

9. Convex optimization problems

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

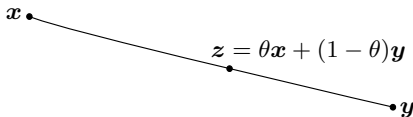
Line segment

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

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

Line segment between x and y :

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

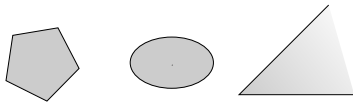


Convex sets

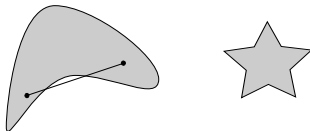
a set $\mathcal{C} \subseteq \mathbb{R}^n$ is *convex* if for any $x, y \in \mathcal{C}$, we have

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

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 x and y is called a *convex combination* of the points x and y

Example 9.1

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

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

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:

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

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

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

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

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

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

- *Hyperplane and halfspaces*: let $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, then, the hyperplane $\mathcal{H} = \{\mathbf{x} \mid \mathbf{a}^T \mathbf{x} = b\}$ and the halfspace $\mathcal{H}^- = \{\mathbf{x} \mid \mathbf{a}^T \mathbf{x} \leq b\}$ are convex sets
- *Balls*: let $\mathbf{c} \in \mathbb{R}^n$, $r > 0$, and $\|\cdot\|$ be an arbitrary norm; then, the open ball

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

and closed ball

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

are convex

Linear matrix inequality

a *linear matrix inequality* (LMI) is represented by:

$$F(\mathbf{x}) = F_0 + \sum_{i=1}^n x_i F_i \leq 0, \quad (9.1)$$

- $\mathbf{x} \in \mathbb{R}^n$, F_0, \dots, F_n are symmetric of size $m \times m$
- the solution set of a linear matrix inequality, $\{\mathbf{x} \mid F(\mathbf{x}) \leq 0\}$, is convex

Example any solution $\mathbf{w}(t)$ to the linear differential equation

$$\dot{\mathbf{w}}(t) = A\mathbf{w}(t), \quad A \in \mathbb{R}^{n \times n},$$

converges to the origin as t approaches infinity if and only if there exists a real symmetric matrix X satisfying the conditions:

$$AX + XA^T < 0, \quad X > 0 \quad (9.2)$$

let us express the variable vector $x \in \mathbb{R}^m$ as:

$$X = x_1 X_1 + x_2 X_2 + \cdots + x_m X_m,$$

where the matrices X_i ($i = 1, 2, \dots, m$) serve as a basis for the linear space spanned by $n \times n$ symmetric matrices (with $m = n(n+1)/2$); for instance, when $n = 2$, we have $m = 3$ and:

$$X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

given this representation, the inequality in (9.2) can be recast as:

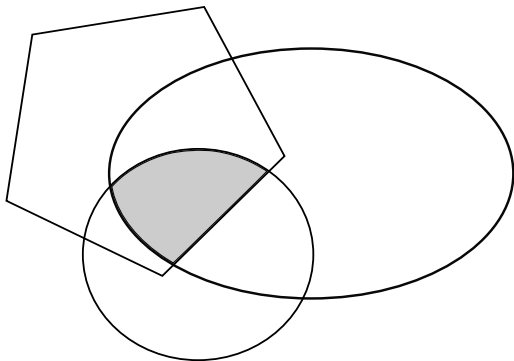
$$F(x) \triangleq \begin{bmatrix} -X & 0 \\ 0 & AX + XA^T \end{bmatrix} < 0,$$

which can then be expressed in the form of (9.1), where $F_0 = 0$ and:

$$F_i = \begin{bmatrix} -X_i & 0 \\ 0 & AX_i + X_i A^T \end{bmatrix}, \quad (i = 1, \dots, m)$$

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

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

is also convex

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

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

is convex

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

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

is convex

Outline

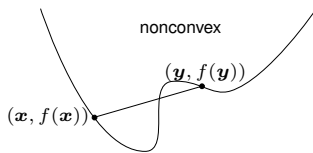
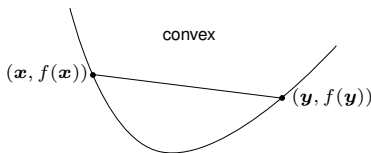
- convex sets
- **convex functions**
- operations preserving convexity
- basic properties
- convex problems

Definition

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *convex* if $\text{dom } f$ is a convex set and

$$f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}), \quad (9.3)$$

for all $\mathbf{x}, \mathbf{y} \in \text{dom } f$, and $0 \leq \theta \leq 1$



- f is *strictly convex* if strict inequality holds in (9.3)
- f is *concave* (*strictly concave*) if $-f$ is convex (strictly convex)
- f is convex over convex set $\mathcal{X} \subseteq \mathbb{R}^n$ if (9.3) holds for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$
- f is convex iff for all $\mathbf{x} \in \text{dom } f$ and $\mathbf{v} \in \mathbb{R}^n$, the function $g(t) = f(\mathbf{x} + t\mathbf{v})$ is convex on its domain $\{t \mid \mathbf{x} + t\mathbf{v} \in \text{dom } f\}$

Example 9.2

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

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

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

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

where the inequality follows from the triangle inequality

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

$$f(0.5\mathbf{x} + 0.5\mathbf{y}) = \frac{9}{4} \not\leq 0.5f(\mathbf{x}) + 0.5f(\mathbf{y}) = 2,$$

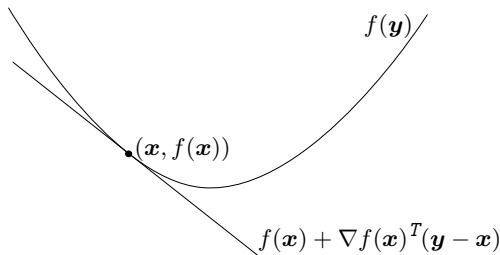
which violates the definition of convexity

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

First-order convexity condition

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

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad (9.4)$$



- f is strictly convex if strict inequality holds
- if $\nabla f(\mathbf{x}) = \mathbf{0}$, then the inequality (9.4) becomes $f(\mathbf{x}) \leq f(\mathbf{y})$ for all $\mathbf{y} \in \text{dom } f$ implying that \mathbf{x} is a global minimizer of f

Second-order convexity condition

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

$$\nabla^2 f(\mathbf{x}) \succeq 0 \tag{9.5}$$

- if $\nabla^2 f(\mathbf{x}) \succ 0$ for all \mathbf{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 $\text{dom } f = \{x \in \mathbb{R} \mid x \neq 0\}$ satisfies $f''(x) = 6/x^4 > 0$ for all $x \in \text{dom } f$, 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)

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

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

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

is positive definite

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

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

is indefinite

Example 9.5

Quadratic over linear: the function

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

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

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

is positive semidefinite over its domain

Example 9.6

Log-sum-exp function: the function

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

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

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

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

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

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

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

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

we get

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

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

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 $g : \mathbb{R}^m \rightarrow \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$; let $f : \mathbb{R}^n \rightarrow \mathbb{R}$

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

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

Example 9.7

- *Negative entropy function:* $f(\mathbf{x}) = \sum_{i=1}^n x_i \log x_i$ is convex over $\text{dom } f = \mathbb{R}_{++}^n = \{\mathbf{x} \mid x_i > 0\}$ 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 $\text{dom } f = \mathbb{R}_{++}$
- $f(\mathbf{x}) = e^{\mathbf{a}^T \mathbf{x} + b}$ where $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ is convex over \mathbb{R}^n ; we can write f as $f(\mathbf{x}) = g(\mathbf{a}^T \mathbf{x} + b)$ 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 $f_2(x_1, x_2) = g(x_1)$
 - 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 $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$ ($\mathbf{c} \neq \mathbf{0}$), and $d \in \mathbb{R}$, then the function

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

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

- we can write f as

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

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

- $g = \sum_{i=1}^m g_i$ where $g_i(\mathbf{y}, t) = \frac{y_i^2}{t}$ 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 $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, k$ are convex, then

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

is convex

Examples

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

$$f_k(\mathbf{x}) = x_{[1]} + \dots + x_{[k]}$$

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

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

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

Composition with a nondecreasing convex function

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

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

- 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 $\text{dom } g$ for g convex (concave)

Proof:

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

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

- $f(\mathbf{x}) = e^{\|\mathbf{x}\|^2}$ is convex since $f(\mathbf{x}) = g(h(\mathbf{x}))$ where
 - $h(\mathbf{x}) = \|\mathbf{x}\|^2$ is a convex function
 - $g(t) = e^t$ is a nondecreasing convex functionmore generally, $e^{h(\mathbf{x})}$ is convex if h is convex
- $f(\mathbf{x}) = (1 + \|\mathbf{x}\|^2)^2$ is a convex function since $f(\mathbf{x}) = g(h(\mathbf{x}))$ where
 - $h(\mathbf{x}) = 1 + \|\mathbf{x}\|^2$ 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 $h(\mathbf{x})^p$ is convex for $p \geq 1$
- if h is convex, then $-\log(-h(\mathbf{x}))$ is convex on $\{\mathbf{x} \mid h(\mathbf{x}) < 0\}$
- if h is concave and positive, then $1/h(\mathbf{x})$ is convex
- if h is concave and positive, then $\log h(\mathbf{x})$ is concave

Vector functions composition

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

$$f(\mathbf{x}) = g(h(\mathbf{x})) = g(h_1(\mathbf{x}), \dots, h_k(\mathbf{x}))$$

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

Example 9.10

- $g(\mathbf{z}) = \log(\sum_{i=1}^k e^{z_i})$ is convex and nondecreasing in each argument; hence, $g(h(\mathbf{x})) = \log(\sum_{i=1}^k e^{h_i(\mathbf{x})})$ is convex when h_i are convex
- suppose $p \geq 1$, and let h_1, \dots, h_k be convex and nonnegative functions; then function given by $(\sum_{i=1}^k h_i(\mathbf{x})^p)^{\frac{1}{p}}$ is convex
to demonstrate this, we introduce the function $g : \mathbb{R}^k \rightarrow \mathbb{R}$ defined as

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

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

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

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

Minimizing over some variables

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

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

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

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

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

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

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

Outline

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

Line restriction and convexity

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

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

- f is convex if and only if, for every $\mathbf{x} \in \text{dom } f$ and all $\mathbf{v} \in \mathbb{R}^n$, 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

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

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

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

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

taking the logarithm, we thus obtain

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

- 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

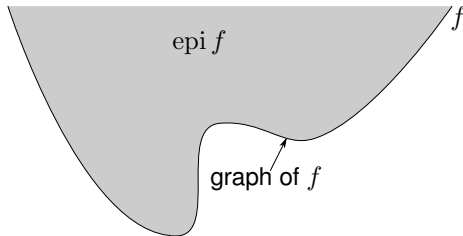
the *graph* of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is described as

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

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

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

- 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 $f : \mathbb{R}^n \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$, represented by

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

where Y is positive definite

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

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

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

Sublevel sets and convexity

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

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

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

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

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

- 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:

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

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

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

- f is convex, being the sum of two convex functions
- the log-sum-exp function, previously established as convex
- the function $h(\mathbf{x}) = (\mathbf{x}^T P \mathbf{x} + 1)^2$, 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 $\mathbf{x}^T P \mathbf{x}$

Outline

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

Definition

Convex optimization problems

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \dots, p \end{array} \quad (9.6)$$

- f and g_i are convex
- $h_j(\mathbf{x})$ are affine, i.e., $h_j(\mathbf{x}) = \mathbf{a}_j^T \mathbf{x} - b_j$ for some $\mathbf{a}_j \in \mathbb{R}^n$ 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 $h(\mathbf{x}) = x_1^2 + x_2^2 - 4$ 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex (convex with convex domain), then, any local minimizer is a global minimizer

Proof:

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

$$f(z) = f(\theta y + (1 - \theta)x^o) \leq \theta f(y) + (1 - \theta)f(x^o) < f(x^o)$$

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

A first-order optimality condition

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

$$\nabla f(\mathbf{x}^*)^T(\mathbf{y} - \mathbf{x}^*) \geq 0, \quad \forall \mathbf{y} \in \mathcal{X} \quad (9.7)$$

(the above condition is difficult to verify in practice)

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

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

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

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

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

Sufficiency of KKT conditions

suppose that there exists points $\mathbf{x}^* \in \mathcal{D}$ (\mathcal{D} is domain of (9.6)), $\boldsymbol{\mu}^* \in \mathbb{R}^m$, and $\boldsymbol{\lambda}^* \in \mathbb{R}^p$ satisfying the KKT conditions

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \mu_i^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^p \lambda_j^* \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

$$g_i(\mathbf{x}^*) \leq 0, \quad i = 1, \dots, m$$

$$A\mathbf{x}^* = \mathbf{b}$$

$$\mu_i^* \geq 0, \quad i = 1, \dots, m$$

$$g_i(\mathbf{x}^*)\mu_i^* = 0, \quad i = 1, \dots, m$$

then, \mathbf{x}^* is a global minimizer of problem (9.6)

Proof: let \mathbf{x} be a feasible solution; note that the function

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

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

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

hence, \mathbf{x}^* is optimal

Slater's constraint qualification

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

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

- 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

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

the KKT conditions are

$$x_1 + \lambda = 0$$

$$x_2 + \lambda = 0$$

$$x_3 + \lambda = 0$$

$$x_1 + x_2 + x_3 = 3$$

the unique optimal solution is $\mathbf{x} = (1, 1, 1)$ 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 $\mathbf{x}^* = (0, 0)$; for this problem Slater condition is not satisfied since we cannot find an \mathbf{x} such that $x_2^2 < 0$; the Lagrangian is

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

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

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

and the KKT conditions are

$$8x_1 - 1 + 2\mu_1 + 2\mu_2x_1 = 0$$

$$2x_2 - 2 + \mu_2 = 0$$

$$\mu_1(2x_1 + x_2 - 1) = 0$$

$$\mu_2(x_1^2 - 1) = 0$$

$$2x_1 + x_2 \leq 1$$

$$x_1^2 \leq 1$$

$$\mu_1, \mu_2 \geq 0$$

- 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 $x = (\frac{1}{16}, \frac{7}{8})$ 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.