

## 9. Convex optimization problems

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

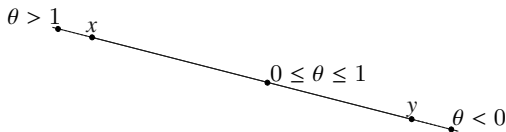
## Line segment

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

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

**Line segment** between  $x$  and  $y$ :

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

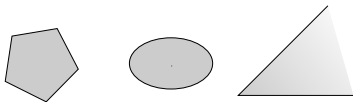


## Convex sets

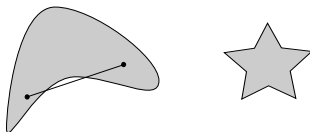
a set  $C \subseteq \mathbb{R}^n$  is *convex* if for any  $x, y \in C$ , we have

$$\theta x + (1 - \theta)y \in C \quad \text{for any } \theta \in [0, 1]$$

*i.e.*, a convex set contains the line segment between any two points in the set



**convex sets**



**nonconvex sets**

a point on line segment between  $x$  and  $y$  is called a *convex combination* of  $x$  and  $y$

## Affine sets

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

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

- a set that contains the line through any two distinct points in the set
- a convex set since it holds for any  $\theta$ , so it holds also for  $\theta \in [0, 1]$
- a point  $\theta x + (1 - \theta)y$  is called an *affine combination* of  $x, y$

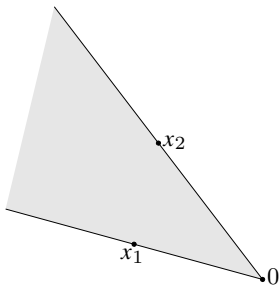
### Examples

- solution set of linear equations  $\{x \mid Ax = b\}$  is affine
- every affine set can be expressed as solution set of linear equations
- the empty set, any single point (singleton), and  $\mathbb{R}^n$  are affine, hence convex
- a line  $\mathcal{L} = \{x_0 + tv \mid t \in \mathbb{R}\}$  with  $x_0, v \in \mathbb{R}^n$  and  $v \neq 0$  is affine and convex

## Convex cones and rays

**Convex cone:**  $C \subseteq \mathbb{R}^n$  is a *convex cone* if for every  $x, y \in C$ ,

$$\theta_1 x + \theta_2 y \in C \quad \text{for all } \theta_1, \theta_2 \geq 0$$



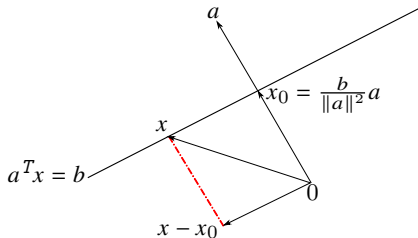
- a point  $\theta_1 x + \theta_2 y$  with  $\theta_1, \theta_2 \geq 0$  is called a *conic (nonnegative) combination*
- an example of a convex cone is the *norm cone*:  $\{(x, t) \mid \|x\| \leq t\} \subseteq \mathbb{R}^{n+1}$
- called *second-order cone* for Euclidean norm, *i.e.*,

$$\{(x, t) \mid \|x\|_2 \leq t\} = \{(x, t) \mid \|x\|_2^2 \leq t^2, t \geq 0\}$$

**Rays:**  $\{x_0 + tv \mid t \geq 0\}$  with  $v \neq 0$ , is convex (not affine); it is a convex cone if  $x_0 = 0$

# Hyperplane

a hyperplane  $\mathcal{H} = \{x \in \mathbb{R}^n \mid a^T x = b\}$  with  $a \neq 0$  is affine and convex



- $a$  is called the *normal vector*
- for any  $x_0 \in \mathcal{H}$  (e.g.,  $x_0 = (b/\|a\|^2)a$ ),  $x \in \mathcal{H}$  if and only if  $x - x_0 \perp a$ :

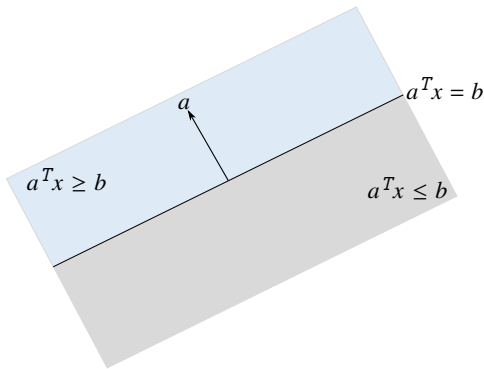
$$a^T x = b = a^T x_0 \implies a^T (x - x_0) = 0$$

## Halfspaces

the hyperplane  $\{x \in \mathbb{R}^n \mid a^T x = b\}$  divides  $\mathbb{R}^n$  in two *halfspaces*

$$\mathcal{H}^- = \{x \in \mathbb{R}^n \mid a^T x \leq b\} \quad \text{and} \quad \mathcal{H}^+ = \{x \in \mathbb{R}^n \mid a^T x \geq b\}$$

a halfspace is convex



## Balls and ellipsoids

**Balls:** for  $x_c \in \mathbb{R}^n$ ,  $r > 0$ , and  $\|\cdot\|$  an arbitrary norm, the open and closed balls

$$\mathcal{B}(x_c, r) = \{x \mid \|x - x_c\| < r\} = \{x_c + ru \mid \|u\| < 1\}$$

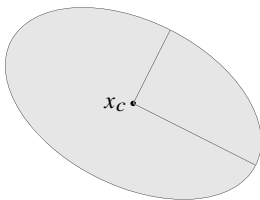
$$\mathcal{B}[x_c, r] = \{x \mid \|x - x_c\| \leq r\} = \{x_c + ru \mid \|u\| \leq 1\}$$

are convex

**Ellipsoids:** an ellipsoid

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

is convex with  $Q \in \mathbb{S}_{++}^n$  positive definite,  $r \in \mathbb{R}^n$ , and  $c \in \mathbb{R}$



also written as  $\{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$  with  $P \in \mathbb{S}_{++}^n$  and center  $x_c \in \mathbb{R}^n$



## Linear matrix inequality

a *linear matrix inequality* (LMI) has the form

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

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

**Example** any solution  $w(t)$  to the linear differential equation

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

converges to the origin iff there exists a real symmetric matrix  $X$  satisfying:

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

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  $X_i$  ( $i = 1, 2, \dots, m$ ) are basis for subspace 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:

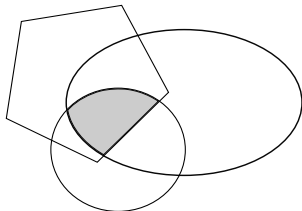
$$F(x) \triangleq \begin{bmatrix} -X & 0 \\ 0 & AX + XA^T \end{bmatrix} < 0,$$

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, scaling, summation

**Intersection:** the intersection of any collection of convex sets is convex



**Scaling:** if  $C$  is a convex set and  $\beta$  is a real number, then the set

$$\beta C = \{\beta y \mid y \in C\} \text{ is also convex}$$

**Summation:** if  $C_1$  and  $C_2$  are convex sets, then the set

$$C_1 + C_2 = \{x_1 + x_2 \mid x_1 \in C_1, x_2 \in C_2\} \text{ is convex}$$

## Affine transformation

let  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$  and let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be the affine function

$$f(x) = Ax + b$$

- the image of a convex set  $C \subseteq \mathbb{R}^n$  under  $f$  is convex

$$C \subseteq \mathbb{R}^n \text{ convex} \implies f(C) = \{Ax + b \mid x \in C\} \text{ is convex}$$

- the inverse image  $f^{-1}(C)$  of a convex set under  $f$  is convex

$$C \subseteq \mathbb{R}^m \text{ convex} \implies f^{-1}(C) = \{x \in \mathbb{R}^n \mid Ax + b \in C\} \text{ is convex}$$

## Examples

- the image of norm ball under affine transformation

$$\{Ax + b \mid \|x\| \leq 1\}$$

- for example, an ellipsoid

$$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\} = \{P^{1/2}u + x_c \mid \|u\|_2 \leq 1\}$$

is the image of the unit Euclidean ball  $\{u \mid \|u\|_2 \leq 1\}$  via  $f(u) = P^{1/2}u + x_c$

- the inverse image of norm ball under affine transformation

$$\{x \mid \|Ax + b\| \leq 1\}$$

- hyperbolic cone

$$\{x \mid x^T P x \leq (c^T x)^2, \ c^T x \geq 0\} \quad \text{with } P \in \mathbb{S}_+^n$$

- inverse image of 2nd cone  $\{(z, t) \mid z^T z \leq t^2, t \geq 0\}$  under  $f(x) = (P^{1/2}x, c^T x)$

- solution set of linear matrix inequality

$$\{x \mid x_1 A_1 + \cdots + x_m A_m \leq B\} \quad \text{with } A_i, B \in \mathbb{S}^p$$

# Outline

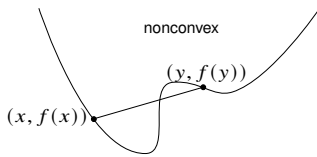
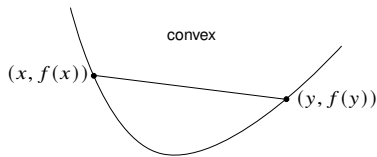
- convex sets
- **convex functions**
- operations preserving convexity
- basic properties
- convex problems

## Definition

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

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) \quad (9.3)$$

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



- $f$  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  $x, y \in \mathcal{X}$

## Examples

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

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

- *norm functions*: any norm  $\| \cdot \|$  is convex:

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

where the inequality follows from the triangle inequality

- $f(x) = x^T Q x$  with  $Q \in \mathbb{S}^n$  and convex  $\text{dom } f$  is convex if

$$(x - y)^T Q (x - y) \geq 0 \quad \text{for all } x, y \in \text{dom } f$$



- the function

$$f(x_1, x_2) = x_1 x_2 \quad \text{with} \quad \text{dom } f = \{x \mid x_1, x_2 \geq 0\}$$

is nonconvex since for  $x = (1, 2)$ ,  $y = (2, 1)$ ,  $\theta = 0.5$ , we have

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

which violates the definition of convexity

- the function

$$f(x) = x \quad \text{over} \quad \text{dom } f = \{x \mid x \neq 1\}$$

is not convex even though it is linear; this is because its domain is nonconvex

## Extended-value extension

extended-value extension  $\tilde{f}$  of  $f$  is

$$\tilde{f}(x) = f(x), \quad x \in \text{dom } f, \quad \tilde{f}(x) = \infty, \quad x \notin \text{dom } f$$

often simplifies notation; for example, the condition

$$0 \leq \theta \leq 1 \implies \tilde{f}(\theta x + (1 - \theta)y) \leq \theta \tilde{f}(x) + (1 - \theta) \tilde{f}(y)$$

(as an inequality in  $\mathbb{R} \cup \{\infty\}$ ), means the same as the two conditions

- $\text{dom } f$  is convex
- for  $x, y \in \text{dom } f$ ,

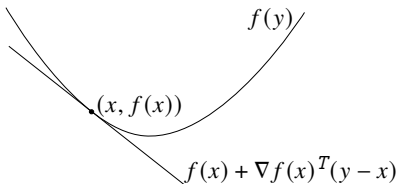
$$0 \leq \theta \leq 1 \implies f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

## First-order convexity condition

suppose  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable (with open domain)

$f$  is convex if and only if its domain is convex and for any  $x, y \in \text{dom } f$

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



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

## Second-order convexity condition

suppose that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is twice differentiable (with open domain)

$f$  is convex if and only if its domain is convex and

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \text{dom } f \quad (9.5)$$

- if  $\nabla^2 f(x) \succ 0$  for all  $x$ , then  $f$  is strictly convex
- converse is not true (e.g.,  $f(x) = x^4$  is strictly convex but  $f''(x) = 0$  at  $x = 0$ )

### Convexity of domain

- $\text{dom } f$  must be convex to use the first or second order convexity characterization
- for example, the function

$$f(x) = 1/x^2 \quad \text{with} \quad \text{dom } f = \{x \in \mathbb{R} \mid x \neq 0\}$$

satisfies  $f''(x) = 6/x^4 > 0$  for all  $x \in \text{dom } f$ , but is not a convex function

## Examples

the following can be shown using the definition or the second order condition

### Convex

- *exponential*:  $e^{\alpha x}$  is convex for any  $\alpha \in \mathbb{R}$
- *powers*:  $x^\alpha$  is convex on  $\mathbb{R}_{++}$  when  $\alpha \geq 1$  or  $\alpha \leq 0$
- *powers of absolute value*:  $|x|^p$  is convex on  $\mathbb{R}$  for  $p \geq 1$
- *negative entropy*:  $x \log x$  defined as 0 for  $x = 0$  is convex on  $\mathbb{R}_+ = \{x \mid x \geq 0\}$

### Concave

- *powers*:  $x^\alpha$  is concave for  $0 \leq \alpha \leq 1$
- *logarithm*:  $\log x$  is concave on  $\mathbb{R}_{++}$

## Example: quadratic functions

$$f(x) = x^T Q x + r^T x + c \quad \text{with } Q = Q^T$$

is convex if and only if  $Q \geq 0$

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

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

is positive definite

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

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

is indefinite

## Example: quadratic over linear

the function

$$f(x, t) = x^2/t \quad \text{with} \quad \text{dom } f = \{(x, t) \mid t > 0\}$$

is convex

this is because the Hessian

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

over its domain

## Example: log-sum-exp function

the function  $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$  is convex over  $\mathbb{R}^n$

- 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

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

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



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

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

- applying Cauchy-Schwarz on the vectors  $a$  and  $b$  with entries

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

we get

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

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

# Outline

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

## Operations that preserves convexity

### Weighted nonnegative sum

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

- $f$  convex if  $f_i$  are convex and  $w_i \geq 0$
- a nonnegative weighted sum of concave functions is concave
- +ve weighted sum of strictly convex (concave)  $f_i$  is strictly convex (concave)

**Composition with affine mapping:** for  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , consider

$$f(x) = g(Ax + b), \quad \text{with} \quad \text{dom } f = \{x \mid Ax + b \in \text{dom } g\}$$

$f$  is convex (concave) if  $g$  is convex (concave)

## Example

- *negative entropy function*

$$f(x) = \sum_{i=1}^n x_i \log x_i, \quad \text{dom } f = \mathbb{R}_{++}^n = \{x \mid x_i > 0\}$$

$f$  is convex since it is the sum of convex functions  $x_i \log x_i$

- logarithmic barrier for linear inequalities

$$f(x) = - \sum_{i=1}^m \log(b_i - a_i^T x), \quad \text{dom } f = \{x \mid a_i^T x < b_i, i = 1, \dots, m\}$$

is convex since it is a sum of convex functions

- for  $a \in \mathbb{R}^n$  and  $b \in \mathbb{R}$

$$f(x) = e^{a^T x + b}$$

is convex over  $\mathbb{R}^n$  since  $f(x) = g(a^T x + b)$  where  $g(t) = e^t$  is a convex function

- the function

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

is convex it is the sum of two convex functions  $f = f_1 + f_2$  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  $f_2(x_1, x_2) = g(x_1)$

- 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:  $e^{x_1 - x_2 + x_3}$ ,  $e^{2x_2}$ ,  $x_1$

## Example: generalized quadratic-over-linear

let  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c \in \mathbb{R}^n$  ( $c \neq 0$ ), and  $d \in \mathbb{R}$ , then the function

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

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

- we can write  $f$  as

$$f(x) = g(Ax + b, c^T x + d), \quad g(y, t) = \frac{\|y\|^2}{t} = \sum_{i=1}^m \frac{y_i^2}{t}$$

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

- $g$  is sum of convex functions  $g_i(y, t) = \frac{y_i^2}{t}$  over  $\{(y_i, t) \mid y_i \in \mathbb{R}, t > 0\}$
- thus  $f$  is convex (composition of convex function with an affine mapping)

## Pointwise maximum

the max of convex functions  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, k$

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

is convex

### Examples

- the piece-wise linear function  $f(x) = \max_{i=1, \dots, k} \{a_i^T x + b_i\}$  is convex
- sum of k largest values*: let  $x_{[i]}$  denote the  $i$ th largest component of  $x$ , then

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

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

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

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

## Pointwise supremum

if  $f(x, y)$  is convex in  $x$  for each  $y \in \mathcal{A}$ , then the function

$$g(x) = \sup_{y \in \mathcal{A}} f(x, y)$$

is convex

### Examples

- the distance to farthest point in a set  $C$ :

$$\sup_{y \in C} \|x - y\|$$

is convex

- the maximum eigenvalue of symmetric matrix  $X \in \mathbb{S}$ :

$$\lambda_{\max}(X) = \sup_{\|y\|=1} y^T X y$$

is convex



## Partial minimization

if  $f(x, y)$  is convex in  $(x, y)$  and  $C$  is a convex set, then the function

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

is convex (provided that  $g(x) > -\infty$  for some  $x$ )

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

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

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

## Summary of minimization/maximization rules

if we include the counterparts for concave functions, there are four rules

### Maximization

$$g(x) = \sup_{y \in C} f(x, y)$$

- $g$  is convex if  $f$  is convex in  $x$  for fixed  $y$ ;  $C$  can be any set
- $g$  is concave if  $f$  is jointly concave in  $(x, y)$  and  $C$  is a convex set

### Minimization

$$g(x) = \inf_{y \in C} f(x, y)$$

- $g$  is convex if  $f$  is jointly convex in  $(x, y)$  and  $C$  is a convex set
- $g$  is concave if  $f$  is concave in  $x$  for fixed  $y$ ;  $C$  can be any set

## Composition with scalar functions

composition of  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $g : \mathbb{R} \rightarrow \mathbb{R}$ :

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

$f$  is convex if  $g$  is convex and one of the following three cases holds

- $h$  is convex, and  $\tilde{g}$  is nondecreasing
- $h$  is concave, and  $\tilde{g}$  is nonincreasing
- $g$  is affine

$f$  is concave if  $g$  is concave and one of the following three cases holds

- $h$  is concave, and  $\tilde{g}$  is nondecreasing
- $h$  is convex, and  $\tilde{g}$  is nonincreasing
- $g$  is affine

## Proof

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

- the first inequality arises from convexity of  $h$  and the nondecreasing nature of  $g$
- the second inequality is a result of the convexity of  $g$

## Examples

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

## Vector functions composition

composition of  $h : \mathbb{R}^n \rightarrow \mathbb{R}^k$  and  $g : \mathbb{R}^k \rightarrow \mathbb{R}$ :

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

$f$  is convex if  $g$  is convex and for each  $i$ , one of the following holds

- $h_i$  is convex and  $\tilde{g}$  nondecreasing in its  $i$ th argument
- $h_i$  is concave and  $\tilde{g}$  nonincreasing in its  $i$ th argument
- $h_i$  is affine

## Examples

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

$$\left( \sum_{i=1}^k h_i(x)^p \right)^{\frac{1}{p}}$$

is convex for  $p \geq 1$  and convex and nonnegative  $h_1, \dots, h_k$

to demonstrate this, consider the function  $g : \mathbb{R}^k \rightarrow \mathbb{R}$

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

- $g(h(x))$  is convex since  $g$  is both convex and nondecreasing in its arguments
- for nonnegative values of  $z$ ,  $g(z)$  simplifies to

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

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

## Example

- $f(x) = \sum_{i=1}^k \log h_i(x)$  is concave if  $h_i$  are concave and positive
- the function

$$f(x, y) = \frac{(x - y)^2}{1 - \max(x, y)}, \quad x < 1, \quad y < 1$$

is convex

- (leaves)  $x$ ,  $y$ , and  $1$  are affine
- $\max(x, y)$  is convex;  $x - y$  is affine
- $1 - \max(x, y)$  is concave
- function  $u^2/v$  is convex, monotone decreasing in  $v$  for  $v > 0$
- $f$  is compos. of  $g(u, v) = \frac{u^2}{v}$  with  $u = x - y, v = 1 - \max(x, y)$ , hence convex



## Perspective function

the *perspective* of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is the function  $g : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$g(x, t) = tf(x/t), \quad \text{dom } g = \{(x, t) \mid x/t \in \text{dom } f, t > 0\}$$

$g$  is convex if  $f$  is convex

### Examples

- $f(x) = x^T x$  is convex, so  $g(x, t) = x^T x/t$  is convex for  $t > 0$
- negative logarithm  $f(x) = -\log x$  is convex, so the relative entropy

$$g(x, t) = t \log t - t \log x$$

is convex on  $\mathbb{R}_{++}^2$

- if  $f$  is convex, then

$$g(x) = (c^T x + d)f((Ax + b)/(c^T x + d))$$

is convex on  $\{x \mid c^T x + d > 0, (Ax + b)/(c^T x + d) \in \text{dom } f\}$

# Outline

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

## Restriction of a convex function to a line

$f : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex if and only if

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

is convex in  $t$  for any  $x \in \text{dom } f$  and  $v \in \mathbb{R}^n$

- $f$  convex if it remains convex when restricted to any line intersecting its domain
- allows us to check convexity of  $f$  by checking convexity of one variable functions

## Example: log-determinant function

$f : \mathbb{S}^n \rightarrow \mathbb{R}$  with  $f(X) = -\log \det X$  is convex over  $\text{dom } f = \mathbb{S}_{++}^n$

### Proof

- let  $X_0 = X_0^{1/2} X_0^{1/2} \in \mathbb{S}_{++}^n$ ,  $V \in \mathbb{R}^{n \times n}$  be symmetric, then

$$\begin{aligned} g(t) &= -\log \det(X_0 + tV) = \log \det(X_0^{1/2} X_0^{1/2} + tV) \\ &= \log \det X_0 + \log \det(I + tX_0^{-1/2} V X_0^{-1/2}) \\ &= \log \det X_0 + \log \prod_i (1 + t\lambda_i) \\ &= \log \det X_0 + \log \sum_{i=1}^n \log(1 + t\lambda_i) \end{aligned}$$

where  $\lambda_i$ , are the eigenvalues of  $X_0^{-1/2} V X_0^{-1/2}$

- 2nd term is sum of convex functions; hence  $g(t)$  is convex and  $f$  is convex

## Sublevel sets and convexity

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

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

- for a convex function  $f$ , the sublevel set  $\mathcal{S}_\gamma$  is also convex:

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

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

- a function can have all its sublevel sets convex, but not be a convex
  - for example,  $f(x) = -e^x$  is not convex on  $\mathbb{R}$  but all its sublevel sets are convex
  - another example is  $f(x) = \log(x)$ , which is concave; with convex sublevel sets  $(0, e^\gamma]$

## Example

let  $P \succeq 0$  is an  $n \times n$  matrix, then the set:

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

is convex since it is the level set of a convex function

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

- the log-sum-exp function, previously established as convex
- $(x^T P x + 1)^2$  is convex since it is equal  $g(x^T P x)$  with  $g(t) = (t + 1)^2$ 
  - $g$  is nondecreasing convex function (defined on  $\mathbb{R}_+$ )
  - $x^T P x$  convex quadratic function
  - convexity follows from composition rule
- $f$  is convex, being the sum of two convex functions

# Epigraph

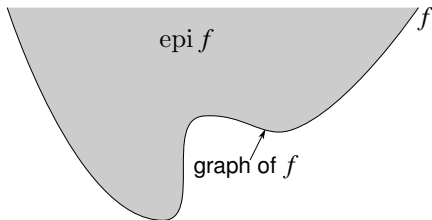
the *graph* of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is the set

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

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

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

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



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

## Example

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

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

where  $Y$  is positive definite

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

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

last line follows from Schur complement criteria for positive semidefiniteness

- the latter condition is an LMI in the variables  $(x, Y, t)$
- hence the epigraph of  $f$  is convex, and consequently  $f$  is convex



# Outline

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

## Definition

### Convex optimization problem in standard form

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

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

### Concave problems

- maximization with concave objective and convex constraints
- a concave problem is also referred to as a convex problem

## Examples

- 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  $h(x) = x_1^2 + x_2^2 - 4$  is not affine

## Example

$$\begin{array}{ll}\text{minimize} & f(x) = x_1^2 + x_2^2 \\ \text{subject to} & g_1(x) = x_1/(1+x_2^2) \leq 0 \\ & h_1(x) = (x_1+x_2)^2 = 0\end{array}$$

- problem has convex objective  $f$
- the feasible set  $\{(x_1, x_2) \mid x_1 = -x_2 \leq 0\}$  is convex
- for our definition, this is not a convex problem ( $g_1$  not convex and  $h_1$  not affine)
- problem is equivalent (but not identical) to the convex problem:

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

## Example

- an investor wants to invest a total value of at most  $d$  into  $n$  possible investments
- let  $x_i$  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

- formulation: determine the investment deposits that maximize expected profit

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

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

## Convexity of feasible and optimal set

- feasible set is convex since it is the intersection of convex sets:

$\text{dom } f$ , sublevel sets  $\{x \mid g_i(x) \leq 0\}$ , and affine sets  $\{x \mid a_j^T x = b_j\}$

- optimal set is convex: any convex combination of optimal  $x_1, x_2$  is feasible and

$$f(\theta x_1 + (1 - \theta)x_2) \leq \theta f(x_1) + (1 - \theta)f(x_2) = p^\star$$

so  $f(\theta x_1 + (1 - \theta)x_2) = p^\star$ , i.e., any convex combination is optimal

## Local minimizers are global minimizers

any locally optimal point of a convex problem is (globally) optimal

### Proof

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

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

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

## First-order optimality condition

- suppose  $f : \mathcal{X} \rightarrow \mathbb{R}$  is convex a convex set  $\mathcal{X} \subset \mathbb{R}^n$
- the point  $x^\star$  is optimal if and only if

$$\nabla f(x^\star)^T(y - x^\star) \geq 0, \quad \forall y \in \mathcal{X} \quad (9.6)$$

(the above condition is difficult to verify in practice)

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

$$\nabla f(x^\star) = 0$$

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

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

hence,  $\nabla f(x) = 0$



## Examples

- $f(x) = x \log x$  with  $\text{dom } f = \mathbb{R}_{++}$ ; setting the derivative to zero

$$f'(x) = \log x + 1 = 0$$

gives the stationary point  $x = 1/e$ ; the second derivative is

$$f''(x) = 1/x > 0 \quad \text{for all } x \in \text{dom } f$$

hence, the function is convex and  $x = 1/e$  is a global minimizer

- *minimization over the nonnegative orthant*

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \geq 0 \end{array}$$

using the optimality condition:

$$x \geq 0, \quad \nabla f(x)^T(y - x) \geq 0 \text{ for all } y \geq 0$$

equivalent to

$$x \geq 0, \quad \nabla f(x) \geq 0, \quad x_i \nabla f(x)_i = 0, \quad i = 1, \dots, n$$

## Sufficiency of KKT conditions

for cvx problems, if there exists  $x^\star \in \mathcal{D}$ ,  $\mu^\star \in \mathbb{R}^m$ ,  $\lambda^\star \in \mathbb{R}^p$  satisfying

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

$$g_i(x^\star) \leq 0, \quad i = 1, \dots, m$$

$$Ax^\star = b$$

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

$$g_i(x^\star) \mu_i^\star = 0, \quad i = 1, \dots, m$$

then,  $x^\star$  is a global minimizer

- there may be optimal points that do not satisfy KKT conditions
- when we discuss duality, we will provide conditions such that the KKT conditions are both necessary and sufficient

## Proof

- let  $x$  be a feasible solution; note that the function

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

is convex since it is the sum of convex functions

- since  $\nabla J(x^*) = 0$ ,  $x^*$  is a minimizer of  $J$  over  $\mathbb{R}^n$ ; thus,

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

- hence,  $x^*$  is optimal

## Example

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

the above problem is convex with an equality constraint; the Lagrangian is

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

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  $x = (1, 1, 1)$  and  $\lambda = -1$

## Example

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

it is easy to see that the solution is  $x^\star = (0, 0)$ ; for this the Lagrangian is

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

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

## References and further readings

- S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- E. K.P. Chong, Wu-S. Lu, and S. H. Zak. *An Introduction to Optimization: With Applications to Machine Learning*. John Wiley & Sons, 2023.
- A. Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with Python and MATLAB*. SIAM, 2023.