of convex sets is convex, the set of points simultaneously satisfying \( m \) inequalities \( f_1(x) \leq c_1, f_2(x) \leq c_2, \ldots, f_m(x) \leq c_m \) where each \( f_i \) is a convex function, defines a convex set. In particular, the polygonal region defined by a set of such inequalities when the \( f_i \) are affine is convex.

From this result, we can obtain an important fact about points at which a convex function attains a minimum.

**Proposition 5.1.14** Let \( C \subset \mathbb{R} \) be a convex set and \( f : C \to \mathbb{R} \) a convex function. Then the set of points \( M \subset C \) at which \( f \) attains its minimum is convex. Moreover, any relative minimum is an absolute minimum.

**Proof:** If the function does not attain its minimum at any point of \( C \), then the set of such points in empty, which is a convex set. So, suppose that the set of points at which the function attains its minimum is non-empty and let \( m \) be the minimal value attained by \( f \). If \( x, y \in M \) and \( \lambda \in [0, 1] \) then certainly \( (1 - \lambda)x + \lambda y \in C \) and so

\[
m \leq f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y) = m,
\]

and so the point \( (1 - \lambda)x + \lambda y \in M \). Hence \( M \), the set of minimal points, is convex.

Now, suppose that \( x^* \in C \) is a relative minimum point of \( f \), but that there is another point \( \hat{x} \in C \) such that \( f(\hat{x}) < f(x^*) \). On the line \( (1 - \lambda)\hat{x} + \lambda x^* \), \( 0 < \lambda < 1 \), we have

\[
f((1 - \lambda)\hat{x} + \lambda x^*) \leq (1 - \lambda)f(\hat{x}) + \lambda f(x^*) < f(x^*),
\]

contradicting the fact that \( x^* \) is a relative minimum point. ■

Again, the example of the simple parabola, shows that the set \( M \) may well contain only a single point, i.e., it may well be that the minimum point is unique. We can guarantee that this is the case for an important class of convex functions.

**Definition 5.1.15** A real-valued function \( f \), defined on a convex set \( C \subset \mathbb{R} \) is said to be **strictly convex** provided, for all \( x, y \in C, x \neq y \) and \( \lambda \in (0, 1) \), we have

\[
f((1 - \lambda)x + \lambda y) < (1 - \lambda)f(x) + \lambda f(y).
\]
Proposition 5.1.16 If $C \subset \mathbb{R}^n$ is a convex set and $f : C \to \mathbb{R}$ is a strictly convex function then $f$ attains its minimum at, at most, one point.

Proof: Suppose that the set of minimal points $M$ is not empty and contains two distinct points $x$ and $y$. Then, for any $0 < \lambda < 1$, since $M$ is convex, we have $(1 - \lambda)x + \lambda y \in M$. But $f$ is strictly convex. Hence

$$m = f((1 - \lambda)x + \lambda y) < (1 - \lambda)f(x) + \lambda f(y) = m,$$

which is a contradiction. ■

If a function is differentiable then, as in the case in elementary calculus, we can give characterizations of convex functions using derivatives. If $f$ is a continuously differentiable function defined on an open convex set $C \subset \mathbb{R}^n$ then we denote its gradient at $x \in C$, as usual, by $\nabla f(x)$. The excess function

$$E(x, y) := f(y) - f(x) - \langle \nabla f(x), y - x \rangle$$

is a measure of the discrepancy between the value of $f$ at the point $y$ and the value of the tangent approximation at $x$ to $f$ at the point $y$. This is illustrated in the next figure.

Now we introduce the notion of a monotone derivative

Definition 5.1.17 The map $x \mapsto \nabla f(x)$ is said to be monotone on $C \subset \mathbb{R}^n$ provided

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0,$$

for all $x, y \in C$.

We can now characterize convexity in terms of the function $E$ and the monotonicity concept just introduced. However, before stating and proving the next theorem, we need a lemma.

Lemma 5.1.18 Let $f$ be a real-valued, differentiable function defined on an open interval $I \subset \mathbb{R}$. Then if the first derivative $f'$ is a non-decreasing function on $I$, the function $f$ is convex on $I$. 
Proof: Choose \( x, y \in I \) with \( x < y \), and for any \( \lambda \in [0, 1] \), define \( z_\lambda := (1 - \lambda)x + \lambda y \). By the Mean Value Theorem, there exist \( u, v \in \mathbb{R} \), \( x \leq v \leq z_\lambda \leq u \leq y \) such that

\[
f(y) = f(z_\lambda) + (y - z_\lambda)f'(u), \quad \text{and} \quad f(z_\lambda) = f(x) + (z_\lambda - x)f'(v).
\]

But, \( y - z_\lambda = y - (1 - \lambda)x - \lambda y = (1 - \lambda)(y - x) \) and \( z_\lambda - x = (1 - \lambda)x + \lambda y - x = \lambda(y - x) \) and so the two expressions above may be rewritten as

\[
f(y) = f(z_\lambda) + \lambda(y - x)f'(u), \quad \text{and} \quad f(z_\lambda) = f(x) + \lambda(y - x)f'(v).
\]

Since, by choice, \( v < u \), and since \( f' \) is non-decreasing, this latter equation yields

\[
f(z_\lambda) \leq f(x) + \lambda(y - x)f'(u).
\]

Hence, multiplying this last inequality by \((1 - \lambda)\) and the expression for \( f(y) \) by \(-\lambda\) and adding we get

\[
(1 - \lambda)f(z_\lambda) - \lambda f(y) \leq (1 - \lambda)f(x) + \lambda(1 - \lambda)(y - x)f'(u) - \lambda f(z_\lambda) - \lambda(1 - \lambda)(y - x)f'(u),
\]

which we then rearrange to yield

\[
(1 - \lambda)f(z_\lambda) + \lambda f(z_\lambda) = f(z_\lambda) \leq (1 - \lambda)f(x) + \lambda f(y),
\]

and this is just the condition for the convexity of \( f \)  

We can now prove a theorem which gives three different characterizations of convexity for continuously differentiable functions.

**Theorem 5.1.19** Let \( f \) be a continuously differentiable function defined on an open convex set \( C \subset \mathbb{R}^n \). Then the following are equivalent:

(a) \( E(x, y) \geq 0 \) for all \( x, y \in C \);

(b) the map \( x \mapsto \nabla f(x) \) is monotone in \( C \);

(c) the function \( f \) is convex on \( C \).
**Proof:** Suppose that (a) holds, i.e. $E(x, y) \geq 0$ on $C \times C$. Then we have both

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle,$$

and

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle = -\langle \nabla f(y), y - x \rangle.$$

Then, from the second inequality, $f(y) - f(x) \leq \langle \nabla f(y), x - y \rangle$, and so

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle = \langle \nabla f(y), y - x \rangle - \langle \nabla f(x), y - x \rangle \geq (f(y) - f(x)) - (f(x) - f(x)) = 0.$$

Hence, the map $x \mapsto \nabla f(x)$ is monotone in $C$.

Now suppose the map $x \mapsto \nabla f(x)$ is monotone in $C$, and choose $x, y \in C$. Define a function $\varphi : [0, 1] \to \mathbb{R}$ by $\varphi(t) := f(x + t(y - x))$. We observe, first, that if $\varphi$ is convex on $[0, 1]$ then $f$ is convex on $C$. To see this, let $u, v \in [0, 1]$ be arbitrary. On the one hand,

$$\varphi((1 - \lambda)u + \lambda v) = f(x + [(1 - \lambda)u + \lambda v](y - x)) = f((1 - [(1 - \lambda)u + \lambda v])x + ((1 - \lambda)u + \lambda v)y),$$

while, on the other hand,

$$\varphi((1 - \lambda)u + \lambda v) \leq (1 - \lambda)f(x + u(y - x)) + f(x + v(y - x)).$$

Setting $u = 0$ and $v = 1$ in the above expressions yields

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y).$$

so the convexity of $\varphi$ on $[0, 1]$ implies the convexity of $f$ on $C$.

Now, choose any $\alpha, \beta$, $0 \leq \alpha < \beta \leq 1$. Then

$$\varphi'(\beta) - \varphi'(\alpha) = \langle (\nabla f(x + \beta(y - x)) - \nabla f(x + \alpha(y - x)), y - x \rangle.$$

Setting $u := x + \alpha(y - x)$ and $v := x + \beta(y - x)$ we have $v - u = (\beta - \alpha)(y - x)$ and so

$$\varphi'(\beta) - \varphi'(\alpha) = \langle (\nabla f(v) - \nabla f(u), v - u \rangle \geq 0.$$
Hence $\varphi'$ is non-decreasing, so that the function $\varphi$ is convex.

Finally, if $f$ is convex on $C$, then, for fixed $x, y \in C$ define

$$h(\lambda) := (1 - \lambda) f(x) + \lambda f(y) - f((1 - \lambda) x + \lambda y).$$

Then $\lambda \mapsto h(\lambda)$ is a non-negative, differentiable function on $[0, 1]$ and attains its minimum at $\lambda = 0$. Therefore $0 \leq h'(0) = E(x, y)$, and the proof is complete. ■

As an immediate corollary, we have

**Corollary 5.1.20** Let $f$ be a continuously differentiable convex function defined on a convex set $C$. If there is a point $x^* \in C$ such that, for all $y \in C$, $\langle \nabla f(x^*), y - x^* \rangle \geq 0$, then $x^*$ is an absolute minimum point of $f$ over $C$.

**Proof:** By the preceding theorem, the convexity of $f$ implies that

$$f(y) - f(x^*) \geq \langle \nabla f(x^*), y - x^* \rangle,$$

and so, by hypothesis,

$$f(y) \geq f(x^*) + \langle \nabla f(x^*), y - x^* \rangle \geq f(x^*).$$

The inequality $E(x, y) \geq 0$ shows that local information about a convex function, given in terms of the derivative at a point, gives us global information in terms of a global underestimator of the function $f$. In a way, this is the key property of convex functions. For example, suppose that $\nabla f(x) = 0$. Then, for all $y \in \text{dom}(f), f(y) \geq f(x)$ so that $x$ is a global minimizer of the convex function $f$.

It is also important to remark that the hypothesis that the convex function $f$ is defined on a convex set is crucial, both for the first order conditions as well as for the second order conditions. Indeed, if we consider the function $f(x) = 1/x^2$ with domain $\{x \in \mathbb{R} | x \neq 0\}$. The usual second order condition $f''(x) > 0$ for all $x \in \text{dom}(f)$ yet $f$ is not convex there so that the second order test fails.

The condition $E(x, y) \geq 0$ can be given an important geometrical interpretation in terms of epigraphs. Indeed if $f$ is convex and $x, y \in \text{dom}(f)$ then for $(x, z) \in \text{epi}(f)$, then
5.1. CONVEX FUNCTIONS

\[ z \geq f(y) \geq f(x) + \nabla f(x)^\top (y - x), \]

can be expressed as

\[
\begin{bmatrix}
\nabla f(x) \\
-1
\end{bmatrix}^\top \left( \begin{bmatrix}
y \\
z
\end{bmatrix} - \begin{bmatrix}
x \\
f(x)
\end{bmatrix} \right) \leq 0.
\]

This shows that the hyperplane defined by \((\nabla f(x), -1)^\top\) supports \(\text{epi}(f)\) at the boundary point \((x, f(x))\).

We now turn to so-called second order criteria for convexity. The discussion involves the Hessian matrix of a twice continuously differentiable function, and depends on the question of whether this matrix is positive semi-definite or even positive definite (for strict convexity) Let us recall some definitions.

**Definition 5.1.21** A real symmetric \(n \times n\) matrix \(A\) is said to be

(a) **Positive definite** provided \(x^\top A x > 0\) for all \(x \in \mathbb{R}^n, x \neq 0\).

(b) **Negative definite** provided \(x^\top A x < 0\) for all \(x \in \mathbb{R}^n, x \neq 0\).

(c) **Positive semidefinite** provided \(x^\top A x \geq 0\) for all \(x \in \mathbb{R}^n, x \neq 0\).

(d) **Negative semidefinite** provided \(x^\top A x \leq 0\) for all \(x \in \mathbb{R}^n, x \neq 0\).

(e) **Indefinite** provided \(x^\top A x\) takes on values that differ in sign.

It is important to be able to determine if a matrix is indeed positive definite. In order to do this, a number of criteria have been developed. Perhaps the most important characterization is in terms of the eigenvalues.

**Theorem 5.1.22** Let \(A\) be a real symmetric \(n \times n\) matrix. Then \(A\) is positive definite if and only if all its eigenvalues are positive.

**Proof:** If \(A\) is positive definite and \(\lambda\) is an eigenvalue of \(A\), then, for any eigenvector \(x\) belonging to \(\lambda\)
\[ x^\top A x = \lambda x^\top x = \lambda \|x\|^2, \]

Hence
\[ \lambda = \frac{x^\top A x}{\|x\|} > 0 \]

Conversely, suppose that all the eigenvalues of \( A \) are positive. Let \( \{x_1, \ldots, x_n\} \) be an orthonormal set of eigenvectors of \( A \). Hence any \( x \in \mathbb{R}^n \) can be written as
\[ x = \alpha_1 x_1 + \alpha_2 b x_2 + \cdots + \alpha_n x_n \]

with
\[ \alpha_i = x^\top x_i \quad \text{for} \quad i = 1, 2, \ldots, n, \quad \text{and} \quad \sum_{i=1}^n \alpha_i^2 = \|x\|^2 > 0. \]

It follows that
\[ x^\top A x = (\alpha_1 x_1 + \cdots + \alpha_n x_n)^\top (\alpha_1 \lambda_1 x_1 + \cdots + \alpha_n \lambda_n x_n) \]
\[ = \sum_{i=1}^n \alpha_i^2 \lambda_i \geq (\min \lambda_i) \|x\|^2 > 0. \]

Hence \( A \) is positive definite. \( \blacksquare \)

In simple cases where we can compute the eigenvalues easily, this is a useful criterion.

**Example 5.1.23** Let
\[ A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}. \]

Then the eigenvalues are the roots of
\[ \det (A - \lambda I) = (2 - \lambda) (5 - \lambda) - 4 = (\lambda - 1) (\lambda - 6). \]

Hence the eigenvalues are both positive and hence the matrix is positive definite. In this particular case it is easy to check directly that \( A \) is positive definite. Indeed
\[(x_1, x_2) \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} 2x_1 - 2x_2 \\ -2x_1 + 5x_2 \end{pmatrix} = 2x_2^2 - 4x_1x_2 + 5x_2^2 = 2 [x_1^2 - 2x_1x_2 + x_2^2] + 4x_2^2 = 2(x_1 - x_2)^2 + 4x_2^2 > 0.

This last theorem has some immediate useful consequences. First, if \( A \) is positive definite, then \( A \) must be nonsingular, since singular matrices have \( \lambda = 0 \) as an eigenvalue. Moreover, since we know that the \( \det(A) \) is the product of the eigenvalues, and since each eigenvalue is positive, then \( \det(A) > 0 \). Finally, we have the following result which depends on the notion of leading principle submatrices.

**Definition 5.1.24** Given any \( n \times N \) matrix \( A \), let \( A_r \) denote the matrix formed by deleting the last \( n - r \) rows and columns of \( A \). Then \( A_4 \) is called the **leading principal submatrix** of \( A \).

**Proposition 5.1.25** If \( A \) is a symmetric positive definite matrix then the leading principal submatrices \( A_1, A_2, \ldots, A_n \) of \( A \) are all positive definite. In particular, \( \det(A_r) > 0 \).

**Proof:** Let \( x_r = (x_1, x_2, \ldots, x_r)^\top \) be any non-zero vector in \( \mathbb{R}^r \). Set

\[
\mathbf{x} = (x_1, x_2, \ldots, x_r, 0, \ldots, 0)^\top.
\]

Since \( x_r^\top A_r x_r = x^\top A x > 0 \), it follows that \( A_r \) is positive definite, by definition.

This proposition is half of the famous criterion of Sylvester for positive definite matrices.

**Theorem 5.1.26** A real, symmetric matrix \( A \) is positive definite if and only if all of its leading principle minors are positive definite.

We will not prove this theorem here but refer the reader to his or her favorite treatise on linear algebra.
Example 5.1.27 Let

\[
A = \begin{pmatrix}
2 & -1 & 0 \\
-1 & 2 & -1 \\
0 & -1 & 2
\end{pmatrix}
\]

Then

\[
A_2 = (2), \quad A_2 = \begin{pmatrix}
2 & -1 \\
-1 & 2
\end{pmatrix}, \quad A_3 = A.
\]

Then

\[
det A_1 = 2, \quad det A_2 = 4 - 1 = 3, \quad \text{and} \quad det A = 4.
\]

Hence, according to Sylvester's criterion, the matrix \(A\) is positive definite.

Now we are ready to look at second order conditions for convexity.

Proposition 5.1.28 Let \(D \subset \mathbb{R}^n\) be an open convex set and let \(f : D \to \mathbb{R}\) be twice continuously differentiable in \(D\). Then \(f\) is convex if and only if the Hessian matrix of \(f\) is positive semidefinite throughout \(D\).

Proof: By Taylor’s Theorem we have

\[
f(y) = f(x) + \left\langle \nabla f(x), y - x \right\rangle + \frac{1}{2} \left\langle y - x, \nabla^2 f(x + \lambda (y - x))(y - x) \right\rangle,
\]

for some \(\lambda \in [0, 1]\). Clearly, if the Hessian is positive semi-definite, we have

\[
f(y) \geq f(x) + \left\langle \nabla f(x), y - x \right\rangle,
\]

which in view of the definition of the excess function, means that \(E(x, y) \geq 0\) which implies that \(f\) is convex on \(D\).

Conversely, suppose that the Hessian is not positive semi-definite at some point \(x \in D\). Then, by the continuity of the Hessian, there is a \(y \in D\) so that, for all \(\lambda \in [0, 1]\),

\[
\left\langle y - x, \nabla^2 f(x + \lambda (y - x))(y - x) \right\rangle < 0,
\]

which, in light of the second order Taylor expansion implies that \(E(x, y) < 0\) and so \(f\) cannot be convex. □
5.1. CONVEX FUNCTIONS

Let us consider, as an example, the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ with $\text{dom}(f) = \mathbb{R}^n$, given by

$$f(x) = \frac{1}{2} x^\top Q x + q^\top x + r,$$

with $Q$ and $n \times n$ symmetric matrix, $q \in \mathbb{R}^n$ and $r \in \mathbb{R}$. Then since as we have seen previously, $\nabla^2 f(x) = Q$, the function $f$ is convex if and only if $Q$ is positive semidefinite. Strict convexity of $f$ is likewise characterized by the positive definiteness of $Q$.

These first and second-order necessary conditions give us methods of showing that a given function is convex. Thus, we either check the definition, Jensen’s inequality, using the equivalence that is given by Theorem 2.1.3, or showing that the Hessian is positive semi-definite. Let us look as some simple examples.

**Example 5.1.29**

(a) The real-valued function defined on $\mathbb{R}_+ \times \ln(x)$. Then, since this function $C^2(\mathbb{R}_+)$ and $f'(x) = \ln(x) + 1$ and $f''(x) = 1/x > 0$, we see that $f$ is (even strictly) convex.

(b) The max function $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbb{R}^n$. Here we can use Jensen’s inequality. Let $\lambda \in [0, 1]$ then

$$f((1 - \lambda)x + \lambda y) = \max_{1 \leq i \leq n}(\lambda x_i + \lambda y_i) \leq \lambda \max_{1 \leq i \leq n} x_i + (1 - \lambda) \max_{1 \leq i \leq n} y_i$$

$$= (1 - \lambda)f(x) + \lambda f(y).$$

(c) The function $q : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ given by $q(x, y) = x^2/y$ is convex. In this case, $\nabla q(x, y) = (2x/y, -x^2/y^2)\top$ while an easy computation shows

$$\nabla^2 q(x, y) = \frac{2}{y^3} \begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix}.$$

Since $y > 0$ and

$$(u_1, u_2) \begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix} (u_1, u_2)^\top = (u_1 y - u_2 x)^2 \geq 0,$$

the Hessian of $q$ is positive definite and the function is convex.
CHAPTER 5. CONVEX FUNCTIONS