9. Convex optimization problems

- convex sets
- convex functions
- operations preserving convexity
- basic properties
- convex problems

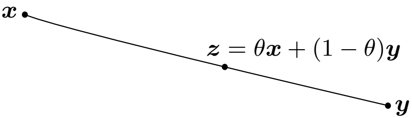
Line segment

a *line* passing through non-equal points $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^n$ has the form

$$\{\boldsymbol{z} \mid \boldsymbol{x} = \theta\boldsymbol{x} + (1-\theta)\boldsymbol{y},\ \theta \in \mathbb{R}\}$$

Line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\{\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y} \mid \theta \in [0,1]\}$$

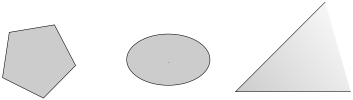


Convex sets

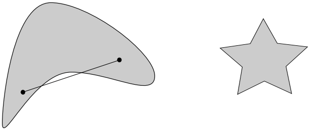
a set $\mathcal{C} \subseteq \mathbb{R}^n$ is *convex* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$, we have

$$\theta \boldsymbol{x} + (1 - \theta)\boldsymbol{y} \in \mathcal{C}$$

for any $\theta \in [0, 1]$, *i.e.*, the line segment between any two points in $\mathcal{C}$ lies in $\mathcal{C}$



convex sets



nonconvex sets

a point on the line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$ is called a *convex combination* of the points $\boldsymbol{x}$ and $\boldsymbol{y}$

Example 9.1

- *Affine sets:* a set $\mathcal{C} \subseteq \mathbb{R}^n$ is *affine* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ and θ, we have

$$\theta \boldsymbol{x} + (1 - \theta) \boldsymbol{y} \in \mathcal{C}$$

  since the above holds for any θ, it holds also for $\theta \in [0, 1]$; hence, affine sets are also convex (the converse is not true)

- the empty set, any single point (singleton), and $\mathbb{R}^n$ are affine, hence convex

- *Lines:* a line in $\mathbb{R}^n$ is a set of the form:

$$\mathcal{L} = \{\boldsymbol{x}_0 + t\boldsymbol{d} \mid t \in \mathbb{R}\}$$

  where $\boldsymbol{x}_0, \boldsymbol{d} \in \mathbb{R}^n$ and $\boldsymbol{d} \neq \boldsymbol{0}$

- *Rays:* a ray $\{\boldsymbol{x}_0 + t\boldsymbol{d} \mid t \geq 0\}$, where $\boldsymbol{d} \neq \boldsymbol{0}$, is convex

- *Ellipsoids:* an ellipsoid is a set of the form

$$\mathcal{E} = \{\boldsymbol{x} \mid \boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{r}^T \boldsymbol{x} + c \leq 0\},$$

  where $Q \in \mathbb{R}^{n \times n}$ is positive definite, $\boldsymbol{r} \in \mathbb{R}^n$, and $c \in \mathbb{R}$; an ellipsoid is a convex set

- *Hyperplane and halfspaces:* let $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, then, the hyperplane $\mathcal{H} = \{\boldsymbol{x} \mid \boldsymbol{a}^T \boldsymbol{x} = b\}$ and the halfspace $\mathcal{H}^- = \{\boldsymbol{x} \mid \boldsymbol{a}^T \boldsymbol{x} \leq b\}$ are convex sets

- *Balls:* let $\boldsymbol{c} \in \mathbb{R}^n$, $r > 0$, and $\| \cdot \|$ be an arbitrary norm; then, the open ball

$$\mathcal{B}(\boldsymbol{c}, r) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{c}\| < r\}$$

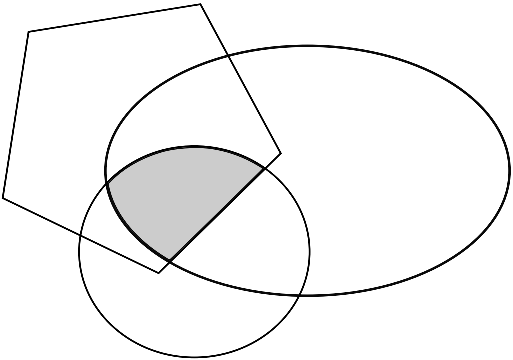
  and closed ball

$$\mathcal{B}[\boldsymbol{c}, r] = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{c}\| \leq r\}$$

  are convex

Intersection of convex sets

the intersection of any collection of convex sets is convex



Properties

- if $\mathcal{C}$ is a convex set and β is a real number, then the set

$$\beta\mathcal{C} = \{\beta \boldsymbol{y} \mid \boldsymbol{y} \in \mathcal{C}\}$$

  is also convex

- if $\mathcal{C}_1$ and $\mathcal{C}_2$ are convex sets, then the set

$$\mathcal{C}_1 + \mathcal{C}_2 = \{\boldsymbol{x}_1 + \boldsymbol{x}_2 \mid \boldsymbol{x}_1 \in \mathcal{C}_1, \boldsymbol{x}_2 \in \mathcal{C}_2\}$$

  is convex

- suppose that $f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}$ where $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$; if $\mathcal{C} \subset \mathbb{R}^n$ is convex, then the image set

$$f(\mathcal{C}) = \{A\boldsymbol{x} + \boldsymbol{b} \mid \boldsymbol{x} \in \mathcal{C}\}$$

  is convex

Outline

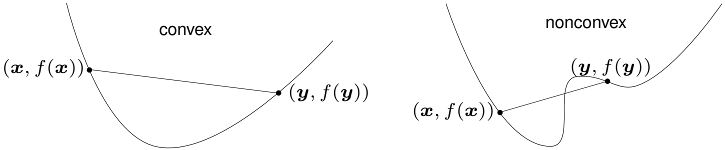
- convex sets
- **convex functions**
- operations preserving convexity
- basic properties
- convex problems

Definition

$f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if $\mathrm{dom}\, f$ is a convex set and

$$f\big(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\big) \le \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}), \tag{9.1}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\, f$, and $0 \le \theta \le 1$



- f is *strictly convex* if strict inequality holds in (9.1)
- f is *concave* (*strictly concave*) if $-f$ is convex (strictly convex)
- f is convex over convex set $\mathcal{X} \subseteq \mathbb{R}^n$ if (9.1) holds for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$
- f is convex iff for all $\boldsymbol{x} \in \mathrm{dom}\, f$ and $\boldsymbol{v} \in \mathbb{R}^n$, the function $g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$ is convex on its domain $\{t \mid \boldsymbol{x} + t\boldsymbol{v} \in \mathrm{dom}\, f\}$

Example 9.2

- *Affine functions:* $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$ where $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, is both convex and concave:

$$\begin{aligned} f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) &= \boldsymbol{a}^T((\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y})) + b \\ &= \theta(\boldsymbol{a}^T\boldsymbol{x} + b) + (1-\theta)(\boldsymbol{a}^T\boldsymbol{y} + b) \\ &= \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) \end{aligned}$$

- *Norm functions:* $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ for any norm $\|\cdot\|$ is convex:

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

  where the inequality follows from the triangle inequality

- consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ where $f(x_1, x_2) = x_1 x_2$ and $\operatorname{dom} f = \{\boldsymbol{x} \mid x_1, x_2 \geq 0\}$; this function is nonconvex over since for $\boldsymbol{x} = (1, 2)$, $\boldsymbol{y} = (2, 1)$, $\theta = 0.5$, we have

$$f(0.5\boldsymbol{x} + 0.5\boldsymbol{y}) = \frac{9}{4} \nleq 0.5 f(\boldsymbol{x}) + 0.5 f(\boldsymbol{y}) = 2,$$

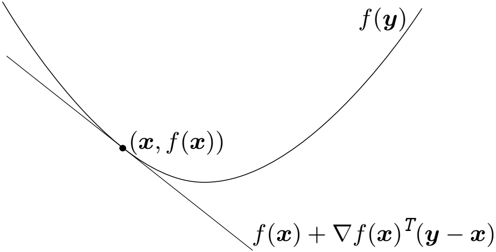
  which violates the definition of convexity

- the function $f(x) = x$ over $\operatorname{dom} f = \{x \mid x \neq 1\}$ is not convex even though it is linear; this is because its domain is nonconvex

First-order convexity condition

if $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, then f is convex if and only if its domain is convex and for any $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\, f$

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) \tag{9.2}$$



- f is strictly convex if strict inequality holds
- if $\nabla f(\boldsymbol{x}) = \mathbf{0}$, then the inequality (9.2) becomes $f(\boldsymbol{x}) \leq f(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathrm{dom}\, f$ implying thay $\boldsymbol{x}$ is a global minimizer of f

Second-order convexity condition

suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, then f is convex if and only if its domain is convex and for all $\boldsymbol{x} \in \operatorname{dom} f$, we have

$$\nabla^2 f(\boldsymbol{x}) \geq 0 \tag{9.3}$$

- if $\nabla^2 f(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$, then f is strictly convex
- converse is not true since $f(x) = x^4$ is strictly convex but has zero second derivative at $x = 0$

Convexity of domain:

- domain of f must be convex to use the first or second order convexity characterization
- for example, the function $f(x) = 1/x^2$ with $\operatorname{dom} f = \{x \in \mathbb{R} \mid x \neq 0\}$ satisfies $f''(x) = 6/x^4 > 0$ for all $x \in \operatorname{dom} f$, but is not a convex function

Example 9.3

convexity or concavity of the following examples can be shown using the definition or the second order condition

- *Exponential:* $e^{\alpha x}$ is convex for any $\alpha \in \mathbb{R}$
- *Powers:* x^α is convex on $\mathbb{R}_{++} = \{x \mid x > 0\}$ when $\alpha \geq 1$ or $\alpha \leq 0$, and concave for $0 \leq \alpha \leq 1$
- *Powers of absolute value:* $|x|^p$ is convex on $\mathbb{R}$ for $p \geq 1$
- *Logarithm:* $\log x$ is concave on $\mathbb{R}_{++}$
- *Negative entropy:* $x \log x$ defined as 0 for $x = 0$ is convex on $\mathbb{R}_+ = \{x \mid x \geq 0\}$

Example 9.4 (Quadratic functions)

$f(\boldsymbol{x}) = \boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{r}^T \boldsymbol{x} + c$ where $Q = Q^T$ is convex if and only if $Q \geq 0$

- $f(\boldsymbol{x}) = 4x_1^2 + 2x_2^2 + 3x_1x_2 + 4x_1 + 5x_2$ is convex since its Hessian

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 8 & 3 \\ 3 & 4 \end{bmatrix}$$

  is positive definite

- $f(\boldsymbol{x}) = 4x_1^2 - 2x_2^2 + 3x_1x_2 + 4x_1 + 5x_2$ is nonconvex since its Hessian

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 8 & 3 \\ 3 & -4 \end{bmatrix}$$

  is indefinite

Example 9.5

Quadratic over linear: the function

$$f(x,t) = x^2/t$$

with $\mathrm{dom}\, f = \{(x,t) \mid t > 0\}$ is convex; this is because the Hessian

$$\nabla^2 f(\boldsymbol{x}) = 2 \begin{bmatrix} 1/t & -x/t^2 \\ -x/t^2 & x^2/t^3 \end{bmatrix} = \frac{2}{t^3} \begin{bmatrix} t \\ -x \end{bmatrix} \begin{bmatrix} t & -x \end{bmatrix} \geq 0,$$

is positive semidefinite over its domain

Example 9.6

Log-sum-exp function: the function

$$f(\boldsymbol{x}) = \log(e^{x_1} + \cdots + e^{x_n})$$

is convex over $\mathbb{R}^n$; we now show this by showing that the Hessian is positive semidefinite

- the partial derivatives of f are:

$$\frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}}$$

  the second partial derivatives are

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{cases} \frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}} - \frac{e^{x_i} e^{x_i}}{(\sum_{k=1}^{n} e^{x_k})^2}, & \text{if } i = j \\ -\frac{e^{x_i} e^{x_j}}{(\sum_{k=1}^{n} e^{x_k})^2}, & \text{if } i \neq j \end{cases}$$

- thus, we can express the Hessian as

$$\nabla^2 f(\boldsymbol{x}) = \operatorname{diag}(\boldsymbol{w}) - \boldsymbol{w}\boldsymbol{w}^T$$

  where $\boldsymbol{w} = \left(\frac{e^{x_1}}{\sum_{k=1}^{n} e^{x_k}}, \ldots, \frac{e^{x_n}}{\sum_{k=1}^{n} e^{x_k}} \right)$

- note that for any $\boldsymbol{v} \in \mathbb{R}^n$, we have

$$\boldsymbol{v}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} = \sum_{i=1}^n w_i v_i^2 - (\boldsymbol{v}^T \boldsymbol{w})^2$$

- applying Cauchy-Schwarz on the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ with entries

$$a_i = \sqrt{w_i} v_i, \quad b_i = \sqrt{w_i}, \quad i = 1, \ldots, n$$

  we get

$$(\boldsymbol{v}^T \boldsymbol{w})^2 = (\boldsymbol{a}^T \boldsymbol{b})^2 \leq \|\boldsymbol{a}\|^2 \|\boldsymbol{b}\|^2 = \left(\sum_{i=1}^n w_i v_i^2 \right) \left(\sum_{i=1}^n w_i \right) = \sum_{i=1}^n w_i v_i^2$$

  it follows that $\boldsymbol{v}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} \geq 0$ for any $\boldsymbol{v} \in \mathbb{R}^n$

Outline

- convex sets
- convex functions
- **operations preserving convexity**
- basic properties
- convex problems

Operations preserving convexity

Weighted nonnegative sum: the function

$$f = w_1 f_1 + \cdots + w_k f_k$$

is convex if f_i are convex and $w_i \geq 0$

- a nonnegative weighted sum of concave functions is concave
- a nonnegative nonzero weighted sum of strictly convex (concave) functions is strictly convex (concave)

Composition with affine mapping: suppose that $g : \mathbb{R}^m \to \mathbb{R}$, $A \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$; let $f : \mathbb{R}^n \to \mathbb{R}$

$$f(\boldsymbol{x}) = g(A\boldsymbol{x} + \boldsymbol{b}),$$

with $\operatorname{dom} f = \{\boldsymbol{x} \mid A\boldsymbol{x} + \boldsymbol{b} \in \operatorname{dom} g\}$; then, f is convex (concave) if g is convex (concave)

Example 9.7

- *Negative entropy function:* $f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i \log x_i$ is convex over $\operatorname{dom} f = \mathbb{R}^n_{++} = \{\boldsymbol{x} \mid x_i > 0\}$ since it is the sum of convex functions $x_i \log x_i$

- $f(x) = -\log(ax+b)$ is convex over $ax+b>0$ since $g(t) = -\log(t)$ is convex over $\operatorname{dom} f = \mathbb{R}_{++}$

- $f(\boldsymbol{x}) = e^{\boldsymbol{a}^T\boldsymbol{x}+b}$ where $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ is convex over $\mathbb{R}^n$; we can write f as $f(\boldsymbol{x}) = g(\boldsymbol{a}^T\boldsymbol{x}+b)$ where $g(t) = e^t$ is a convex function; hence, f is convex

- consider the function

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2 + e^{x_1}$$

  - we can write f as $f = f_1 + f_2$ with

$$f_1(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2, \quad f_2(x_1, x_2) = e^{x_1}$$

  - f_1 is convex since $\nabla^2 f(x_1, x_2) = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$ is positive semidefinite
  - f_2 is also convex since $g(t) = e^t$ is convex and $f_2(x_1, x_2) = g(x_2)$
  - hence, f is convex since it is the sum of two convex functions.

- the function

$$f(x_1, x_2, x_3) = e^{x_1 - x_2 + x_3} + e^{2x_2} + x_1$$

  is convex over $\mathbb{R}^3$; it is the sum of three convex functions: $e^{x_1 - x_2 + x_3}$, e^{2x_2}, and x_1

Example 9.8

Generalized quadratic-over-linear: let $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{c} \in \mathbb{R}^n$ ($\boldsymbol{c} \neq \boldsymbol{0}$), and $d \in \mathbb{R}$, then the function

$$f(\boldsymbol{x}) = \frac{\|A\boldsymbol{x} + \boldsymbol{b}\|^2}{\boldsymbol{c}^T\boldsymbol{x} + d}$$

is convex over $\operatorname{dom} f = \{\boldsymbol{x} \mid \boldsymbol{c}^T\boldsymbol{x} + d > 0\}$

- we can write f as

$$f(\boldsymbol{x}) = g(A\boldsymbol{x} + \boldsymbol{b}, \boldsymbol{c}^T\boldsymbol{x} + d), \qquad g(\boldsymbol{y}, t) = \frac{\|\boldsymbol{y}\|^2}{t}$$

  with $\operatorname{dom} f = \{(\boldsymbol{y}, t) \mid \boldsymbol{y} \in \mathbb{R}^m,\ t > 0\}$

- $g = \sum_{i=1}^m g_i$ where $g_i(\boldsymbol{y}, t) = \frac{y_i^2}{t}$ is convex over $\{(y_i, t) \mid y_i \in \mathbb{R},\ t > 0\}$; thus, g is convex since it is the sum of convex function

- thus f is convex (composition of convex function with an affine mapping)

Pointwise maximum of convex functions

if $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, k$ are convex, then

$$f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), \dots, f_k(\boldsymbol{x})\}$$

is convex

Examples

- *Maximum function:* $f(\boldsymbol{x}) = \max\{x_1, x_2, \dots, x_n\}$ is convex because it is the maximum of n linear (hence convex) functions
- *Sum of k largest values:* let $x_{[i]}$ denote the ith largest component of $\boldsymbol{x}$, then the function

$$f_k(\boldsymbol{x}) = x_{[i]} + \cdots + x_{[k]}$$

is convex; to see this, note that we can rewrite f_k as

$$f_k(\boldsymbol{x}) = \max\{x_{i_1} + \cdots + x_{i_k} \mid i_1, \dots, i_k \in \{1, 2, \dots, n\} \text{ are different}\}$$

hence, f_k is a maximum of linear functions, hence convex

Composition with a nondecreasing convex function

let $h : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ and define $f = g \circ h : \mathbb{R}^n \to \mathbb{R}$:

$$f(\boldsymbol{x}) = g(h(\boldsymbol{x})), \quad \mathrm{dom}\, f = \{\boldsymbol{x} \in \mathrm{dom}\, h \mid h(\boldsymbol{x}) \in \mathrm{dom}\, g\}$$

- f is convex if h is convex, and g is convex and nondecreasing (over the range of h)
- f is convex if h is concave, and $\tilde{g}$ is convex and nonincreasing
- f is concave if h is concave, and $\tilde{g}$ is concave and nondecreasing
- f is concave if h is convex, and $\tilde{g}$ is concave and nonincreasing

here $\tilde{g}$ denotes the extended-value extension of the function g, which assigns the value ∞ $(-\infty)$ to points not in $\mathrm{dom}\, g$ for g convex (concave)

Proof:

$$
\begin{aligned}
f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) &= g\big(h(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y})\big) \\
&\leq g\big(\theta h(\boldsymbol{x}) + (1-\theta)h(\boldsymbol{y})\big) \\
&\leq \theta g\big(h(\boldsymbol{x})\big) + (1-\theta)g\big(h(\boldsymbol{y})\big) \\
&= \theta f(\boldsymbol{x}) + (1-\theta)f(\boldsymbol{x}),
\end{aligned}
$$

where the first inequality arises from the convexity of h and the nondecreasing nature of g; the second inequality is a result of the convexity of $\tilde{g}$

Example 9.9

- $f(\boldsymbol{x}) = e^{\|\boldsymbol{x}\|^2}$ is convex since $f(\boldsymbol{x}) = g(h(\boldsymbol{x}))$ where
  - $h(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ is a convex function
  - $g(t) = e^t$ is a nondecreasing convex function

  more generally, $e^{h(\boldsymbol{x})}$ is convex if h is convex

- $f(\boldsymbol{x}) = (1 + \|\boldsymbol{x}\|^2)^2$ is a convex function since $f(\boldsymbol{x}) = g(h(\boldsymbol{x}))$ where
  - $h(\boldsymbol{x}) = 1 + \|\boldsymbol{x}\|^2$ is convex
  - $g(t) = t^2$, which is convex and nondecreasing over h (*i.e.*, the interval $[1, \infty)$)

- if h is convex and nonnegative, then $h(\boldsymbol{x})^p$ is convex for $p \geq 1$

- if h is convex, then $-\log(-h(\boldsymbol{x}))$ is convex on $\{\boldsymbol{x} \mid h(\boldsymbol{x}) < 0\}$

- if h is concave and positive, then $1/h(\boldsymbol{x})$ is convex

- if h is concave and positive, then $\log h(\boldsymbol{x})$ is concave

Vector functions composition

the aforementioned principle can be extended to functions that take a vector as their argument:

$$f(\boldsymbol{x}) = g(h(\boldsymbol{x})) = g(h_1(\boldsymbol{x}), \ldots, h_k(\boldsymbol{x}))$$

- $h_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, k$, are convex
- if the function $g : \mathbb{R}^k \to \mathbb{R}$ is convex and non-decreasing in every argument, given that $\operatorname{dom} h_i = \mathbb{R}^n$ and $\operatorname{dom} g = \mathbb{R}^k$, then the function $f(\boldsymbol{x}) = g\left(h_1(\boldsymbol{x}), \ldots, h_k(\boldsymbol{x})\right)$ is also convex

Example 9.10

- $g(\boldsymbol{z}) = \log(\sum_{i=1}^{k} e^{z_i})$ is convex and nondecreasing in each argument; hence, $g(h(\boldsymbol{x})) = \log(\sum_{i=1}^{k} e^{h_i(\boldsymbol{x})})$ is convex when h_i are convex

- suppose $p \geq 1$, and let $h_1, \ldots, h_k$ be convex and nonnegative functions; then function given by $\left(\sum_{i=1}^{k} h_i(\boldsymbol{x})^p\right)^{\frac{1}{p}}$ is convex

  to demonstrate this, we introduce the function $g : \mathbb{R}^k \to \mathbb{R}$ defined as

$$g(\boldsymbol{z}) = (\sum_{i=1}^{k} \max\{z_i, 0\}^p)^{\frac{1}{p}},$$

  with $\operatorname{dom} g = \mathbb{R}^k$; since this function is both convex and nondecreasing in its arguments, $g(h(\boldsymbol{x}))$ is also convex in x; for nonnegative values of z, $g(z)$ simplifies to

$$(\sum_{i=1}^{k} z_i^p)^{\frac{1}{p}},$$

  leading us to conclude that $(\sum_{i=1}^{k} h_i(\boldsymbol{x})^p)^{\frac{1}{p}}$ is convex

Minimizing over some variables

suppose that $f : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}$ is convex in $(\boldsymbol{x}, \boldsymbol{y})$ and $\mathcal{C}$ is a convex set; then, the function

$$g(\boldsymbol{x}) = \min_{\boldsymbol{y} \in \mathcal{C}} f(\boldsymbol{x}, \boldsymbol{y})$$

is convex (provided that $g(\boldsymbol{x}) > \infty$ for some $\boldsymbol{x}$); the domain of g is

$$\operatorname{dom} g = \{\boldsymbol{x} \mid (\boldsymbol{x}, \boldsymbol{y}) \in \operatorname{dom} f \text{ for some } \boldsymbol{y} \in \mathcal{C}\}$$

Example: for a convex set $\mathcal{C} \subset \mathbb{R}^n$, the *distance function* defined as

$$d(\boldsymbol{x}, \mathcal{C}) = \min\{\|\boldsymbol{x} - \boldsymbol{y}\| \mid \boldsymbol{y} \in \mathcal{C}\}$$

is convex because $f(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ is convex in both $(\boldsymbol{x}, \boldsymbol{y})$

Outline

- convex sets
- convex functions
- operations preserving convexity
- **basic properties**
- convex problems

Line restriction and convexity

suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and define

$$g(t) = f(\boldsymbol{x} + t\boldsymbol{v}), \quad \operatorname{dom} g = \{t \mid \boldsymbol{x} + t\boldsymbol{v} \in \operatorname{dom} f\}$$

- f if and only if, for every $\boldsymbol{x} \in \operatorname{dom} f$ and all $\boldsymbol{v} \in \mathbb{R}^n$, the function $g(t)$ is convex over its domain
- this means that function is convex if it remains convex when restricted to any line intersecting its domain

Example 9.11

the *log-determinant* function $f(X) = -\log\det X$ is convex over the domain of symmetric, positive definite matrices

to verify this let $X_0 \in \mathbb{R}^{n\times n}$ be a positive definite matrix, $V \in \mathbb{R}^{n\times n}$ be symmetric, and consider the scalar-valued function

$$g(t) = -\log\det\left(X_0 + tV\right)$$

since $X_0 > 0$, it can be factored (matrix square-root factorization) as $X_0 = X_0^{1/2}X_0^{1/2}$, hence

$$\begin{aligned}\det\left(X_0 + tV\right) &= \det\left(X_0^{1/2}X_0^{1/2} + tV\right)\\ &= \det X_0 \det\left(I + tX_0^{-1/2}VX_0^{-1/2}\right)\\ &= \det X_0 \prod_{i=1,\ldots,n}\left(1 + t\lambda_i(Z)\right)\end{aligned}$$

where $\lambda_i(Z)$, are the eigenvalues of the matrix $Z = X_0^{-1/2}VX_0^{-1/2}$

taking the logarithm, we thus obtain

$$g(t) = -\log\det X_0 + \sum_{i=1}^{n} -\log\left(1 + t\lambda_i(Z)\right)$$

- the first term in the previous expression is a constant
- the second term is the sum of convex functions
- hence $g(t)$ is convex for any positive definite matrix $X_0 \in \mathbb{R}^{n\times n}$, and symmetric $V \in \mathbb{R}^{n\times n}$
- it follows that $-\log\det X$ is convex over the domain of positive definite matrices

Epigraph

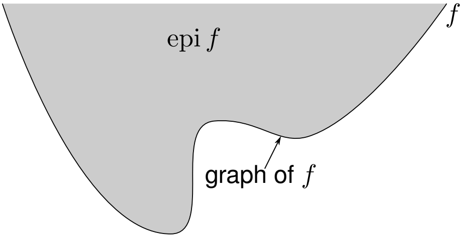
the *graph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is described as

$$\{(\boldsymbol{x}, f(\boldsymbol{x})) \mid \boldsymbol{x} \in \operatorname{dom} f\} \subset \mathbb{R}^{n+1}$$

The *epigraph* of $f : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\operatorname{epi}(f) = \{(\boldsymbol{x}, s) \mid \boldsymbol{x} \in \operatorname{dom} f,\ f(\boldsymbol{x}) \leq s\} \subset \mathbb{R}^{n+1}$$

- the epigraph encompasses the points situated on or above the graph of f



- a function is convex if and only if its epigraph constitutes a convex set

Example 9.12

consider the function $f : \mathbb{R}^n \times \mathbb{R}^{n \times n} \to \mathbb{R}$, represented by

$$f(\boldsymbol{x}, Y) = \boldsymbol{x}^T Y^{-1} \boldsymbol{x}$$

where Y is positive definite

we can determine the convexity of f is by examining its epigraph:

$$\begin{aligned} \operatorname{epi} f &= \{(\boldsymbol{x}, Y, t) \mid Y \geq 0, \boldsymbol{x}^T Y^{-1} \boldsymbol{x} \leq t\} \\ &= \left\{(\boldsymbol{x}, Y, t) \mid \begin{bmatrix} Y & \boldsymbol{x} \\ \boldsymbol{x}^T & t \end{bmatrix} \geq 0, Y > 0\right\}, \end{aligned}$$

utilizing the Schur complement criteria for a block matrix's positive semidefiniteness; the latter condition is linear matrix inequality (LMI) in the variables $(\boldsymbol{x}, Y, t)$, signifying that the epigraph of f is convex

Sublevel sets and convexity

the sublevel set of $f : \mathbb{R}^n \to \mathbb{R}$ at level γ is defined as

$$\mathcal{S}_\gamma = \{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq \gamma\}$$

- for a convex function f, the sublevel set $\mathcal{S}_\gamma$ is also convex; to see this, observe that

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) \leq \gamma$$

  for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}_\gamma$

- a function can have all its sublevel sets convex, but not be a convex
  - for example, $f(x) = -e^x$ is not convex on $\mathbf{R}$ (indeed, it is strictly concave) but all its sublevel sets are convex
  - another example is the function $f(x) = \ln(x)$, which is concave; however, its sublevel sets, which are intervals of the form $(0, e^\gamma]$, are convex

Example 9.13

the set:

$$\mathcal{C} = \left\{ \boldsymbol{x} \mid (\boldsymbol{x}^T P \boldsymbol{x} + 1)^2 + \ln\left(\sum_{i=1}^{n} e^{x_i} \right) \leq 3 \right\},$$

where $P \geq 0$ is an $n \times n$ matrix, is convex since it is the level set of a convex function

$$f(\boldsymbol{x}) = \left(\boldsymbol{x}^T P \boldsymbol{x} + 1\right)^2 + \ln\left(\sum_{i=1}^{n} e^{x_i} \right)$$

- f is convex, being the sum of two convex functions
- the log-sum-exp function, previously established as convex
- the function $h(\boldsymbol{x}) = \left(\boldsymbol{x}^T P \boldsymbol{x} + 1\right)^2$, which is convex since it can be represented as a composition of the nondecreasing convex function $g(t) = (t+1)^2$ (defined on $\mathbb{R}_+$) with the convex quadratic function $\boldsymbol{x}^T P \boldsymbol{x}$

Outline

- convex sets
- convex functions
- operations preserving convexity
- basic properties
- **convex problems**

Definition

Convex optimization problems

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, p \end{array} \tag{9.4}$$

- f and g_i are convex
- $h_j(\boldsymbol{x})$ are affine, *i.e.*, $h_j(\boldsymbol{x}) = \boldsymbol{a}_j^T \boldsymbol{x} - b_j$ for some $\boldsymbol{a}_j \in \mathbb{R}^n$ and $b_j \in \mathbb{R}$
- the feasible set is convex since it is the intersection of convex sets

Concave problems

- when the problem is a maximization with concave objective and convex constraints, then the problem is said to be *concave optimization problem*
- a concave problem is also referred to as a convex problem

Example 9.14

- the problem

$$\begin{array}{ll} \text{minimize} & -2x_1 + x_2 \\ \text{subject to} & x_1^2 + x_2^2 \leq 4 \end{array}$$

  is convex

- the problem

$$\begin{array}{ll} \text{minimize} & -2x_1 + x_2 \\ \text{subject to} & x_1^2 + x_2^2 = 4 \end{array}$$

  is nonconvex since the equality constraint function $h(\boldsymbol{x}) = x_1^2 + x_2^2 - 4$ is not affine

Example 9.15

- an investor wants to invest a total value of at most d into n possible investment opportunities

- if x_i is investment deposit for investment i; in economy it is frequently assumed that $f_i(x_i)$ have forms:

$$f_i(x_i) = \alpha_i(1 - e^{-\beta_i x_i}), \quad f_i(x_i) = \alpha_i \log(1 + \beta_i x_i), \quad f_i(x_i) = \frac{\alpha_i x_i}{x_i + \beta_i}$$

  with $\alpha_i, \beta_i > 0$; the above functions are concave

- we want to determine the investment deposits that maximize expected profit; we can formulate the optimization problem:

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{n} f_i(x_i) \\ \text{subject to} & \sum_{i=1}^{n} x_i \leq d \\ & x_i \geq 0, \quad i = 1, \dots, n \end{array}$$

  this is a convex problem (we can transform max into min)

Local minimizers are global minimizers

if the function $f : \mathbb{R}^n \to \mathbb{R}$ is convex (convex with convex domain), then, any local minimizer is a global minimizer

Proof:

- if $\boldsymbol{x}^o$ is a local minimizer of f, then $f(\boldsymbol{x}^o) \le f(\boldsymbol{z})$ for all points $\boldsymbol{z}$ with $\|\boldsymbol{z} - \boldsymbol{x}^o\| \le R$
- assume that there exists a feasible $\boldsymbol{y}$ such that $f(\boldsymbol{y}) < f(\boldsymbol{x}^o)$ so that $\boldsymbol{x}^o$ is not a global minimizer
- since $f(\boldsymbol{y}) < f(\boldsymbol{x}^o)$, we have $\|\boldsymbol{y} - \boldsymbol{x}^o\| > R$; let $\boldsymbol{z} = \theta \boldsymbol{y} + (1-\theta)\boldsymbol{x}^o$, from convexity definition, we have

$$f(\boldsymbol{z}) = f\big(\theta \boldsymbol{y} + (1-\theta)\boldsymbol{x}^o\big) \le \theta f(\boldsymbol{y}) + (1-\theta) f(\boldsymbol{x}^o) < f(\boldsymbol{x}^o)$$

- for $\theta = R/2\|\boldsymbol{y} - \boldsymbol{x}^o\|$, we have $\|\boldsymbol{z} - \boldsymbol{x}^o\| = R/2 < R$; this implies that there is a point $\boldsymbol{z}$ close to $\boldsymbol{x}^o$ such that $f(\boldsymbol{z}) < f(\boldsymbol{x}^o)$; this contradicts that $\boldsymbol{x}^o$ is a local minimizer
- hence, there is no feasible $\boldsymbol{y}$ such that $f(\boldsymbol{y}) < f(\boldsymbol{x}^o)$, *i.e.*, $\boldsymbol{x}^o$ is a global minimizer

A first-order optimality condition

suppose that a convex function $f : \mathcal{X} \to \mathbb{R}$ is defined on a convex set $\mathcal{X} \subset \mathbb{R}^n$; the point $\boldsymbol{x}^\star$ is optimal if and only if

$$\nabla f(\boldsymbol{x}^\star)^T(\boldsymbol{y} - \boldsymbol{x}^\star) \geq 0, \quad \forall\, \boldsymbol{y} \in \mathcal{X} \tag{9.5}$$

(the above condition is difficult to verify in practice)

Unconstrained case: for $\mathcal{X} = \mathbb{R}^n$, the above condition reduces to

$$\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$$

to see this suppose that $\boldsymbol{x} \in \operatorname{dom} f$ is optimal and let $\boldsymbol{y} = \boldsymbol{x} - t\nabla f(\boldsymbol{x})$, which is in the domain of f for sufficiently small t (since domain is open by definition); note that

$$\nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) = -t\|\nabla f(\boldsymbol{x})\|^2 \geq 0$$

hence, $\nabla f(\boldsymbol{x}) = \mathbf{0}$

Sufficiency of KKT conditions

suppose that there exists points $\boldsymbol{x}^\star \in \mathcal{D}$ ($\mathcal{D}$ is domain of (9.4)), $\boldsymbol{\mu}^\star \in \mathbb{R}^m$, and $\boldsymbol{\lambda}^\star \in \mathbb{R}^p$ satisfying the KKT conditions

$$\begin{aligned}
\nabla f(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \mu_i^\star \nabla g_i(\boldsymbol{x}^\star) + \sum_{j=1}^{p} \lambda_j^\star \nabla h_j(\boldsymbol{x}^\star) &= \mathbf{0} \\
g_i(\boldsymbol{x}^\star) &\leq 0, \quad i = 1, \dots, m \\
A\boldsymbol{x}^\star &= \boldsymbol{b} \\
\mu_i^\star &\geq 0, \quad i = 1, \dots, m \\
g_i(\boldsymbol{x}^\star)\mu_i^\star &= 0, \quad i = 1, \dots, m
\end{aligned}$$

then, $\boldsymbol{x}^\star$ is a global minimizer of problem (9.4)

Proof: let x be a feasible solution; note that the function

$$J(\boldsymbol{x}) = L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i^\star g_i(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_j^\star h_j(\boldsymbol{x})$$

is convex since it is the sum of convex functions; since $\nabla J(\boldsymbol{x}^\star) = \mathbf{0}$, $\boldsymbol{x}^\star$ is a minimizer of J over $\mathbb{R}^n$; thus,

$$\begin{aligned} f(\boldsymbol{x}^\star) &\overset{\text{kkt}}{=} f(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \mu_i^\star g_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \lambda_j^\star h_j(\boldsymbol{x}^\star) \\ &= J(\boldsymbol{x}^\star) \\ &\leq J(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i^\star g_i(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_j^\star h_j(\boldsymbol{x}) \\ &\leq f(\boldsymbol{x}) \end{aligned}$$

hence, $\boldsymbol{x}^\star$ is optimal

Slater's constraint qualification

Slater's condition is satisfied if there exists an $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ such that

$$g_i(\hat{\boldsymbol{x}}) < 0, \quad i = 1, \ldots, m, \quad A\hat{\boldsymbol{x}} = \boldsymbol{b}$$

- if Slater condition holds, then the KKT conditions are necessary and sufficient for optimality
- we can weaken Slater condition if some g_i are affine by only requiring the non-affine functions to hold with strict inequality

Example 9.16

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\
\text{subject to} & x_1 + x_2 + x_3 = 3
\end{array}
$$

the above problem is convex with an equality constraint, hence, the KKT conditions are necessary and sufficient for optimality; the Lagrangian is

$$L(\boldsymbol{x}, \lambda) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \lambda(x_1 + x_2 + x_3 - 3)$$

the KKT conditions are

$$
\begin{aligned}
x_1 + \lambda &= 0 \\
x_2 + \lambda &= 0 \\
x_3 + \lambda &= 0 \\
x_1 + x_2 + x_3 &= 0
\end{aligned}
$$

the unique optimal solution is $\boldsymbol{x} = (1, 1, 1)$ and $\lambda = -1$

Example 9.17

$$\begin{array}{ll} \text{minimize} & x_1^2 - x_2 \\ \text{subject to} & x_2^2 \leq 0 \end{array}$$

it is easy to see that the solution is $\boldsymbol{x}^\star = (0, 0)$; for this problem Slater condition is not satisfied since we cannot find an $\boldsymbol{x}$ such that $x_2^2 < 0$; the Lagrangian is

$$L(\boldsymbol{x}, \mu) = \tfrac{1}{2}x_1^2 - x_2 + \mu x_2^2$$

the KKT conditions are

$$\begin{aligned} 2x_1 &= 0 \\ -1 + 2\mu x_2 &= 0 \\ \mu x_2^2 &= 0 \\ x_2^2 &\leq 0 \\ \mu &\geq 0 \end{aligned}$$

the above nonlinear system of equations is infeasible

Example 9.18

$$\begin{array}{ll} \text{minimize} & 4x_1^2 + x_2^2 - x_1 - 2x_2 \\ \text{subject to} & 2x_1 + x_2 \leq 1 \\ & x_1^2 \leq 1 \end{array}$$

Slater's condition is satisfied for $\hat{\boldsymbol{x}} = (0, 0)$, hence, the KKT conditions are necessary and sufficient for optimality; the Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{\mu}) = 4x_1^2 + x_2^2 - x_1 - 2x_2 + \mu_1(2x_1 + x_2 - 1) + \mu_2(x_1^2 - 1)$$

and the KKT conditions are

$$\begin{aligned} 8x_1 - 1 + 2\mu_1 + 2\mu_2 x_1 &= 0 \\ 2x_2 - 2 + \mu_2 &= 0 \\ \mu_1(2x_1 + x_2 - 1) &= 0 \\ \mu_2(x_1^2 - 1) &= 0 \\ 2x_1 + x_2 &\leq 1 \\ x_1^2 &\leq 1 \\ \mu_1, \mu_2 &\geq 0 \end{aligned}$$

- for $\mu_1 = \mu_2 = 0$, the KKT system will be infeasible
- for $\mu_1, \mu_2 > 0$, the KKT system will be infeasible
- for $\mu_1 = 0, \mu_2 > 0$, the KKT system will be infeasible
- for $\mu_1 > 0, \mu_2 = 0$, we will get $(x_1, x_2, \mu_1) = (\frac{1}{16}, \frac{7}{8}, \frac{1}{4})$
- hence, from convexity $\boldsymbol{x} = (\frac{1}{16}, \frac{7}{8})$ is the optimal unique solution

References and further readings

- Amir Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB,* SIAM, 2014.
- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization,* Cambridge University Press, 2004.
- Edwin KP Chong and Stanislaw H Zak. *An Introduction to Optimization,* John Wiley & Sons, 2013.