**Learning objectives.**

1. Verify convexity using the chord inequality definition.
2. Recognize convexity of twice differentiable functions using the positive-semidefinite Hessian test.
3. Disprove convexity by finding a violation of the definition or a point at which the Hessian is not positive semidefinite.

# 1 Convex functions

A general optimization problem can have many local minimizers:

$$\min f_0(x) \text{ s.t. } f_1(x) \leq 0, \ldots, f_m(x) \leq 0.$$

But if $f_0, f_1, \ldots, f_m$ are *convex*, then **every local minimum is global**. In many cases, the global minima can be efficiently found.

Let $D \subseteq \mathbb{R}^n$ be a convex set. A function $f : D \to \mathbb{R}$ is *convex* if

Graphically, the chord joining any two points on the graph lies on or above the graph.

**Question 1.** What if our problem is a *maximization*?

**Question 2.** What if our problem is to *maximize a convex function*? For example, $\max_x f(x)$ for $f$ convex (and isn't concave)?

Examples of convex functions? Fitting a curve to points. Fitting a straight line is a simple special case.

**Theorem 1. Polynomial regression.** Given data $(x_i, y_i)$ for $i = 1, \ldots, m$, define the degree-$d$ polynomial function

$$p_\theta(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_d x^d,$$

where $\theta = (c_0, c_1, \ldots, c_d)$ are the fit coefficients, and define the least-squares fit

$$f(\theta) = \frac{1}{2} \sum_{i=1}^{m} (p_\theta(x_i) - y_i)^2.$$

Then, $f$ is a convex function of $\theta$.

**Example 2.** Verify whether $f(x) = x^2$ is convex on $\mathbb{R}$.

A complete convexity proof should take three steps: 1) State a criterion, 2) substitute the specific function, and 3) manipulate the expression until the criterion is verified.

**Example 3.** Verify whether $f(x) = -x^2$ is convex on $\mathbb{R}$.

# 2 The second-order convexity test

## 2.1 Functions of scalars

The chord test is surprisingly complicated, even for very simple functions. A much easier test for convexity is the following.

**Theorem 4. Scalar curvature.** Let $f : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function of a scalar. Then

$$f \text{ is convex} \iff f''(x) \geq 0 \quad \text{for every } x \in \mathbb{R}.$$

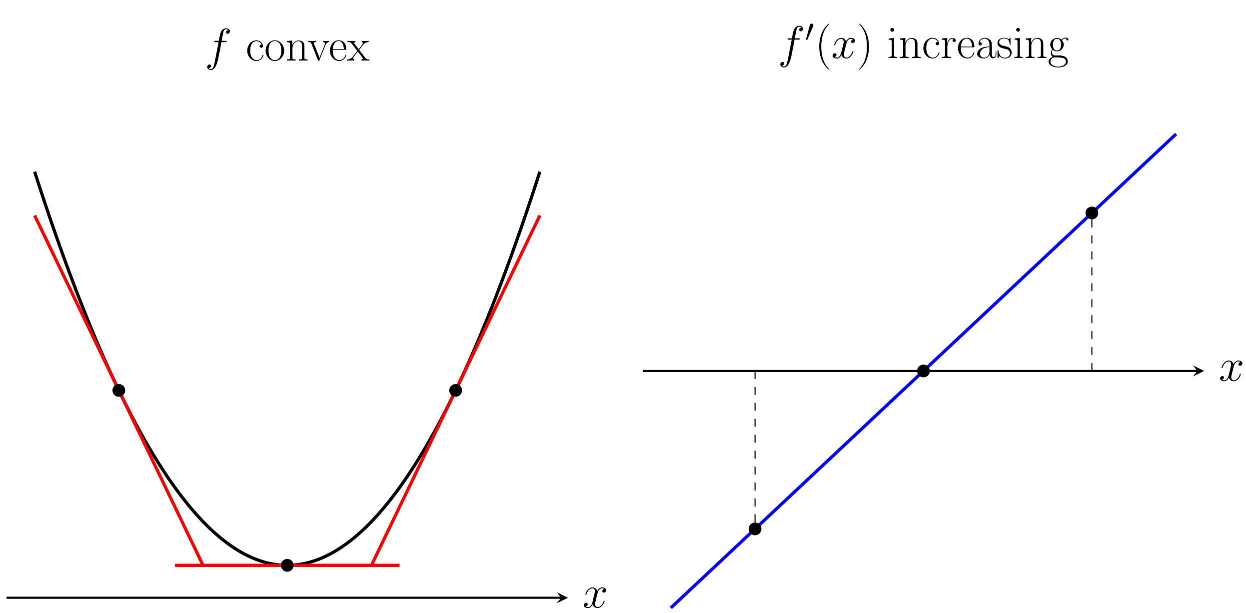


Fig. 1 A geometric illustration of the scalar curvature test. On the left, the tangent slopes of a convex function increase from left to right. On the right, the derivative $f'(x)$ is therefore increasing, which is equivalent to $f''(x) \geq 0$.

The second-order test is not always applicable. Functions like

$$|x|, \quad \mathrm{ReLU}(x) = \max\{x, 0\}, \quad \max\{x^2, (x-1)^2\}$$

are convex but do not have second derivatives. But when it is applicable, the proof is often extremely short and easy.

**Example 5.** Determine whether the following functions are convex on $\mathbb{R}$.

1. $f(x) = x^2$
2. $f(x) = \exp(-x)$
3. $f(x) = x^3$

## 2.2 Functions of vectors

How to generalize the second-order test?

**Lemma 6.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. Write its restriction to the line $at + b$ as follows

$$\phi_{a,b}(t) := f(at + b) \text{ for } a, b \in \mathbb{R}^n$$

Then $f$ is convex if and only if all of its line restrictions are convex

$$f \text{ is convex } \iff \phi_{a,b}(t) \text{ is convex for every } a, b \in \mathbb{R}^n.$$

**Theorem 7. Vector curvature.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function of a vector. Then

$$f \text{ is convex } \iff z^T \nabla^2 f(x) z \geq 0 \quad \text{for every } x, z \in \mathbb{R}^n.$$

## 2.3 Composition rules

These rules are valid even when $f, g, h$ are *not* twice differentiable. However, the proofs are particularly short when they are.

**Theorem 8. Nonnegative sum.** If $g, h : \mathbb{R}^n \to \mathbb{R}$ are twice differentiable and convex and $a, b \geq 0$, then

$$f(x) = ag(x) + bh(x)$$

is convex.

**Theorem 9. Affine composition.** Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. If $f : \mathbb{R}^m \to \mathbb{R}$ is twice differentiable and convex, then $g : \mathbb{R}^n \to \mathbb{R}$ defined by $g(x) = f(Ax + b)$ is also convex.

**Question 3.** If $f(x)$ and $g(x)$ are convex, does it mean that $f(x)g(x)$ is convex?

# 3 Positive semidefinite and definite

For $A = A^T \in \mathbb{R}^{n\times n}$, we write $A \succeq 0$ and say that $A$ is *positive semidefinite* (PSD) if

$$z^T A z \geq 0 \qquad \text{for every } z \in \mathbb{R}^n.$$

Second-order test rephrased.

**Theorem 10. Vector curvature.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function of a vector. Then

$$f \text{ is convex} \iff \nabla^2 f(x) \succeq 0 \quad \text{for every } x \in \mathbb{R}^n.$$

Similarly, we write $A \preceq 0$ and say that $A$ is *negative semidefinite* when $-A$ is positive semidefinite. Then,

$$f \text{ is concave} \iff \nabla^2 f(x) \preceq 0 \quad \text{for every } x \in \mathbb{R}^n.$$

**Question 4.** If $A \not\succeq 0$, does that mean $A \preceq 0$?

A symmetric matrix may be indefinite. Likewise, a function may be neither convex nor concave.

We write $A \succ 0$ and say that $A$ is *positive definite* if

$$z^T A z > 0 \qquad \text{for every } z \in \mathbb{R}^n, z \neq 0.$$

We write $A \succeq B$ to mean that $A - B \succeq 0$. We say that a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is $\mu$-strongly convex if, for a fixed $\mu > 0$, we have

$$\nabla^2 f(x) \succeq \mu I \text{ for all } x \in \mathbb{R}^n.$$

Gradient descent converges exponentially faster under strong convexity than convexity alone.

**Proposition 11.** If $A \succ 0$, then $A \succeq \mu I$ for some $\mu > 0$.

**Question 5.** Are all positive definite matrices also positive semidefinite?

**Question 6.** What is a negative definite matrix?

# 4 Convexity of Feature Regression

The convexity of polynomial regression is a special case of the following more general result with polynomial features $\phi_j(x) = x^j$.

**Theorem 12. Feature regression.** Given data $(x_i, y_i)$ for $i = 1, \ldots, m$, define the regression function

$$p_\theta(x) = c_0\phi_0(x) + c_1\phi_1(x) + \cdots + c_d\phi_d(x),$$

where $\theta = (c_0, c_1, \ldots, c_d)$ are the fit coefficients, and define the least-squares fit

$$f(\theta) = \frac{1}{2}\sum_{i=1}^{m}(p_\theta(x_i) - y_i)^2.$$

Then, $f$ is a convex function of $\theta$.

# 5 Testing for PD and PSD

## 5.1 Connection to eigenvalues

Let $A \in \mathbb{R}^{n\times n}$. A scalar $\lambda$ is an *eigenvalue* of $A$ if there is a nonzero vector $v$ such that

$$Av = \lambda v.$$

The vector $v$ is an *eigenvector*. Generally, eigenvalues can be complex, eigenvectors need not be orthogonal, and there need not be enough eigenvectors to form a basis. Symmetric matrices are special.

**Theorem 13. Spectral theorem.** For a real symmetric matrix $A = A^T$,

$$A = \underbrace{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}}_{U} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} - & u_1^T & - \\ - & u_2^T & - \\ & \vdots & \\ - & u_n^T & - \end{bmatrix},$$

with **real** eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and **orthonormal** matrix-of-eigenvectors $U$, meaning that

$$\begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix} \begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix}^T = \sum_i \begin{bmatrix} | \\ u_i \\ | \end{bmatrix} \begin{bmatrix} | \\ u_i \\ | \end{bmatrix}^T = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix},$$

$$\begin{bmatrix} | \\ u_i \\ | \end{bmatrix}^T \begin{bmatrix} | \\ u_j \\ | \end{bmatrix} = \left\langle \begin{bmatrix} | \\ u_i \\ | \end{bmatrix}, \begin{bmatrix} | \\ u_j \\ | \end{bmatrix} \right\rangle = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The matrix $U$ implements a **change of basis** where quadratic forms become weighted sum of squares.

**Theorem 14.** For $A = A^T \in \mathbb{R}^{n \times n}$,

$$A \succeq 0 \iff \lambda_i(A) \geq 0 \text{ for every } i.$$

## 5.2 Simple PSD tests

In this class, we use the following simple tests to verify PSD by hand.

**Proposition 15.**

$$a \succeq 0 \iff a \geq 0$$

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0 \iff a \geq 0, \quad c \geq 0, \quad ac - b^2 \geq 0$$

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \succ 0 \iff a > 0, \quad ac - b^2 > 0$$

$$\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix} \succeq 0 \iff A \succeq 0, C \succeq 0$$

Numerical PSD checks typically use a symmetric eigendecomposition or a pivoted $LDL^T$ factorization.

**Example 16. Classify three $2 \times 2$ matrices.** Classify each matrix as positive definite, positive semidefinite but not positive definite, or not positive semidefinite:

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

It is very dangerous to test positive semidefinite by looking at submatrices.

**Example 17. False shortcut.** Show that checking all $2 \times 2$ principal submatrices is not enough to prove that a $3 \times 3$ matrix is positive semidefinite. Use

$$A = \begin{bmatrix} 3 & -2 & -2 \\ -2 & 3 & -2 \\ -2 & -2 & 3 \end{bmatrix}.$$