These notes correspond to Section 2.3 in the text.

Convex Functions

We are now prepared describe the usefulness of the convex sets introduced in the previous section. For certain functions defined on convex sets, it can be very easy to determine whether they have a global minimizer, and if so, to compute it. A class of functions that has this property is introduced through the following definition.

Definition Let $C \subseteq \mathbb{R}^n$ be a convex set, and let $f : C \to \mathbb{R}$. Then $f(\mathbf{x})$ is *convex* on C if

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0, 1]$. If

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in C$, $\mathbf{x} \ne \mathbf{y}$, and $\lambda \in (0, 1)$, then we say that $f(\mathbf{x})$ is *strictly convex.*
    We also say that $f(\mathbf{x})$ is *concave* on C if

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \ge \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0, 1]$. If

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) > \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in C$, $\mathbf{x} \ne \mathbf{y}$, and $\lambda \in (0, 1)$, then we say that $f(\mathbf{x})$ is *strictly concave.*

Intuitively, the graph of a convex function lies on or below any chord between two points on the graph. For a single-variable function $f(x)$, the following two other characterizations of convexity are helpful:

1. The graph of a convex function $f(x)$ lies above each of its tangent lines. That is, if $f(x)$ is convex on an interval I and $x_1, x_2 \in I$, then

$$f(x_1) + f'(x_1)(x_2 - x_1) \le f(x_2).$$

2. If $f(x)$ is twice differentiable on an interval I, then $f(x)$ is convex on I if and only if $f''(x) \geq 0$ on I. Furthermore, if $f''(x) > 0$ on I, then $f(x)$ is strictly convex, though the converse is not necessarily true. Similarly, $f(x)$ is concave if and only if $f''(x) \leq 0$ on I, and strictly concave if $f''(x) < 0$ on I.

Example Let $f(x) = x^2$. Then $f(x)$ is strictly convex on $\mathbb{R}$, as $f''(x) = 2$ on $\mathbb{R}$. $\square$

Example Let $\mathbf{a} \in \mathbb{R}^n$ be a fixed vector, and let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(\mathbf{x}) = (\mathbf{a} \cdot \mathbf{x})^2.$$

Then, by the linearity of the dot product,

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) = (\mathbf{a} \cdot [\lambda \mathbf{x} + (1 - \lambda)\mathbf{y})^2 = [\lambda \mathbf{a} \cdot \mathbf{x} + (1 - \lambda)\mathbf{a} \cdot \mathbf{y}]^2.$$

By the convexity of $\varphi(t) = t^2$ from the previous example, we have

$$[\lambda t_1 + (1 - \lambda)t_2]^2 \leq \lambda t_1^2 + (1 - \lambda)t_2^2,$$

with $t_1 = \mathbf{a} \cdot \mathbf{x}$ and $t_2 = \mathbf{a} \cdot \mathbf{y}$. We conclude that

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}),$$

and therefore $f(\mathbf{x})$ is convex on $\mathbb{R}^n$. $\square$

One essential consequence of convexity is given by the following theorem.

Theorem If $f(\mathbf{x})$ is a convex function defined on an open convex set C, then $f(\mathbf{x})$ is continuous on C.

The definition of a convex function can be generalized to apply to convex combinations of any number of points.

Theorem Let $C \subseteq \mathbb{R}^n$ be a convex set, and let $f : C \to \mathbb{R}$ be convex on C. If $\lambda_1, \lambda_2, \ldots, \lambda_k \in [0, 1]$, $\sum_{i=1}^{k} \lambda_i = 1$, and $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)} \in C$, then

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

If $f(\mathbf{x})$ is strictly convex on C and $\lambda_i > 0$ for $i = 1, 2, \ldots, n$, then the above inequality holds, with equality if and only if all of the $\mathbf{x}^{(i)}$ are equal.

The main benefit of knowing whether a function is convex, as far as optimization is concerned, is provided by the following theorem.

Theorem Let $C \subseteq \mathbb{R}^n$ be a convex set, and let $f : C \to \mathbb{R}$ be convex on C. Then any local minimizer of $f(\mathbf{x})$ if a global minimizer. Furthermore, if $f(\mathbf{x})$ is strictly convex on C, then any local minimizer of $f(\mathbf{x})$ is the unique strict global minimizer of $f(\mathbf{x})$ on C.

One difficulty with applying the preceding theorem is that it can be very difficult to determine whether a function is convex using the definition of convexity directly. The following theorem is a first step toward the development of simpler methods for checking convexity.

Theorem Let $D \subseteq \mathbb{R}^n$ be convex, and let $f : D \to \mathbb{R}$ have continuous first partial derivatives on D. Then $f(\mathbf{x})$ is convex on D if and only if

$$f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in D$. In addition, $f(\mathbf{x}$ is strictly convex on D if and only if

$$f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) < f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in D$, $\mathbf{x} \neq \mathbf{y}$.

The following corollary to the preceding theorem makes it very easy to classify critical points of convex functions.

Corollary Let $D \subseteq \mathbb{R}^n$ be convex, and let $f : D \to \mathbb{R}$ be convex and have continuous first partial derivatives on D. Then any critical point of $f(\mathbf{x})$ in D is a global minimizer of $f(\mathbf{x})$ on D.

The following theorem is very useful for determining whether a function is convex or strictly convex.

Theorem Let $C \subseteq \mathbb{R}^n$ be open and convex, and let $f : C \to \mathbb{R}$ have continuous second partial derivatives on C.

1. If the Hessian $Hf(\mathbf{x})$ of $f(\mathbf{x})$ is positive semidefinite on C, then $f(\mathbf{x})$ is convex on C.

2. If $Hf(\mathbf{x})$ is positive definite on C, then $f(\mathbf{x})$ is strictly convex on C.

Example Let

$$f(x, y, z) = 2x^2 + y^2 + z^2 + 2yz.$$

This function has the Hessian

$$Hf(x, y, z) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix},$$

which has minors $\Delta_1 = 4, \Delta_2 = 8$, and $\Delta_3 = 0$, so $Hf(x, y, z)$ is positive semidefinite. Therefore, $f(x, y, z)$ is convex on $\mathbb{R}^3$. We cannot conclude that $f(x, y, z)$ is strictly convex on $\mathbb{R}^3$, as $Hf(x, y, z)$ is not positive definite, and in fact it is not strictly convex, as $f(x, y, z)$ can be rewritten as

$$f(x, y, z) = 2x^2 + (y + z)^2,$$

which shows that $f(x, y, z) = 0$ along the line $x = 0$, $y = -z$. The graph of a strictly convex function must lie strictly *below* the interior of a chord between any two points on the graph, which cannot occur if the graph contains any line segment. □

A partial converse of the preceding theorem holds: if $f(\mathbf{x})$ is convex on C, then $Hf(\mathbf{x})$ is positive semidefinite on C. However, if $f(\mathbf{x})$ is strictly convex on C, then $Hf(\mathbf{x})$ is not necessarily positive definite; it may be positive semidefinite.

Example Let $f(x) = x^4$. Then $f(x)$ is strictly convex on $\mathbb{R}$, but its second derivative $f''(x) = 12x^2$ is not positive definite, as $f''(0) = 0$. □

The following theorem also is very useful for determining whether a function is convex, by allowing the problem to be reduced to that of determining convexity for several simpler functions.

Theorem

1. If $f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x})$ are convex functions defined on a convex set $C \subseteq \mathbb{R}^n$, then

$$f(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x}) + \cdots + f_k(\mathbf{x})$$

   is convex on C. Furthermore, if at least one of the functions $f_i(\mathbf{x})$, $i = 1, 2, \ldots, k$, is strictly convex on C, then $f(\mathbf{x})$ is strictly convex on C.

2. If $f(\mathbf{x})$ is convex on a convex set $C \subseteq \mathbb{R}^n$, and if $\alpha > 0$, then $\alpha f(\mathbf{x})$ is convex on C.

3. If $f(\mathbf{x})$ is strictly convex on a convex set $C \subseteq \mathbb{R}^n$, and if $\alpha > 0$, then $\alpha f(\mathbf{x})$ is strictly convex on C.

4. If $f(\mathbf{x})$ is convex on a convex set $C \subseteq \mathbb{R}^n$, and if $g(y)$ is an increasing convex function defined on the range of $f(\mathbf{x}$, then the composition $g(f(\mathbf{x}))$ is convex on C.

5. If $f(\mathbf{x})$ is strictly convex on a convex set $C \subseteq \mathbb{R}^n$, and if $g(y)$ is a strictly increasing convex function defined on the range of $f(\mathbf{x}$, then the composition $g(f(\mathbf{x}))$ is strictly convex on C.

Example Let $f(x, y, z) = e^{x^2 + y^2 + z^2}$. This function is strictly convex on $\mathbb{R}^3$, as it is a composition of a strictly increasing convex function $g(y) = e^y$ with a function $h(x, y, z) = x^2 + y^2 + z^2$ that has a positive definite Hessian $Hh(x, y, z) = 2I$, and is therefore strictly convex on $\mathbb{R}^3$. □

Example Let $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(\mathbf{x}) = \sum_{i=1}^{k} c_i e^{\mathbf{a}^{(i)} \cdot \mathbf{x}},$$

where $c_1, c_2, \ldots, c_k > 0$ and $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(k)} \in \mathbb{R}^n$. This is because $f(\mathbf{x})$ is a positive linear combination of compositions of an increasing function $g(y) = e^y$ and a linear function, which is convex. □

4

Example Let
$$f(x, y) = x^2 - 4xy + 5y^2 - \ln(xy), \quad x, y > 0.$$
Then $f(x, y)$ is strictly convex on the first quadrant
$$Q_1 = \{(x, y) \in \mathbb{R}^2 \mid x, y > 0\},$$
because it is a positive linear combination of the strictly convex function
$$f_1(x, y) = x^2 - 4xy + 5y^2,$$
for which
$$Hf_1(x, y) = \begin{bmatrix} 2 & -4 \\ -4 & 10 \end{bmatrix}$$
is positive definite, in view of its minors $\Delta_1 = 2$ and $\Delta_2 = 4$, and the strictly convex function
$$f_2(x, y) = -\ln(xy) = -lnx - lny,$$
for which
$$Hf_2(x, y) = \begin{bmatrix} x^{-2} & 0 \\ 0 & y^{-2} \end{bmatrix}$$
is also positive definite on Q_1. It should be noted that it is only necessary for one of $f_1(x, y)$ and $f_2(x, y)$ to be *strictly* convex, for $f(x, y)$ to be strictly convex, as long as both functions are at least convex. $\square$

Exercises

1. Chapter 2, Exercise 1ac

2. Chapter 2, Exercise 2ad

3. Chapter 2, Exercise 3

4. Chapter 2, Exercise 7

5. Chapter 2, Exercise 8 (bonus)

6. Chapter 2, Exercise 12