Convex Functions

Pontus Giselsson

Learning goals

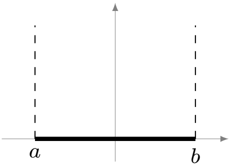
- Know convex function definition
- Understand extended-valued functions and domain
- Know about epigraphs and connection between convex hull and convex envelope
- Able to decide if function is convex from
  - First and second order conditions
  - Convexity preserving operations
- Understand strict convexity, strong convexity, and smoothness

Convex Functions

Extended-valued functions and domain

- We consider extended-valued functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} =: \overline{\mathbb{R}}$
- Example: Indicator function of interval $[a, b]$

$$\iota_{[a,b]}(x) = \begin{cases} 0 & \text{if } a \leq x \leq b \\ \infty & \text{else} \end{cases}$$



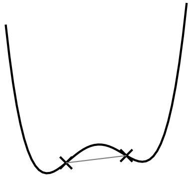
- The (effective) domain of $f \ : \ \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is the set

$$\text{dom } f = \{x \in \mathbb{R}^n : f(x) < \infty\}$$

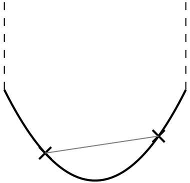
- (Will always assume $\mathrm{dom} f \neq \emptyset$, this is called proper)

Convex functions

- Graph below line connecting any two pairs $(x, f(x))$ and $(y, f(y))$



nonconvex function



convex function

- Function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *convex* if for all $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$:

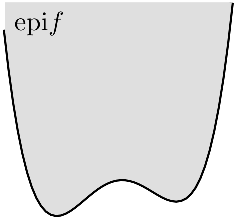
$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

(in extended valued arithmetics)

- A function f is *concave* if $-f$ is convex

Graphs and epigraphs

- The *epigraph* of a function f is the set of points above graph



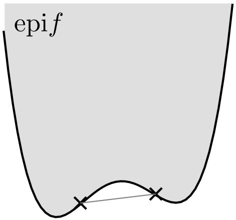
- Mathematical definition:

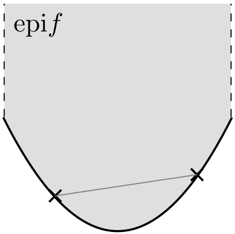
$$\mathrm{epi} f = \{(x, r) \mid f(x) \leq r\}$$

- The epigraph is a set in $\mathbb{R}^n \times \mathbb{R}$

Epigraphs and convexity

- Let $f \ : \ \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$
- Then f is convex if and only $\mathrm{epi} f$ is a convex set in $\mathbb{R}^n \times \mathbb{R}$

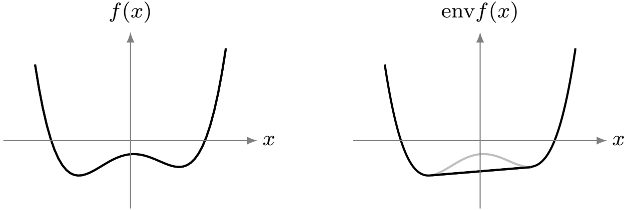




- f is called closed (lower semi-continuous) if $\mathrm{epi} f$ is closed set

Convex envelope

- Convex envelope of f is largest convex minorizer

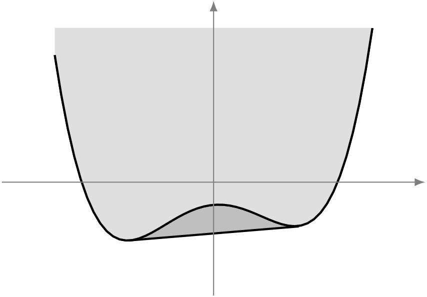


- Definition: The convex envelope $\mathrm{env} f$ satisfies: $\mathrm{env} f$ convex,

$$\mathrm{env} f \le f \qquad \text{and} \qquad \mathrm{env} f \ge g \text{ for all convex } g \le f$$

Convex envelope and convex hull

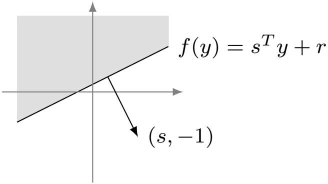
- Epigraph of convex envelope of f is convex hull of $\mathrm{epi} f$



- $\mathrm{epi} f$ in light gray, $\mathrm{epi}\,\mathrm{env} f$ includes dark gray

Affine functions

- Affine functions $f : \mathbb{R}^n \to \mathbb{R}$ cut $\mathbb{R}^n \times \mathbb{R}$ in two halves



- s defines slope of function
- Upper halfspace is epigraph with normal vector $(s, -1)$:

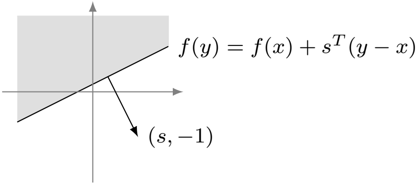
$$\mathrm{epi} f = \{(y,t) : t \geq s^T y + r\} = \{(y,t) : -r \geq (s,-1)^T (y,t)\}$$

Affine functions – Reformulation

- Pick any fixed $x \in \mathbb{R}^n$; affine $f(y) = s^T y + r$ can be written as

$$f(y) = f(x) + s^T(y - x)$$

(since $r = f(x) - s^T x$)



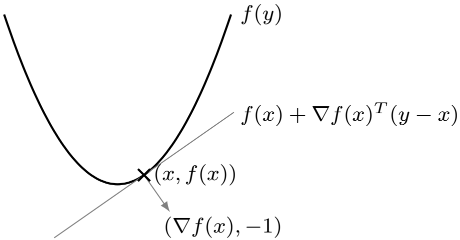
- We see affine function of this form important for convexity

First-order condition for convexity

- A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

  for all $x, y \in \mathbb{R}^n$



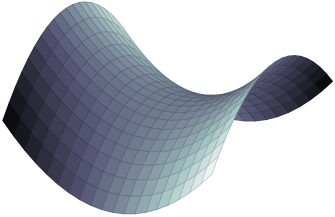
- Function f has for all $x \in \mathbb{R}^n$ an affine minorizer that:
  - has slope s defined by ∇f
  - coincides with function f at x
  - is supporting hyperplane to epigraph of f
  - defines normal $(\nabla f(x), -1)$ to epigraph of f

Second-order condition for convexity

- A twice differentiable function is convex if and only if

$$\nabla^2 f(x) \succeq 0$$

  for all $x \in \mathbb{R}^n$ (i.e., the Hessian is positive semi-definite)
- "The function has non-negative curvature"
- Nonconvex example: $f(x) = x^T \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right] x$ with $\nabla^2 f(x) \not\succeq 0$



Conclude convexity

For simple functions like

- indicator function

$$\iota_S(x) = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{else} \end{cases}$$

  convex function if and only if S convex set
- norms: $\|x\|$
- (shortest) distance to convex set: $\text{dist}_S(x) = \inf_{y \in S}(\|y - x\|)$
- affine functions: $f(x) = s^T x + r$
- quadratics: $f(x) = \frac{1}{2} x^T Q x$ with Q positive semi-definite matrix
- matrix fractional function: $f(x, Y) = x^T Y^{-1} x$

convexity concluded from definition or 1st or 2nd order conditions

Example – Convexity of norms

Show that $f(x) := \|x\|$ is convex

- Norms satisfy the triangle inequality

$$\|u + v\| \leq \|u\| + \|v\|$$

- Let $z = \theta x + (1 - \theta)y$ for arbitrary x, y and $\theta \in [0, 1]$:

$$\begin{aligned} f(z) &= \|\theta x + (1 - \theta)y\| \\ &\leq \|\theta x\| + \|(1 - \theta)y\| \\ &= \theta\|x\| + (1 - \theta)\|y\| \\ &= \theta f(x) + (1 - \theta)f(y) \end{aligned}$$

  which is definition of convexity

- Proof uses triangle inequality and $\theta \in [0, 1]$

Operations that preserve convexity

For more complicated functions, use convexity preserving operations:

- Positive sum
- Composition with matrix
- Image of function under affine mapping
- Supremum of convex functions
- A composition rule

Positive sum

- Assume that f_j are convex for $j = \{1, \ldots, m\}$
- Assume that there exists x such that $f_j(x) < \infty$ for all j
- Then positive sum

$$f = \sum_{j=1}^{m} t_j f_j$$

with $t_j > 0$ is convex

Composition with matrix

- Let f be convex and L be a matrix, then

$$(f \circ L)(x) := f(L(x))$$

  is convex

Image of function under linear mapping

- The image function $Lf \ : \ \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is defined as

$$(Lf)(x) := \inf_y \{f(y) \ : \ Ly = x\}$$

where $L \ : \ \mathbb{R}^m \to \mathbb{R}^n$ is a matrix and $f \ : \ \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$

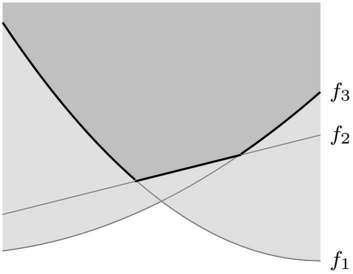
- Convex if f convex and bounded below for all x on inverse image

Supremum of convex functions

- Point-wise supremum of convex functions from family $\{f_j\}_{j \in J}$:

$$f(x) := \sup_j \{f_j(x) \ : \ j \in J\}$$

- Supremum is over functions in family for fixed x
- Example:



- Convex since intersection of convex epigraphs

Composition rule

- Consider the function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ defined as

$$f(x) = h(g(x))$$

  where $h : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is convex and $g : \mathbb{R}^n \to \mathbb{R}$
- Suppose that one of the following holds:
  - h is nondecreasing and g is convex
  - h is nonincreasing and g is concave
  - g is affine

  Then f is convex

Vector composition rule

- Consider the function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ defined as

$$f(x) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

  where $h : \mathbb{R}^k \to \mathbb{R} \cup \{\infty\}$ is convex and $g_i : \mathbb{R}^n \to \mathbb{R}$
- Suppose that for each $i \in \{1, \ldots, k\}$ one of the following holds:
  - h is nondecreasing in the ith argument and g_i is convex
  - h is nonincreasing in the ith argument and g_i is concave
  - g_i is affine

  Then f is convex

Convexity: Example 1

Show that: $f(x) := e^{\|Lx-b\|_2^4}$ is convex where L is matrix b vector:

- Let

$$g_1(u_1) = \|u_1\|_2, \quad g_2(u_2) = \begin{cases} 0 & \text{if } u_2 \le 0 \\ u_2^4 & \text{if } u_2 \ge 0 \end{cases}, \quad g_3(u_3) = e^{u_3}$$

  then $f(x) = g_3(g_2(g_1(Lx-b)))$
- $g_1(Lx-b)$ convex: convex g_1 and $Lx-b$ affine
- $g_2(g_1(Lx-b))$ convex: cvx nondecreasing g_2 and cvx $g_1(Lx-b)$
- $f(x)$ convex: convex nondecreasing g_3 and convex $g_2(g_1(Lx-b))$

Convexity: Example 2

Show that the *conjugate* $f^*(s) := \sup_{x \in \mathbb{R}^n} (s^T x - f(x))$ is convex:

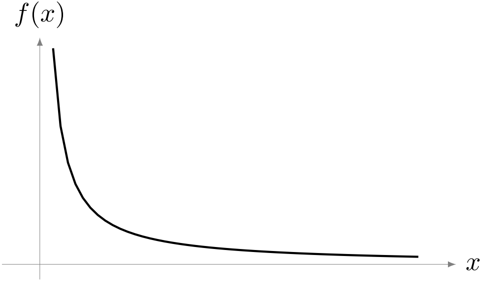
- Define (uncountable) index set J and x_j such that $\cup_{j \in J} x_j = \mathbb{R}^n$
- Define $r_j := f(x_j)$ and affine (in s): $a_j(s) := s^T x_j - r_j$
- Therefore $f^*(s) = \sup_j (a_j(s) : j \in J)$
- Convex since supremum over family of convex (affine) functions
- Note convexity of f^* not dependent on convexity of f

Strict convexity

- A function is strictly convex if

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$$

  for all $x \neq y$ and $\theta \in (0,1)$
- Convexity definition with strict inequality
- No flat (affine) regions
- Example: $f(x) = 1/x$ for $x > 0$

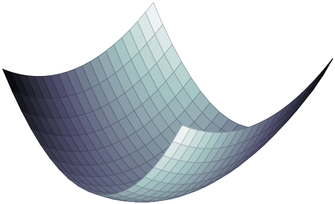


Strong convexity

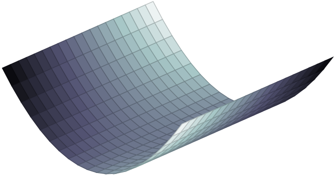
- Let $\sigma > 0$
- A function f is *σ-strongly convex* if $f - \frac{\sigma}{2}\|\cdot\|_2^2$ is convex
- Alternative equivalent definition of σ-strong convexity:

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) - \frac{\sigma}{2}\theta(1-\theta)\|x-y\|^2$$

  holds for every $x, y \in \mathbb{R}^n$ and $\theta \in [0,1]$
- Strongly convex functions are strictly convex and convex
- Example: f 2-strongly convex since $f - \|\cdot\|_2^2$ convex:



$f(x)$



$f(x) - \|x\|_2^2$

Uniqueness of minimizers

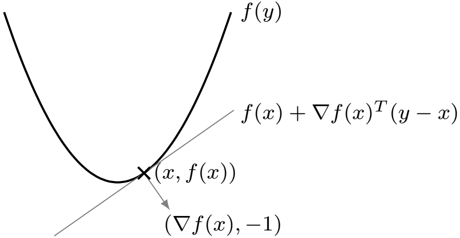
- Strictly (strongly) convex functions have unique minimizers
- Strictly convex functions may not have a minimizing point
- Strongly convex functions always have a unique minimizing point

First-order condition for strict convexity

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable
- f is strictly convex if and only if

$$f(y) > f(x) + \nabla f(x)^T(y - x)$$

  for all $x, y \in \mathbb{R}^n$ where $x \neq y$



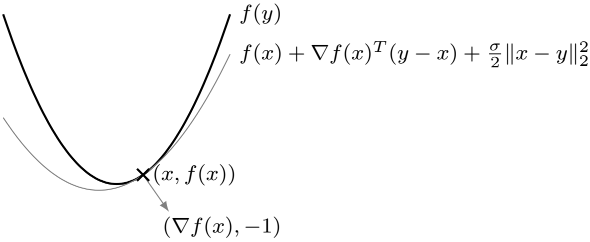
- Function f has for all $x \in \mathbb{R}^n$ an affine minorizer that:
  - has slope s defined by ∇f
  - coincides with function f *only* at x
  - is supporting hyperplane to epigraph of f
  - defines normal $(\nabla f(x), -1)$ to epigraph of f

First-order condition for strong convexity

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable
- f is σ-strongly convex with $\sigma > 0$ if and only if

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\sigma}{2}\|x - y\|_2^2$$

  for all $x, y \in \mathbb{R}^n$



- Function f has for all $x \in \mathbb{R}^n$ a quadratic minorizer that:
  - curvature defined by σ
  - coincides with function f at x
  - defines normal $(\nabla f(x), -1)$ to epigraph of f

Second-order condition for strict/strong convexity

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable

- f is strictly convex if

$$\nabla^2 f(x) \succ 0$$

  for all $x \in \mathbb{R}^n$ (i.e., the Hessian is positive definite)

- f is σ-strongly convex if and only if

$$\nabla^2 f(x) \succeq \sigma I$$

  for all $x \in \mathbb{R}^n$

Examples of strictly/strongly convex functions

Strictly convex

- $f(x) = -\log(x) + \iota_{>0}(x)$
- $f(x) = 1/x + \iota_{>0}(x)$
- $f(x) = e^x$
- $f(x) = e^{-x}$

Strongly convex

- $f(x) = \frac{\lambda}{2}\|x\|_2^2$
- $f(x) = \frac{1}{2}x^T Qx$ where Q positive definite
- $f(x) = f_1(x) + f_2(x)$ where f_1 strongly convex and f_2 convex
- $f(x) = \frac{1}{2}x^T Qx + \iota_C(x)$ where Q positive definite and C convex

Proof of two examples

Strict convexity of $f(x) = e^{-x}$:

- $\nabla f(x) = -e^{-x}$, $\nabla^2 f(x) = e^{-x} > 0$ for all $x \in \mathbb{R}$

Strong convexity of $f(x) = \frac{1}{2}x^T Q x$ with Q positive definite

- $\nabla f(x) = Qx$, $\nabla^2 f(x) = Q \succeq \lambda_{\min}(Q) I$ where $\lambda_{\min}(Q) > 0$

Smoothness

- A function is called β-*smooth* if its gradient is β-Lipschitz:

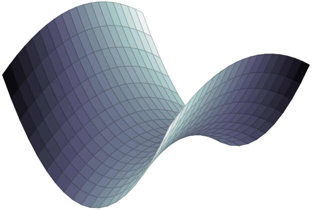
$$\|\nabla f(x) - \nabla f(y)\|_2 \leq \beta \|x - y\|_2$$

  for all $x, y \in \mathbb{R}^n$ (it is not necessarily convex)
- Alternative equivalent definition of β-smoothness

$$f(\theta x + (1-\theta)y) \geq \theta f(x) + (1-\theta) f(y) - \frac{\beta}{2}\theta(1-\theta)\|x-y\|^2$$

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y) + \frac{\beta}{2}\theta(1-\theta)\|x-y\|^2$$

  hold for every $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$
- Smoothness does not imply convexity
- Example:

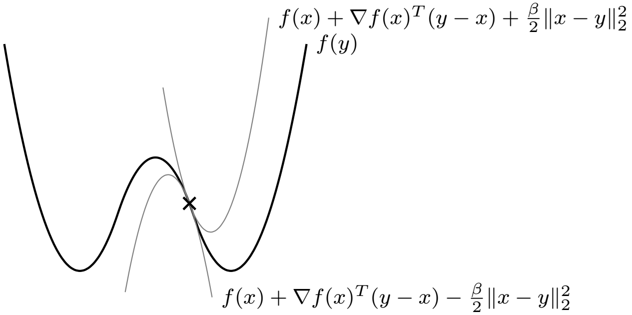


First-order condition for smoothness

- f is β-smooth with $\beta \geq 0$ if and only if

$$f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{\beta}{2}\|x-y\|_2^2$$
$$f(y) \geq f(x) + \nabla f(x)^T(y-x) - \frac{\beta}{2}\|x-y\|_2^2$$

for all $x, y \in \mathbb{R}^n$



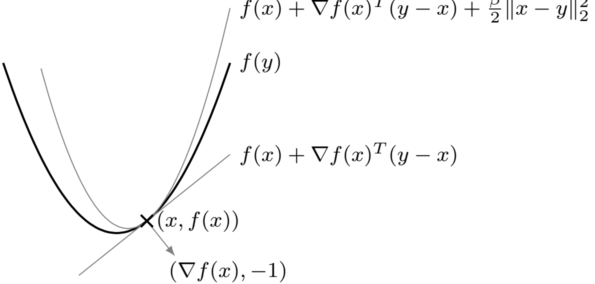
- Quadratic upper/lower bounds with curvatures defined by β
- Quadratic bounds coincide with function f at x

First-order condition for smooth convex

- f is β-smooth with $\beta \geq 0$ and convex if and only if

$$f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{\beta}{2}\|x-y\|_2^2$$
$$f(y) \geq f(x) + \nabla f(x)^T(y-x)$$

for all $x, y \in \mathbb{R}^n$



- Quadratic upper bounds and affine lower bound
- Bounds coincide with function f at x
- Quadratic upper bound is called *descent lemma*

Second-order condition for smoothness

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable

- f is β-smooth if and only if

$$-\beta I \preceq \nabla^2 f(x) \preceq \beta I$$

  for all $x \in \mathbb{R}^n$

- f is β-smooth and convex if and only if

$$0 \preceq \nabla^2 f(x) \preceq \beta I$$

  for all $x \in \mathbb{R}^n$

Convex Optimization Problems

Composite optimization form

- We will consider optimization problem on composite form

$$\underset{x}{\text{minimize}}\, f(Lx) + g(x)$$

  where f and g convex function and L a matrix
- Convex problem due to convexity preserving operations
- Can model constrained problems via indicator function
- This model format is suitable for many algorithms