Curvature, Convexity, and Separation

Lecture 5

Maria Titova

August 10, 2026

Econ 8000 — Math Camp

Motivation

Lecture 4 used a linear map to approximate a differentiable function near a point.

This lecture is about curvature, which a second-order approximation describes.

If f is twice continuously differentiable near x_0, we add a quadratic term:

$$f(x_0 + h) \approx f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2.$$

When $f'(x_0) = 0$, we cannot say whether the graph bends up or down.

- The sign of $f''(x_0)$ does.
- In several variables, the Hessian H_f generalizes $f''(x_0)$.

Second-order approximation and curvature

C^2 functions

Definition 1.1 (C^2 function)

Let $U \subseteq \mathbb{R}^n$ be open.

A function $f : U \to \mathbb{R}$ is C^2 *on* U if each partial derivative $\partial f / \partial x_i$ is C^1 on U.

We also say that f is *twice continuously differentiable on* U.

Equivalently, all second partial derivatives of f exist and are continuous on U.

Hessian

Definition 1.2 (Hessian)

Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be differentiable.

If $\nabla f : U \to \mathbb{R}^n$ is differentiable at $x \in U$, the *Hessian* of f at x is

$$H_f(x) = \mathrm{D}(\nabla f)(x),$$

the Jacobian of its gradient.

It is the $n \times n$ matrix whose (i, j) entry is

$$\big(H_f(x)\big)_{ij} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x).$$

Computing a Hessian

Example 1.3 (Computing a Hessian)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x_1, x_2) = x_1^2 x_2.$$

Compute $H_f(x)$.

Computing a Hessian

Example 1.3 (Computing a Hessian)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x_1, x_2) = x_1^2 x_2.$$

Compute $H_f(x)$.

Solution.

In Lecture 4 we computed ∇f.

The Hessian is the Jacobian of the gradient, so

$$\nabla f(x) = \begin{pmatrix} 2x_1x_2 \\ x_1^2 \end{pmatrix}, \qquad H_f(x) = \begin{pmatrix} 2x_2 & 2x_1 \\ 2x_1 & 0 \end{pmatrix}.$$

□

Equality of mixed partials

Theorem 1.4 (Equality of mixed partials)

Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be C^2. Then

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(x)$$

for every $x \in U$ and every i, j.

Thus $H_f(x)$ is symmetric.

- The result requires continuity of the second partial derivatives.

Second-order Taylor expansion

Theorem 1.5 (Second-order Taylor expansion)

Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \to \mathbb{R}$ be C^2, and let $x_0 \in U$.

For changes h such that $x_0 + th \in U$ for every $t \in [0,1]$,

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2} h^\top H_f(x_0) h + r(h),$$

where

$$\frac{|r(h)|}{\|h\|^2} \longrightarrow 0 \quad \text{as } h \to 0.$$

Quadratic approximation

This has the form of the linear approximation in Lecture 4, with one more term.

The expression

$$f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2} h^\top H_f(x_0) h$$

is the *quadratic approximation* to $f(x_0 + h)$.

The first-order approximation stops after the linear term.

- Its remainder is negligible relative to $\|h\|$.

The second-order approximation keeps the quadratic term.

- Its remainder is negligible relative to $\|h\|^2$.

Quadratic forms

The quadratic term of the Taylor approximation has the form $x^\top Ax$.

- Expressions of this form are essentially quadratic polynomials in several variables.
- Like quadratic functions in one dimension, they have a constant Hessian.

Definition 1.6 (Quadratic form)

Let A be a symmetric $n \times n$ matrix. The function

$$Q_A(x) = x^\top Ax, \qquad x \in \mathbb{R}^n,$$

is the *quadratic form* associated with A.

The Hessian of a quadratic form

The gradient and Hessian of $Q_A(x)$ are:

$$\nabla Q_A(x) = 2Ax, \qquad H_{Q_A}(x) = 2A.$$

If you don't believe this, write $Q_A(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$ and use $a_{ij} = a_{ji}$:

$$\frac{\partial Q_A}{\partial x_k}(x) = 2 \sum_{j=1}^n a_{kj} x_j, \qquad \frac{\partial}{\partial x_l}\left(\frac{\partial Q_A}{\partial x_k}\right)(x) = 2a_{kl}.$$

Definiteness

The sign of $x^\top Ax$ describes the possible shapes of the curvature in $\mathbb{R}^n$.

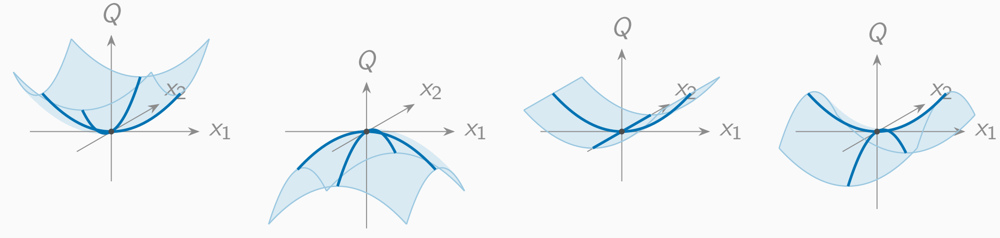
Definition 1.7 (Definiteness)

Let A be a symmetric $n \times n$ matrix.

$$\begin{aligned}
A \text{ is } \textit{positive definite} \quad &\iff \quad x^\top Ax > 0 && \text{for every } x \neq 0,\\
A \text{ is } \textit{positive semidefinite} \quad &\iff \quad x^\top Ax \geq 0 && \text{for every } x,\\
A \text{ is } \textit{negative definite} \quad &\iff \quad x^\top Ax < 0 && \text{for every } x \neq 0,\\
A \text{ is } \textit{negative semidefinite} \quad &\iff \quad x^\top Ax \leq 0 && \text{for every } x.
\end{aligned}$$

The matrix is *indefinite* if its quadratic form takes both positive and negative values.

Four quadratic forms in two variables



(a) positive definite: $x_1^2 + x_2^2$

(b) negative definite: $-x_1^2 - x_2^2$

(c) positive semidefinite: x_1^2

(d) indefinite: $x_1^2 - x_2^2$

Figure 1.1

- The thick curves show Q along the x_1 axis and along the x_2 axis

Definiteness describes every slice

Positive definite means that every nonzero-direction slice bends upward.

- So the origin is the unique global minimum.

Negative definite means that every such slice bends downward.

- So the origin is the unique global maximum.

A positive semidefinite form may be flat in some directions.

An indefinite form bends upward in some directions and downward in others.

- The origin is then neither a maximum nor a minimum.

Two tests for definiteness

Checking definiteness from the definition requires evaluating $x^\top Ax$ at every $x \neq 0$.

- That is not very practical.

Instead, we use two tests that are easier to apply.

- The first uses eigenvalues.
- The second uses determinants of submatrices.

The diagonal case

For a diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

we can write:

$$x^\top D x = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2,$$

so definiteness is determined by the signs of the diagonal entries.

Diagonalization and the spectral theorem

Recall from Lecture 3 that a matrix A with n independent eigenvectors factors as

$$A = PDP^{-1},$$

where the columns of P are the eigenvectors and D carries the corresponding eigenvalues.

The *spectral theorem* strengthens this result for symmetric matrices:

- They have n independent eigenvectors.
- The eigenvector matrix can be chosen with $P^{-1} = P^\top$, so
$$A = PDP^\top.$$
- Thus, with $z = P^\top x$, the quadratic form reduces to the diagonal case:
$$x^\top Ax = z^\top Dz.$$

Eigenvalue test

Theorem 1.8 (Eigenvalue test)

Let A be a symmetric $n \times n$ matrix. Then

$$\begin{aligned} A \text{ is positive definite} &\iff \text{every eigenvalue of } A \text{ is positive,} \\ A \text{ is positive semidefinite} &\iff \text{every eigenvalue of } A \text{ is nonnegative,} \\ A \text{ is negative definite} &\iff \text{every eigenvalue of } A \text{ is negative,} \\ A \text{ is negative semidefinite} &\iff \text{every eigenvalue of } A \text{ is nonpositive.} \end{aligned}$$

The matrix is indefinite if and only if it has eigenvalues of both signs.

Principal minors

For each nonempty $I \subseteq \{1, \ldots, n\}$, let A_I retain the rows and columns indexed by I. Its determinant $\det A_I$ is a *principal minor* of order $|I|$.

For example, when $I = \{1, 3\}$,

$$\det A_{\{1,3\}} = \det \begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix} = a_{11}a_{33} - a_{13}a_{31}.$$

The *leading principal minors* use $I = \{1, \ldots, k\}$ and grow from the top-left corner. For a 3×3 matrix,

$$\Delta_1 = a_{11}, \qquad \Delta_2 = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad \Delta_3 = \det A.$$

An $n \times n$ matrix has $2^n - 1$ principal minors, of which n are leading.

Principal-minor tests

Theorem 1.10 (Principal-minor tests)

Let A be symmetric $n \times n$ with leading principal minors $\Delta_1, \ldots, \Delta_n$.

1. *A is positive definite if and only if $\Delta_k > 0$ for every $k = 1, \ldots, n$.*
2. *A is negative definite if and only if $(-1)^k \Delta_k > 0$ for every $k = 1, \ldots, n$.*
3. *A is positive semidefinite if and only if every principal minor is nonnegative.*
4. *A is negative semidefinite if and only if every principal minor of order k satisfies*
$$(-1)^k \det A_I \geq 0.$$

Leading minors are not always enough

The first two parts are commonly called *Sylvester's criterion*.

Notice the difference between the definite and semidefinite cases.

- Definite matrices require only the n leading principal minors.
- Semidefinite matrices require all $2^n - 1$ nonempty principal minors.

The 2×2 tests

Example 1.11 (The 2×2 tests)

Classify the definiteness of $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ in terms of its principal minors.

Example 1.11 (The 2×2 tests)

Classify the definiteness of $A = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ in terms of its principal minors.

Solution.

Its leading principal minors are a and $ac - b^2$.

All of its principal minors are a, c, and $ac - b^2$. Therefore

$$\begin{aligned}
A \text{ is positive definite} &\iff a > 0 \text{ and } ac - b^2 > 0, \\
A \text{ is negative definite} &\iff a < 0 \text{ and } ac - b^2 > 0, \\
A \text{ is positive semidefinite} &\iff a \geq 0,\ c \geq 0,\ \text{and } ac - b^2 \geq 0, \\
A \text{ is negative semidefinite} &\iff a \leq 0,\ c \leq 0,\ \text{and } ac - b^2 \geq 0.
\end{aligned}$$

$\square$

Weak leading-minor inequalities are insufficient

Example 1.12 (Weak leading-minor inequalities are insufficient)

Let $A = \operatorname{diag}(0, -1)$, whose leading minors satisfy $\Delta_1 = \Delta_2 = 0$.

Do these weak inequalities establish positive semidefiniteness?

Weak leading-minor inequalities are insufficient

Example 1.12 (Weak leading-minor inequalities are insufficient)

Let $A = \mathrm{diag}(0, -1)$, whose leading minors satisfy $\Delta_1 = \Delta_2 = 0$.

Do these weak inequalities establish positive semidefiniteness?

Solution.

No: the principal minor indexed by $I = \{2\}$ is -1.

So the leading minors alone do not establish positive semidefiniteness.

The missing principal minor detects the negative direction e_2.

- Along it, $e_2^\top A e_2 = -1$.

$\square$

A three-variable quadratic form

Example 1.13 (A three-variable quadratic form)

Let $Q : \mathbb{R}^3 \to \mathbb{R}$ be given by $Q(x) = -x_1^2 - x_2^2 - x_3^2 + x_1x_2 + x_2x_3$.

Determine the definiteness of its Hessian.

A three-variable quadratic form

Example 1.13 (A three-variable quadratic form)

Let $Q : \mathbb{R}^3 \to \mathbb{R}$ be given by $Q(x) = -x_1^2 - x_2^2 - x_3^2 + x_1x_2 + x_2x_3$.

Determine the definiteness of its Hessian.

Solution.

Its Hessian and leading principal minors are

$$H_Q = \begin{pmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}, \qquad \Delta_1 = -2, \quad \Delta_2 = 3, \quad \Delta_3 = -4.$$

Thus $(-1)^k \Delta_k$ equals 2, 3, and 4, respectively.

By the principal-minor test, H_Q is negative definite. $\square$

An interaction term

Example 1.15 (An interaction term)

Fix $b \in \mathbb{R}^2$ and $\gamma \in \mathbb{R}$, and define $q : \mathbb{R}^2 \to \mathbb{R}$ by

$$q(x) = b \cdot x - \left(x_1^2 + \gamma x_1 x_2 + x_2^2\right).$$

Classify the definiteness of H_q in terms of γ.

Example 1.15 (An interaction term)

Fix $b \in \mathbb{R}^2$ and $\gamma \in \mathbb{R}$, and define $q : \mathbb{R}^2 \to \mathbb{R}$ by

$$q(x) = b \cdot x - \left(x_1^2 + \gamma x_1 x_2 + x_2^2\right).$$

Classify the definiteness of H_q in terms of γ.

Solution.

Here H_q is constant, and direct differentiation gives

$$H_q = \begin{pmatrix} -2 & -\gamma \\ -\gamma & -2 \end{pmatrix}.$$

The leading principal minors are -2 and $4 - \gamma^2$. $\square$

The three cases

Example 1.15 (An interaction term)

H_q has leading principal minors -2 and $4 - \gamma^2$.

Example 1.15 (An interaction term)

H_q has leading principal minors -2 and $4 - \gamma^2$.

Solution.

By the principal-minor test, H_q is negative definite when $|\gamma| < 2$.

When $|\gamma| = 2$, both order-one principal minors are negative and $\det H_q = 0$.

- So H_q is negative semidefinite.

When $|\gamma| > 2$, the determinant is negative, so H_q is indefinite.

- Both eigenvalues are real, and their negative product forces opposite signs.

The interaction term can preserve, flatten, or overturn the downward curvature. □

Cheaper checks

Two exact tests are now available: the eigenvalue signs and the principal-minor signs.

Often a full classification is unnecessary, and a cheaper observation settles the question.

Quick checks

Check	Verdict
$a_{ii} \leq 0$ for every i	necessary for negative semidefiniteness, not sufficient
$a_{ii} < 0$ for every i	necessary for negative definiteness, not sufficient
$(-1)^k \Delta_k \geq 0$ for every k	necessary for negative semidefiniteness, not sufficient
$(-1)^k \Delta_k > 0$ for every k	equivalent to negative definiteness
some x has $x^\top Ax > 0$	counterexample to negative semidefiniteness
some $x \neq 0$ has $x^\top Ax = 0$	counterexample to negative definiteness
negative semidefinite and $\det A \neq 0$	sufficient for negative definiteness

- A is symmetric with entries a_{ij} and leading principal minors Δ_k.

Convex sets and concave functions

Definition 2.1 (Convex combination and convex set)

Let $x, y \in \mathbb{R}^n$.

- A *convex combination* of x and y is a point $(1-t)x + ty$ with $t \in [0, 1]$.
  - The set of all of them is the *line segment* joining x and y.
- A set $C \subseteq \mathbb{R}^n$ is *convex* if

$$(1-t)x + ty \in C \qquad \text{for every } x, y \in C \text{ and every } t \in [0, 1].$$

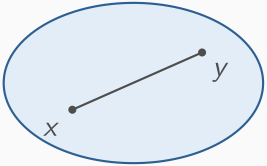
Convex and nonconvex sets

A set is convex exactly when the segments joining its points stay inside it.

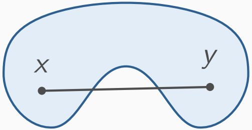
Examples:

- Every affine subspace and every halfspace $\{x : a \cdot x \leq c\}$ is convex.
- $\mathbb{R}^n$ itself and the orthants $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ are convex.
- The unit circle $\{x \in \mathbb{R}^2 : \|x\| = 1\}$ is not convex: it contains e_1 and $-e_1$ but not their midpoint 0.

Examples in $\mathbb{R}^2$



(a) convex



(b) not convex

Figure 2.1

Preservation rules

Theorem 2.2 (Preservation rules)

The following constructions preserve convexity.

1. *The intersection of any collection of convex sets is convex.*
2. *The Cartesian product of convex sets is convex.*
3. *The image of a convex set under an affine map is convex.*
4. *The inverse image of a convex set under an affine map is convex.*

Here, an *affine map* $T : \mathbb{R}^n \to \mathbb{R}^m$ has the form $T(x) = Ax + b$, where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$.

Linear restrictions define a convex set

Any finite list of linear equalities and inequalities defines a convex set.

Example 2.3 (Linear restrictions define a convex set)

Let A be $m \times n$, let B be $p \times n$, and let $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^p$.

Show that $C = \{x \in \mathbb{R}^n : Ax \leq b,\ Bx = c\}$ is convex, with $Ax \leq b$ coordinatewise.

Linear restrictions define a convex set

Any finite list of linear equalities and inequalities defines a convex set.

Example 2.3 (Linear restrictions define a convex set)

Let A be $m \times n$, let B be $p \times n$, and let $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^p$.

Show that $C = \{x \in \mathbb{R}^n : Ax \leq b,\ Bx = c\}$ is convex, with $Ax \leq b$ coordinatewise.

Solution.

Each row has the form $a \cdot x = d$ or $a \cdot x \leq d$ for some $a \in \mathbb{R}^n$ and $d \in \mathbb{R}$.

The set it defines is the inverse image of $\{d\}$ or $(-\infty, d]$ under $x \mapsto a \cdot x$.

The inverse-image rule makes each restriction convex.

The intersection rule then makes C convex. $\square$

Next we define convexity and concavity for functions.

Definition 2.4 (Concave and convex functions)

Let $C \subseteq \mathbb{R}^n$ be convex and let $f : C \to \mathbb{R}$.

- f is *concave* if
$$f\big((1-t)x + ty\big) \geq (1-t)f(x) + tf(y)$$
for every $x, y \in C$ and every $t \in [0,1]$.
- f is *strictly concave* if that inequality is strict whenever $x \neq y$ and $t \in (0,1)$.
- f is *convex* if $-f$ is concave, and *strictly convex* if $-f$ is strictly concave.

Reading the chord inequality

The right-hand side is the height of the chord joining $(x, f(x))$ and $(y, f(y))$.

- The height is measured at $(1 - t)x + ty$.
- So a concave function lies on or above each of its chords.

This is the global version of bending downward.

Strict concavity additionally forbids the graph from containing a line segment.

An affine function $f(x) = a \cdot x + b$ satisfies the definition with equality.

- It is therefore both concave and convex, and neither strictly.

The chord inequality

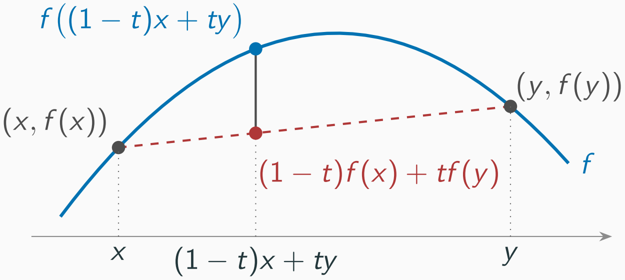


Figure 2.2

- f is strictly concave; the figure shows the inequality for a concrete pair of x and y.

Superlevel sets

Definition 2.5 (Superlevel set)

Let $C \subseteq \mathbb{R}^n$, let $f : C \to \mathbb{R}$, and let $\alpha \in \mathbb{R}$.

The *superlevel set* of f at level α is

$$S_\alpha(f) = \{x \in C : f(x) \geq \alpha\}.$$

Theorem 2.6 (Superlevel sets of a concave function)

If C is convex and $f : C \to \mathbb{R}$ is concave, then $S_\alpha(f)$ is convex for every $\alpha \in \mathbb{R}$.

A concave function lies above its chords, so if both endpoints of a segment sit above the horizontal line at height α, the whole segment does too.

Quasiconcavity

A weaker notion of concavity delivers the same convexity of superlevel sets.

Definition 2.7 (Quasiconcavity)

Let $C \subseteq \mathbb{R}^n$ be convex.

A function $f : C \to \mathbb{R}$ is *quasiconcave* if $S_\alpha(f)$ is convex for every $\alpha \in \mathbb{R}$.

Superlevel sets in one variable

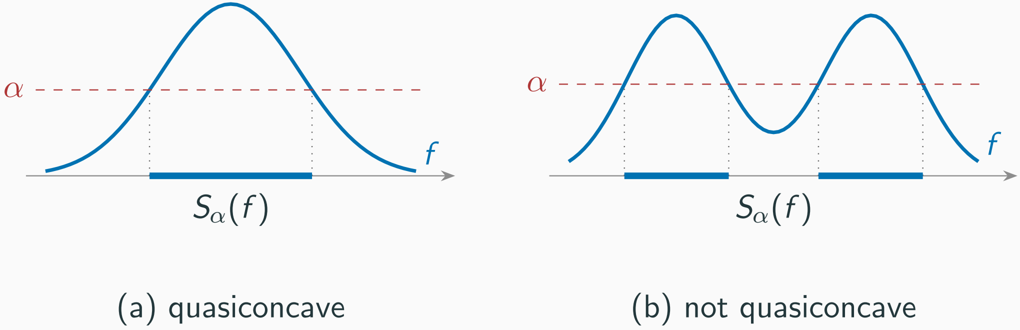


Figure 2.3

- On the right, the superlevel set is a union of two disjoint intervals, not convex.

Minimum characterization

Theorem 2.8 (Minimum characterization)

A function $f : C \to \mathbb{R}$ on a convex set $C \subseteq \mathbb{R}^n$ is quasiconcave if and only if

$$f\big((1-t)x + ty\big) \geq \min\{f(x), f(y)\}$$

for every $x, y \in C$ and every $t \in [0,1]$.

Quasiconcave but not concave

Concavity implies quasiconcavity, but the converse fails.

Example 2.9 (Quasiconcave but not concave)

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$.

Is it quasiconcave? Is it concave?

Quasiconcave but not concave

Concavity implies quasiconcavity, but the converse fails.

Example 2.9 (Quasiconcave but not concave)

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$.

Is it quasiconcave? Is it concave?

Solution.

Since f is increasing, each superlevel set is the interval $[\alpha^{1/3}, \infty)$, which is convex.

So f is quasiconcave, but it is not concave: at the midpoint of 0 and 2,

$$f(1) = 1 < 4 = \tfrac{1}{2} f(0) + \tfrac{1}{2} f(2).$$

$\square$

Increasing transformations

Quasiconcavity (unlike concavity) survives any increasing transformation.

Theorem 2.10 (Increasing transformations)

Let $C \subseteq \mathbb{R}^n$ be convex and let $f : C \to \mathbb{R}$ be quasiconcave.

If $\varphi : \mathbb{R} \to \mathbb{R}$ is nondecreasing, then $\varphi \circ f$ is quasiconcave.

Increasing transformations

Quasiconcavity (unlike concavity) survives any increasing transformation.

Theorem 2.10 (Increasing transformations)

Let $C \subseteq \mathbb{R}^n$ be convex and let $f : C \to \mathbb{R}$ be quasiconcave.

If $\varphi : \mathbb{R} \to \mathbb{R}$ is nondecreasing, then $\varphi \circ f$ is quasiconcave.

Proof.

Applying φ to the minimum inequality preserves it, so

$$\varphi\Big(f\big((1-t)x+ty\big)\Big) \geq \varphi(\min\{f(x), f(y)\}) = \min\big\{\varphi(f(x)), \varphi(f(y))\big\},$$

the equality because φ sends the smaller number to the smaller image. $\square$

Concavity criteria, global optimality, and separation

How to tell whether a function is concave

Suppose f is defined on an open convex set. We will use two tests for concavity.

1. **Hessian test:** If f is C^2, check whether $H_f(x)$ is negative semidefinite at every x.
2. **Tangent test:** If f is differentiable, check whether its graph lies below every tangent plane.

Hessian characterization

Theorem 3.5 (Hessian characterization)

Let $U \subseteq \mathbb{R}^n$ be open and convex, and let $f : U \to \mathbb{R}$ be C^2.

1. *f is concave if and only if $H_f(x)$ is negative semidefinite for every $x \in U$.*
2. *If $H_f(x)$ is negative definite for every $x \in U$, then f is strictly concave.*

The corresponding statements for convexity reverse every sign.

Negative definiteness is sufficient but not necessary for strict concavity.

- For example, $f(x) = -x^4$ is strictly concave on $\mathbb{R}$, but $f''(0) = 0$.

Tangent characterization

Theorem 3.4 (Tangent characterization)

Let $U \subseteq \mathbb{R}^n$ be open and convex, and let $f : U \to \mathbb{R}$ be differentiable.

Then f is concave if and only if

$$f(y) \leq f(x) + \nabla f(x) \cdot (y - x) \qquad \text{for every } x, y \in U.$$

- At x, the right-hand side is the tangent plane.
- Concavity requires $f(y)$ to lie below that plane for every y.

A concave function lies below its tangents

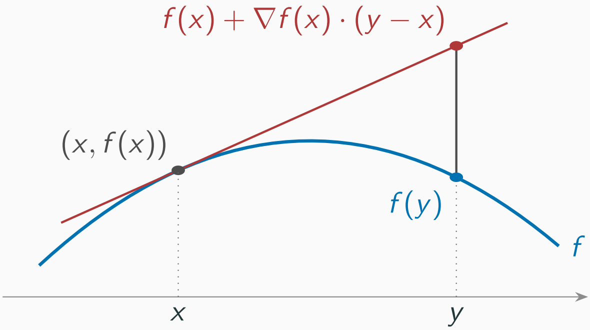


Figure 3.2

A sufficient gradient condition

The inequality below says that no feasible direction from x^* has a positive directional derivative.

Corollary 3.6 (First-order sufficiency)

Let $U \subseteq \mathbb{R}^n$ be open and convex, and let $C \subseteq U$ be convex.

Let $f : U \to \mathbb{R}$ be differentiable and concave.

If $x^* \in C$ satisfies $\nabla f(x^*) \cdot (x - x^*) \leq 0$ for every $x \in C$, then x^* maximizes f on C.

A sufficient gradient condition

The inequality below says that no feasible direction from x^* has a positive directional derivative.

Corollary 3.6 (First-order sufficiency)

Let $U \subseteq \mathbb{R}^n$ be open and convex, and let $C \subseteq U$ be convex.

Let $f : U \to \mathbb{R}$ be differentiable and concave.

If $x^* \in C$ satisfies $\nabla f(x^*) \cdot (x - x^*) \leq 0$ for every $x \in C$, then x^* maximizes f on C.

Proof.

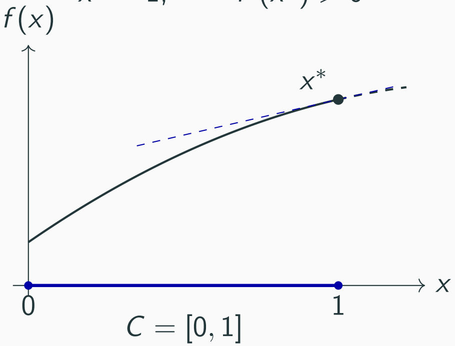
For $x \in C$, the tangent characterization at x^* and then the hypothesis give

$$f(x) \leq f(x^*) + \nabla f(x^*) \cdot (x - x^*) \leq f(x^*).$$

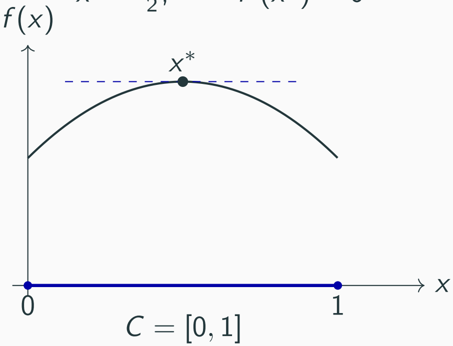
$\square$

The gradient condition in one dimension

$x^* = 1, \qquad f'(x^*) > 0$



$x^* = \frac{1}{2}, \qquad f'(x^*) = 0$



In both cases,

$$f'(x^*)(x - x^*) \leq 0 \qquad \text{for every } x \in C.$$

Uniqueness under strict concavity

Corollary 3.6 (Uniqueness under strict concavity)

Let $U \subseteq \mathbb{R}^n$ be open and convex, let $C \subseteq U$ be convex, and let $f : U \to \mathbb{R}$ be differentiable and strictly concave.

If $x^* \in C$ satisfies $\nabla f(x^*) \cdot (x - x^*) \leq 0$ for every $x \in C$, then x^* is the only maximizer of f on C.

Uniqueness under strict concavity

Corollary 3.6 (Uniqueness under strict concavity)

Let $U \subseteq \mathbb{R}^n$ be open and convex, let $C \subseteq U$ be convex, and let $f : U \to \mathbb{R}$ be differentiable and strictly concave.

If $x^* \in C$ satisfies $\nabla f(x^*) \cdot (x - x^*) \leq 0$ for every $x \in C$, then x^* is the only maximizer of f on C.

Proof.

Suppose $x \neq y$ both maximize f on C, with common value m.

Their midpoint lies in C, and strict concavity gives

$$f\left(\tfrac{1}{2}x + \tfrac{1}{2}y\right) > \tfrac{1}{2}m + \tfrac{1}{2}m = m,$$

which contradicts that m is the maximum. $\square$

The usual first-order condition

For an interior solution, the familiar first-order condition is

$$\nabla f(x^*) = 0.$$

If f is concave, any feasible stationary point is a global maximizer. If f is strictly concave, it is the only one.

At a boundary solution, the gradient need not equal zero.

Maximizing a strictly concave quadratic

Example 3.7 (A quadratic objective)

Fix $b \in \mathbb{R}^2$ and $|\gamma| < 2$. Maximize over $\mathbb{R}^2$ the strictly concave function

$$q(x) = b \cdot x - \left(x_1^2 + \gamma x_1 x_2 + x_2^2\right)$$

Solution.

$$\nabla q(x) = b - \begin{pmatrix} 2 & \gamma \\ \gamma & 2 \end{pmatrix} x.$$

The first-order condition $\nabla q(x^*) = 0$ gives

$$\begin{pmatrix} 2 & \gamma \\ \gamma & 2 \end{pmatrix} x^* = b, \qquad \text{so} \qquad x^* = \frac{1}{4-\gamma^2} \begin{pmatrix} 2 & -\gamma \\ -\gamma & 2 \end{pmatrix} b.$$

By Corollary 4, x^* is the unique global maximizer. $\square$

Hyperplane separation

Definition 3.8 (Hyperplane)

For $a \in \mathbb{R}^n \setminus \{0\}$ and $c \in \mathbb{R}$, the *hyperplane* with normal a at level c is

$$H(a, c) = \{x \in \mathbb{R}^n : a \cdot x = c\},$$

and its two *closed halfspaces* are $\{x : a \cdot x \leq c\}$ and $\{x : a \cdot x \geq c\}$.

Theorem 3.9 (Separating hyperplane)

Let $A, B \subseteq \mathbb{R}^n$ be nonempty, disjoint, and convex. Then there exist $a \neq 0$ and $c \in \mathbb{R}$ such that

$$a \cdot x \leq c \leq a \cdot y \qquad \text{for every } x \in A \text{ and every } y \in B.$$

Thus A and B lie in opposite closed halfspaces. The inequalities may be weak.