Optimization Techniques
Lecture 2
Hale Gonce Köçken

Gradient, Hessian, Convexity of functions, Definiteness

Gradient

- When f is real-valued (i.e.f:$R^n \rightarrow R$) the derivative Df(x) is a 1xn matrix, i.e. it is a row vector. Its transpose is called the gradient of the function:

$$\nabla f(x) = Df(x)^T$$

- which is a column vector, i.e. in R^n. Its components are the partial derivatives of f:

$$\nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}, i = 1,2,\dots,n$$

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right]^T$$

The direction of ∇f is the orientation in which the directional derivative has the largest value and $|\nabla f|$ is the value of that directional derivative.

Hessian Matrix -Second Derivative

Given $f : \mathbb{R}^n \to \mathbb{R}$, if ∇f is differentiable, we say that f is *twice differentiable*, and we write the derivative of ∇f as

$$D^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

The matrix $D^2 f(\boldsymbol{x})$ is called the *Hessian* matrix of f at $\boldsymbol{x}$, and is often also denoted $\boldsymbol{F}(\boldsymbol{x})$.

Find the Hessian Matrix of the function $f(x,y) = x^2y + xy^3$ **.**

We need to first find the first partial derivatives of f. We have that:

$$\frac{\partial f}{\partial x} = 2xy + y^3 \quad , \quad \frac{\partial f}{\partial y} = x^2 + 3xy^2$$

We then calculate the second partial derivatives of f:

$$\frac{\partial^2 f}{\partial x^2} = 2y \quad , \quad \frac{\partial^2 f}{\partial y \partial x} 2x + 3y^2 \quad , \quad \frac{\partial^2 f}{\partial x \partial y} = 2x + 3y^2 \quad , \quad \frac{\partial^2 f}{\partial y^2} = 6xy$$

Therefore the Hessian Matrix of f is:

$$\mathcal{H}(x,y) = \begin{bmatrix} 2y & 2x + 3y^2 \\ 2x + 3y^2 & 6xy \end{bmatrix}$$

Example 6.1 Let $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$.

Theorem 21.2 *A function $f : \Omega \to \mathbb{R}$ defined on a convex set $\Omega \subset \mathbb{R}^n$ is convex if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ and all $\alpha \in (0,1)$, we have*

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}).$$

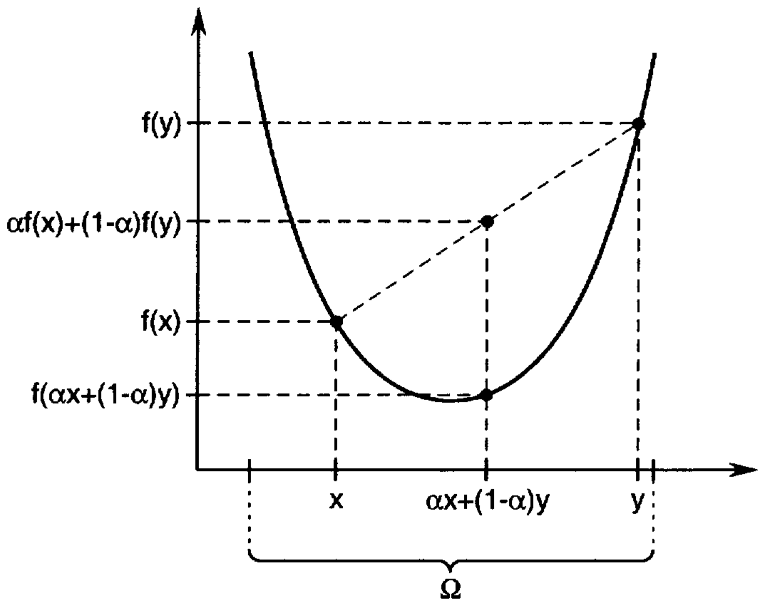


Figure 21.5 Geometric interpretation of Theorem 21.2

The line segment between any two points on the graph lies on or above the graph

Definition 21.4 A function $f : \Omega \to \mathbb{R}$ on a convex set $\Omega \subset \mathbb{R}^n$ is *strictly convex* if for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, $\boldsymbol{x} \neq \boldsymbol{y}$, and $\alpha \in (0,1)$, we have

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) < \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}).$$

Definition 21.5 A function $f : \Omega \to \mathbb{R}$ on a convex set $\Omega \subset \mathbb{R}^n$ is (strictly) *concave* if $-f$ is (strictly) convex. ■

Definition 21.1 The *graph* of $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, is the set of points in $\Omega \times \mathbb{R} \subset \mathbb{R}^{n+1}$ given by

$$\{[\boldsymbol{x}, f(\boldsymbol{x})]^T : \boldsymbol{x} \in \Omega\}.$$

Definition 21.2 The *epigraph* of a function $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, denoted epi$(f)$, is the set of points in $\Omega \times \mathbb{R}$ given by

$$\text{epi}(f) = \{[\boldsymbol{x}, \beta]^T : \boldsymbol{x} \in \Omega, \beta \in \mathbb{R}, \beta \geq f(\boldsymbol{x})\}.$$

The epigraph epi(f) of a function f is simply the set of points in $\Omega \times \mathbb{R}$ on or above the graph of f (see Figure 21.4). We can also think of epi(f) as a subset of $\mathbb{R}^{n+1}$.

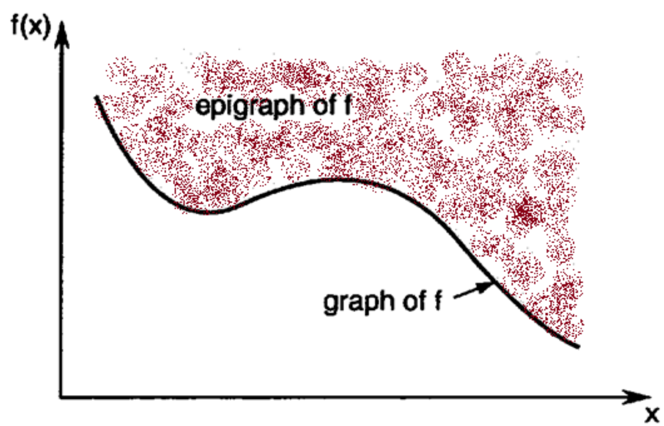


Figure 21.4 The graph and epigraph of a function $f : \mathbb{R} \to \mathbb{R}$

The hypograph of a function $f: \Omega \to R$, denoted *hyp(f)*, is the set of points in $\Omega\, x\, R$ given by

$$hyp(f) = \{[x, \beta]^T:\ x \in \Omega, \beta \in R, \beta \leq f(x)\}$$

Theorem (Characterization the convexity of a function in terms of graphs)

(a) A function f defined on a convex set is concave if and only if its hypograpf hyp f is convex.

(b) A function f defined on a convex set S is convex if and only if its epigraph epi f is convex.

Theorem 21.4 *Let $f : \Omega \to \mathbb{R}$, $f \in C^2$, be defined on an open convex set $\Omega \subset \mathbb{R}^n$. Then, f is convex on Ω if and only if for each $\boldsymbol{x} \in \Omega$, the Hessian $\boldsymbol{F}(\boldsymbol{x})$ of f at $\boldsymbol{x}$ is a positive semidefinite matrix.* □

DEFINITENESS of a MATRIX

- A matrix $H \in R^{nxn}$ (H is a real symmetric *nxn* matrix) is said to be positive definite if for all $h \neq 0$, $h^T H\,h > 0$.

- **The First Method (Definition)**

$$\forall h \in R^n, h \neq 0$$

$i)$ $h^T H\,h \geq 0 \;\Rightarrow H$ is positive semidefinite

$h^T H\,h > 0 \;\Rightarrow H$ is positive definite

$ii)$ $h^T H\,h \leq 0 \;\Rightarrow H$ is negative semidefinite

$h^T H\,h < 0 \;\Rightarrow H$ is negative definite

$iii)$ Otherwise, H is indefinite matrix

The Second Method (Eigenvalues)

- Let the eigenvalues of *H* be $\lambda_1, \lambda_2, \ldots, \lambda_n$.

i) If all $\lambda_i \geq 0$; i=1,2,...,n, then *H* is positive semidefinite

If all $\lambda_i > 0$; i=1,2,...,n, then *H* is positive definite

ii) If all $\lambda_i \leq 0$; i=1,2,...,n, then *H* is negative semidefinite

If all $\lambda_i < 0$; i=1,2,...,n, then *H* is negative definite

iii) Otherwise, *H* is indefinite matrix

The Third Method (Principal Leading Minors)**

A minor of the matrix A of order k is **principal** if it is obtained by deleting $n-k$ rows and the $n-k$ columns with the same numbers.

Note that the definition does not specify which $n-k$ rows and columns to delete, only that their indices must be the same.

The **leading principal minor** of A of order k is the minor of order k obtained by deleting the **last** $n-k$ rows and the **last** $n-k$ columns.

Example 3 *For a general* 3×3 *matrix,*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

All principal minors:

There is one third order principal minor, namely $|A|$.

There are three second order principal minors:

$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, *formed by deleting column 3 and row 3;*

$\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$, *formed by deleting column 2 and row 2;*

$\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, *formed by deleting column 1 and row 1*

And there are three first order principal minors:

$|a_{11}|$, *formed by deleting the last two rows and columns*

$|a_{22}|$, *formed by deleting the first and third rows and columns*

$|a_{33}|$, *formed by deleting the first two rows and columns*

The Leading principal minors:

The first order principal minor:

The second order principal minor:

The third order principal minor:

The algorithm for testing the definiteness of a symmetric matrix:

Let A *be* a symmetric $n \times n$ matrix, denote

its leading principal minors by Δ_i for $i \in \{1, 2, \ldots, n\}$ and

its principal minors by D_i for each order $i \in \{1, 2, \ldots, n\}$.

Remark: The matrix A has $\binom{n}{i}$ principal minors of order i.

The algorithm for testing the definiteness of a symmetric matrix:

1. A is **positive definite** if and only if all its n leading principal minors are positive, that is $\Delta_i > 0$ for all $i \in \{1, 2, \ldots, n\}$.
2. A is **negative definite** if and only if its n leading principal minors alternate in sign beginning by negative: $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \Delta_4 > 0, \ldots$
3. A is **positive semidefinite** if and only if all its principal minors are nonnegative, that is all $D_i \geq 0$ for all $i \in \{1, 2, \ldots, n\}$ considering all principal minors.
4. A is **negative semidefinite** if and only if every principal minors of odd order is nonpositive (≤ 0) and every principal minors of even order is nonnegative (≥ 0), that is $(-1)^i D_i \geq 0$ for all $i \in \{1, 2, \ldots, n\}$ considering all principal minors.
5. Otherwise A is **indefinite**.

Note:

- In the first two cases, it is enough to check the inequality for all the leading principal minors.
- If some $i-$th order leading principal minor of A is nonzero but does not fit either of the sign patterns in case 1 and case 2, then A is indefinite.
- When some leading principal minor of A is zero, but the others fit one of the patterns in case 1 and case 2: the matrix is not definite, but may or may not be semidefinite. In this case, we must unfortunately check not only the principle leading minors, but ***every*** principal minor (cases 3 and 4: we must check all principles minors that is for each i with $1 \le i \le n$ and for each of the $\binom{n}{i}$ principal minors of order i.)

Example: Determine the definiteness of the following matrix with three methods

$$H = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Example: Determine the definiteness of the following matrices

$$A=\begin{bmatrix}6 & 4\\ 4 & 5\end{bmatrix},\ B=\begin{bmatrix}-3 & 0 & 0\\ 0 & -2 & 0\\ 0 & 0 & -1\end{bmatrix}, C=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1/2\end{bmatrix}$$

We have that $D_1 = 6 > 0$, and $D_2 = \begin{vmatrix}6 & 4\\ 4 & 5\end{vmatrix} = 30 - 16 = 14 > 0$.

Therefore, A is a positive definite matrix.

We have that $D_1 = -3 < 0$ and $D_2 = \begin{vmatrix}-3 & 0\\ 0 & -2\end{vmatrix} = 6 > 0$. The matrix A is known as a diagonal matrix,

and the determinant $D_3 = \begin{vmatrix}-3 & 0 & 0\\ 0 & -2 & 0\\ 0 & 0 & -1\end{vmatrix}$

can be computed as the product of the entries in the main diagonal, that is $D_3 = (-3)(-2)(-1) = -6 < 0$.

Since $D_1, D_3 < 0$ and $D_2 > 0$, we have that A is a negative definite matrix.

$$C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & \frac{1}{2} \end{bmatrix}$$

Theorem 21.4 *Let $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, be defined on an open convex set $\Omega \subset \mathbb{R}^n$. Then, f is convex on Ω if and only if for each $\boldsymbol{x} \in \Omega$, the Hessian $\boldsymbol{F}(\boldsymbol{x})$ of f at $\boldsymbol{x}$ is a positive semidefinite matrix.* □

Note that by definition of concavity, a function $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, is concave over the convex set $\Omega \subset \mathbb{R}^n$ if and only if for all $\boldsymbol{x} \in \Omega$, the Hessian $\boldsymbol{F}(\boldsymbol{x})$ of f is negative semidefinite.

positive semidefinite – convex
positive definite – strictly convex
negative semidefinite – concave
negative definite – strictly concave

Some exercises on convexity of sets and functions

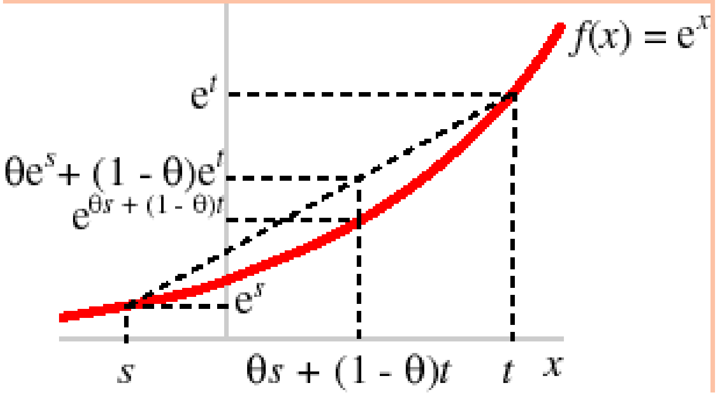
1. By drawing diagrams, determine which of the following sets is convex.

a. $\{(x, y): y = e^x\}$.

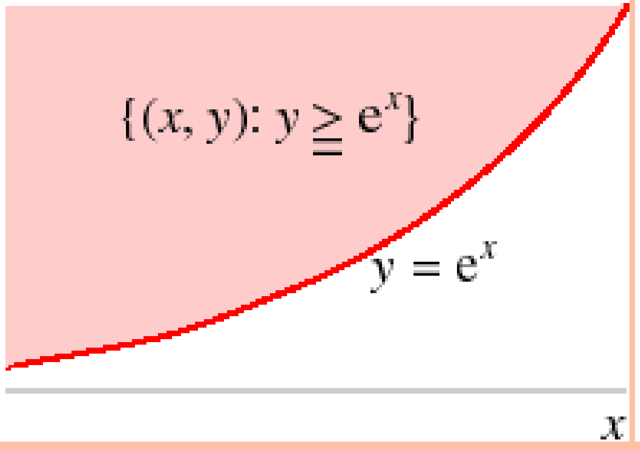
b. $\{(x, y): y \geq e^x\}$.

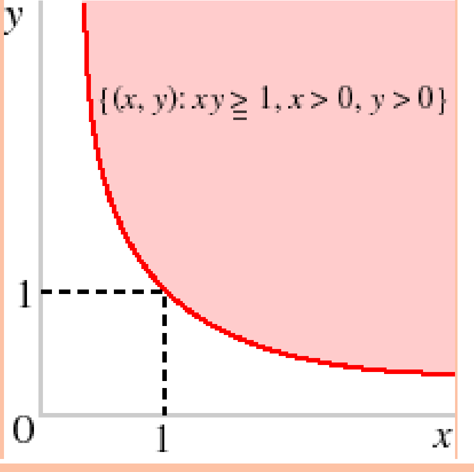
c. $\{(x, y): xy \geq 1, x > 0, y > 0\}$.

a. Not convex, because $e^{\theta x+(1-\theta)u} \neq \theta e^x + (1-\theta)e^u$, as illustrated in the following figure.



b. Convex, because $e^{\theta x+(1-\theta)u} < \theta e^x + (1-\theta)e^u$ (see the following figure).





c. Convex, because if $xy \geq 1$ and $uv \geq 1$ then $(\theta x + (1-\theta)u)(\theta y + (1-\theta)v) \geq 1$ (see figure).

2. For each of the following functions, determine which, if any, of the following conditions the function satisfies: concavity, strict concavity, convexity, strict convexity. (Use whatever technique is most appropriate for each case.)

a. $f(x, y) = x + y$.

b. $f(x, y) = x^2$. [Note: f is a function of two variables.]

c. $f(x, y) = x + y - e^x - e^{x+y}$.

d. $f(x, y, z) = x^2 + y^2 + 3z^2 - xy + 2xz + yz$.

a. $f((1-\lambda)x + \lambda x',(1-\lambda)y + \lambda y') = (1-\lambda)x + \lambda x' + (1-\lambda)y + \lambda y' = (1-\lambda)f(x,y) + \lambda f(x',y')$, so the function is concave and convex, but not strictly concave or strictly convex. Or you can calculate the Hessian, from which you can conclude that the function is both concave and convex, and then argue as above that the function is not strictly concave or strictly convex. [Note: the fact that some of the minors are zero does not imply that the function is not strictly concave or strictly convex, although in fact it is not.] Or you can appeal to the fact that the function is linear to conclude that it is concave and convex.

b. The Hessian shows that the function is convex (all principal minors are nonnegative). The Hessian does not satisfy the sufficient condition for strict convexity, but this does not imply that the function is in fact not strictly convex. However, since, for example, $f(1, 1) = f(1, 2) = f(1, 3)$, the function is in fact not strictly convex. (More generally, for all x, y, and y' we have $f((1-\lambda)(x,y) + \lambda(x,y')) = f(x,(1-\lambda)y + \lambda y') = x^2 = (1-\lambda)f(x,y) + \lambda f(x,y')$.)

c. The Hessian is

$$\begin{pmatrix} -e^x - e^{x+y} & -e^{x+y} \\ -e^{x+y} & -e^{x+y} \end{pmatrix}$$

Since $-e^x - e^{x+y} < 0$ for all (x,y) and the determinant is $(-e^x - e^{x+y})(-e^{x+y}) - (-e^{x+y})(-e^{x+y}) = e^{2x+y} > 0$ for all (x,y) the function is strictly concave. (Or you can argue that since e^u is increasing and convex and $x + y$ is convex, e^{x+y} is convex and thus $-e^{x+y}$ is concave, and similarly for $-e^x$; then you need to make a separate argument for strict concavity.)

d. The Hessian is

$$\begin{pmatrix} 2 & -1 & 2 \\ -1 & 2 & 1 \\ 2 & 1 & 6 \end{pmatrix}$$

the leading principal minors of which are $2 > 0$, $3 > 0$, and $4 > 0$, so that the function is strictly convex.

3. Let $f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2 + 3x_1 - 2x_2 + 1$. Is f convex, concave, or neither?

4. Let $f(x_1, x_2) = x_1^3 + 2x_1^2 + 2x_1x_2 + (1/2)x_2^2 - 8x_1 - 2x_2 - 8$. Find the range of values of (x_1, x_2) for which f is convex, if any.

The Hessian matrix of f is

$$\begin{pmatrix} 6x_1 + 4 & 2 \\ 2 & 1 \end{pmatrix}.$$

This matrix is positive semidefinite if $6x_1 + 4 \geq 0$ and $6x_1 \geq 0$, or if $x_1 \geq 0$. Thus f is convex for $x_1 \geq 0$ (and all x_2).

5. Determine the values of a (if any) for which the function $2x^2 + 2xz + 2ayz + 2z^2$ is concave and the values for which it is convex.

The Hessian of the function is

$$\begin{pmatrix} 4 & 0 & 2 \\ 0 & 0 & 2a \\ 2 & 2a & 4 \end{pmatrix}.$$

The first-order minors are 4, 0, and 4, the second-order minors are 0, 12, and $-4a^2$, and the determinant is $-16a^2$. Thus for $a = 0$ the Hessian is positive semidefinite, so that the function is convex; for other values of a the Hessian is indefinite, so that the function is neither concave nor convex.

6. Show that the function $-w^2 + 2wx - x^2 - y^2 + 4yz - z^2$ (in the four variables w, x, y, and z) is not concave.

The Hessian is

$$\begin{pmatrix} -2 & 2 & 0 & 0 \\ 2 & -2 & 0 & 0 \\ 0 & 0 & -2 & 4 \\ 0 & 0 & 4 & -2 \end{pmatrix}$$

The second order principal minor obtained by deleting the first and second rows and columns is -12, so the matrix is not negative semidefinite. Thus the function is not concave.