# Review of Prerequisite Mathematics

# Introduction to Elementary Optimization

## Definition (The Jacobian)

Let $V$ be open in $\mathbb{R}^n$, $\mathbf{x} \in V$, and $g_i : V \to \mathbb{R}$, $i = 1, 2, \dots, m$ be $C^1$ on $V$. The Jacobian of $\mathbf{g}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is defined as

$$\mathrm{D}\,\mathbf{g}(\mathbf{x}) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \dots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \dots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \dots & \frac{\partial g_m}{\partial x_n} \end{pmatrix} (\mathbf{x})$$

## Theorem (The Chain Rule)

Suppose that $\mathbf{f}$ and $\mathbf{g}$ are vector functions. If $\mathbf{g}$ is differentiable at $\mathbf{a}$ and $\mathbf{f}$ is differentiable at $\mathbf{g}(\mathbf{a})$, then $\mathbf{f} \circ \mathbf{g}$ is differentiable at $\mathbf{a}$ and

$$\mathrm{D}(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = \mathrm{D}\,\mathbf{f}(\mathbf{g}(\mathbf{a}))\ \mathrm{D}\,\mathbf{g}(\mathbf{a})$$

More explicitly, if $f$ is a differentiable function of $x_1, x_2, \dots, x_n$, and each $x_j$ is a differentiable function of $t_1, t_2, \dots, t_m$, $n, m \geqslant 1$; then $f$ is a differentiable function of $t_1, t_2, \dots, t_m$ with

$$\frac{\partial f}{\partial t_i} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t_i}$$

## Example

Let $w = f(xz, yz)$, where $f$ is a differentiable function. Prove that
$x\dfrac{\partial w}{\partial x} + y\dfrac{\partial w}{\partial y} = z\dfrac{\partial w}{\partial z}$.

# Solution I

Write $u(x,y,z) = xz$ and $v(x,y,z) = yz$ so that $w(x,y,z) = f\big(u(x,y,z), v(x,y,z)\big)$. By the chain rule,

$$\frac{\partial w}{\partial x}(x,y,z) = \frac{\partial}{\partial x}\big[f\big(u(x,y,z), v(x,y,z)\big)\big]$$

$$= \frac{\partial f}{\partial u}\big(u(x,y,z), v(x,y,z)\big)\frac{\partial u}{\partial x}(x,y,z) + \frac{\partial f}{\partial v}\big(u(x,y,z), v(x,y,z)\big)\frac{\partial v}{\partial x}(x,y,z)$$

$$= z\frac{\partial f}{\partial u}(xz, yz)$$

$$\frac{\partial w}{\partial y}(x,y,z) = \frac{\partial}{\partial y}\big[f\big(u(x,y,z), v(x,y,z)\big)\big]$$

$$= \frac{\partial f}{\partial u}\big(u(x,y,z), v(x,y,z)\big)\frac{\partial u}{\partial y}(x,y,z) + \frac{\partial f}{\partial v}\big(u(x,y,z), v(x,y,z)\big)\frac{\partial v}{\partial y}(x,y,z)$$

$$= z\frac{\partial f}{\partial v}(xz, yz)$$

# Solution II

$$
\begin{aligned}
&\frac{\partial w}{\partial z}(x,y,z) = \frac{\partial}{\partial z}\left[f(u(x,y,z),v(x,y,z))\right] \\
&= \frac{\partial f}{\partial u}(u(x,y,z),v(x,y,z))\frac{\partial u}{\partial z}(x,y,z) + \frac{\partial f}{\partial v}(u(x,y,z),v(x,y,z))\frac{\partial v}{\partial z}(x,y,z) \\
&= x\frac{\partial f}{\partial u}(xz,yz) + y\frac{\partial f}{\partial v}(xz,yz)
\end{aligned}
$$

So

$$
\begin{aligned}
\frac{\partial w}{\partial x} + y\frac{\partial w}{\partial y} &= xz\frac{\partial f}{\partial u}(xz,yz) + yz\frac{\partial f}{\partial v}(xz,yz) \\
&= z\left[x\frac{\partial f}{\partial u}(xz,yz) + y\frac{\partial f}{\partial v}(xz,yz)\right] = z\frac{\partial w}{\partial z}
\end{aligned}
$$

# Unconstrained Optimization Problems

## Theorem

Given $S \subseteq \mathbb{R}^n$ and continuous $f : S \to \mathbb{R}$; if $S$ is compact, then

$$M = \sup \{f(\mathbf{x}) : \mathbf{x} \in S\} \quad \text{and} \quad m = \inf \{f(\mathbf{x}) : \mathbf{x} \in S\}$$

are finite real numbers. Moreover, there exists points $\mathbf{x}_\mathsf{M}$, $\mathbf{x}_\mathsf{m} \in S$ such that $M = f(\mathbf{x}_\mathsf{M})$ and $m = f(\mathbf{x}_\mathsf{m})$.

## Definition

Given $S \subseteq \mathbb{R}^n$, $f : S \to \mathbb{R}$ and $B(\mathbf{x}, h) \equiv \{\mathbf{y} \in \mathbb{R}^n \mid |\mathbf{y} - \mathbf{x}| < h\}$, $f$ achieves

- global maximum $f(\mathbf{x}_\mathsf{M})$ at $\mathbf{x}_\mathsf{M} \in S$: $f(\mathbf{x}_\mathsf{M}) \geqslant f(\mathbf{x})$, $\forall\, \mathbf{x} \in S$.
- global minimum $f(\mathbf{x}_\mathsf{m})$ at $\mathbf{x}_\mathsf{m} \in S$: $f(\mathbf{x}_\mathsf{m}) \leqslant f(\mathbf{x})$, $\forall\, \mathbf{x} \in S$.
- local maximum $f(\mathbf{x}_0)$ at $\mathbf{x}_0 \in S$: $\exists\, h_0 > 0$ s.t. $f(\mathbf{x}_0) \geqslant f(\mathbf{x})$, $\forall\, \mathbf{x} \in B(\mathbf{x}_0, h_0) \cap S$.
- local minimum $f(\mathbf{x}_1)$ at $\mathbf{x}_1 \in S$: $\exists\, h_1 > 0$ s.t. $f(\mathbf{x}_1) \leqslant f(\mathbf{x})$, $\forall\, \mathbf{x} \in B(\mathbf{x}_1, h_1) \cap S$.

## Theorem (necessary conditions for extremum)

Given $S \subseteq \mathbb{R}^n$ and differentiable $f : S \to \mathbb{R}$, if $f$ achieves extremum at an interior $\mathbf{c} \in S$, then $\nabla f(\mathbf{c}) = \mathbf{0}$.

## Proof

If $\mathbf{c} = (c_1, c_2, \dots, c_n)$, let

$$g_j(t) \equiv f(c_1, c_2, \dots, c_{j-1}, t, c_{j+1}, \dots, c_n), \quad j = 1, 2, \dots, n$$

For $f$ achieves extremum at $\mathbf{c}$, $f(\mathbf{c}) = g_j(c_j)$, $g_j$ achieves extremum at $c_j \implies g_j'(t)\big|_{t=c_j} = 0 \implies D_j f(\mathbf{c}) = 0 \; \forall \, j$, so $\nabla f(\mathbf{c}) = \mathbf{0}$.

## Theorem

Given $S \subseteq \mathbb{R}^n$, if $f : S \to \mathbb{R}$ achieves extremum at $\mathbf{c} \in S$, then $\mathbf{c}$ can possibly be a

- critical point: $\nabla f(\mathbf{c}) = \mathbf{0}$.
- singular point: $f$ is non-differentiable at $\mathbf{c}$.
- boundary point of $S$.

## Definition (Hessian Matrix)

Given $S \subseteq \mathbb{R}^n$, an interior point $\mathbf{c}$ of $S$, and a differentiable function $f : S \to \mathbb{R}$,

$$\mathbf{H}(f, \mathbf{c}) = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{pmatrix}, \quad f_{ij} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{c}), \quad i, j = 1, 2, \ldots, n.$$

## Definition (Matrix Positive/Negative Definiteness)

Given an $n \times n$ real symmetric matrix $\mathbf{A}$. For any $\mathbf{v} \in \mathbb{R}^n \neq \mathbf{0}$, $\mathbf{A}$ is

- positive-definite: $\mathbf{v}\mathbf{A}\mathbf{v}^\top > 0$
- negative-definite: $\mathbf{v}\mathbf{A}\mathbf{v}^\top < 0$
- positive-semidefinite: $\mathbf{v}\mathbf{A}\mathbf{v}^\top \geqslant 0$
- negative-semidefinite: $\mathbf{v}\mathbf{A}\mathbf{v}^\top \leqslant 0$

## Theorem

Given a real symmetric matrix $\mathbf{A}$.

- $\mathbf{A}$ is positive-definite $\iff$ all eigenvalues of $\mathbf{A}$ are positive.
- $\mathbf{A}$ is positive-semidefinite $\iff$ all eigenvalues of $\mathbf{A}$ are nonnegative.

## Definition (Minor)

Given an $n \times n$ matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ and minor

$$\mathbf{A}\begin{pmatrix} i_1, i_2, \cdots, i_k \\ j_1, j_2, \cdots, j_k \end{pmatrix} = \begin{vmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \cdots & a_{i_1 j_k} \\ a_{i_2 j_1} & a_{i_2 j_2} & \cdots & a_{i_2 j_k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_k j_1} & a_{i_k j_2} & \cdots & a_{i_k j_k} \end{vmatrix}, 1 \leqslant k \leqslant n,$$

$1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$, $1 \leqslant j_1 < j_2 < \cdots < j_k \leqslant n$.

- $\Delta_k \equiv \mathbf{A}\begin{pmatrix} i_1, i_2, \cdots, i_k \\ i_1, i_2, \cdots, i_k \end{pmatrix}$ is the $k$-th order principal minor of $A$.
- $M_k \equiv \mathbf{A}\begin{pmatrix} 1, 2, \cdots, k \\ 1, 2, \cdots, k \end{pmatrix}$ is the $k$-th order leading principal minor of $A$.

## Theorem (Criteria for Matrix Positive/Negative Definiteness)

Given an $n \times n$ real symmetric matrix $\mathbf{A}$, then $\forall\, k \leqslant n$, $\mathbf{A}$ is

- positive-definite $\iff M_k > 0$
- negative-definite $\iff (-1)^k M_k > 0$
- positive-semidefinite $\iff \Delta_k \geqslant 0$
- negative-semidefinite $\iff (-1)^k \Delta_k \geqslant 0$

# Equalities Constrained Optimization: The Lagrange Multipliers Method

## Theorem

Given an open set $S \subseteq \mathbb{R}^n$, differentiable functions $f : S \to \mathbb{R}$ and $g_j : S \to \mathbb{R}$, $j = 1, 2, \ldots, m$, $m < n$, and $X_0 = \{\mathbf{x} \in S \mid g_j(\mathbf{x}) = 0, \, j = 1, 2, \ldots, m\}$. If $f$ has an extremum at $\mathbf{x}_0 \in S \cap X_0$ and $\det\left(D_i g_j(\mathbf{x}_0)\right) \neq 0$, then

$$\exists \, \lambda_1, \lambda_2, \ldots, \lambda_m \quad \text{s.t.} \quad D_i f(\mathbf{x}_0) + \sum_{j=1}^{m} \lambda_j D_i g_j(\mathbf{x}_0) = 0, \quad i = 1, 2, \ldots, n$$

## Remark

Let $\mathcal{L} \equiv f + \sum_{j=1}^{m} \lambda_j \, g_j$, the sufficient condition can be rewritten as

$$\begin{aligned} D_i \, \mathcal{L}(\mathbf{x}_0) &= 0, \quad i = 1, 2, \ldots, n \\ g_j(\mathbf{x}_0) &= 0, \quad j = 1, 2, \ldots, m \end{aligned}$$

## Example

Find the maximum and minimum values of $x^2 - 10x - y^2$ on $x^2 + 4y^2 = 16$.

## Solution

Let $\mathcal{L} = x^2 - 10x - y^2 + \lambda\left(x^2 + 4y^2 - 16\right)$, then

$$\frac{\partial \mathcal{L}}{\partial x} = 2x - 10 + 2\lambda x = 0 \implies x - 5 + \lambda x = 0 \tag{1}$$

$$\frac{\partial \mathcal{L}}{\partial y} = -2y + 8\lambda y = 0 \implies -y + 4\lambda y = 0 \tag{2}$$

$$x^2 + 4y^2 - 16 = 0 \tag{3}$$

From (2) $(1 - 4\lambda)y = 0$, so $y = 0 \ \vee \ \lambda = \frac{1}{4}$. If $y = 0$, from (3) $x = \pm 4$; if $\lambda = \frac{1}{4}$, from (1) $(1 + \lambda)\, x = 5 \implies x = 4$, substituting into (3) gives $y = 0$. Therefore, the extremum points are $(x, y) = (4, 0)$, $(-4, 0)$; $x^2 - 10x - y^2$ has a maximum value of $56$ (at $(x, y) = (-4, 0)$), and a minimum value of $-24$ (at $(x, y) = (4, 0)$).

## Example

Find the maximum and minimum values of $f(x,y,z) = (x+z)\,e^y$ on $x^2+y^2+z^2=6$.

## Solution

Let $\mathcal{L} = (x+z)\,e^y + \lambda\,(x^2+y^2+z^2-6)$, then

$$\frac{\partial \mathcal{L}}{\partial x} = e^y + 2\lambda x = 0 \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial y} = (x+z)\,e^y + 2\lambda y = 0 \tag{5}$$

$$\frac{\partial \mathcal{L}}{\partial z} = e^y + 2\lambda z = 0 \tag{6}$$

$$x^2+y^2+z^2-6 = 0 \tag{7}$$

From (4), (6) $2\,\lambda(x-z)=0$, so $\lambda = 0 \;\vee\; x = z$. If $\lambda = 0$, then from (4) $e^y = 0$ which is impossible, so $x = z$. From (4) $e^y = -2\lambda x$, substituting into (5) $2x\,(-2\lambda x) + 2\lambda y = 0 \implies y = 2x^2$, substituting into (7) gives $x^2 + 4x^4 + x^2 = 6 \implies (4x^2+6)(x^2-1) = 0 \implies x = \pm 1$. Therefore, the extremum points are $(x,y,z) = (1,2,1),\ (-1,2,-1)$; $(x+z)\,e^y$ has a maximum value of $2e^2$ (at $(x,y,z) = (1,2,1)$), and a minimum value of $-2e^2$ (at $(x,y,z) = (-1,2,-1)$).

## Example

If $L$ is the curve of intersection of $z^2 = x^2 + y^2$ and $x - 2z = 3$, find the point on $L$ that is closest to the origin and the shortest distance.

## Solution

The objective is $x^2 + y^2 + z^2$ with constraints $x^2 + y^2 - z^2 = 0$ and $x - 2z - 3 = 0$. Let $\mathcal{L} = x^2 + y^2 + z^2 + \lambda_1 (x^2 + y^2 - z^2) + \lambda_2 (x - 2z - 3)$, then

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + 2\lambda_1 x + \lambda_2 = 0 \implies 2(1 + \lambda_1)x + \lambda_2 = 0 \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial y} = 2y + 2\lambda_1 y = 0 \implies (1 + \lambda_1)y = 0 \tag{9}$$

$$\frac{\partial \mathcal{L}}{\partial z} = 2z - 2\lambda_1 z - 2\lambda_2 = 0 \implies (1 - \lambda_1)z - \lambda_2 = 0 \tag{10}$$

$$x^2 + y^2 - z^2 = 0 \tag{11}$$

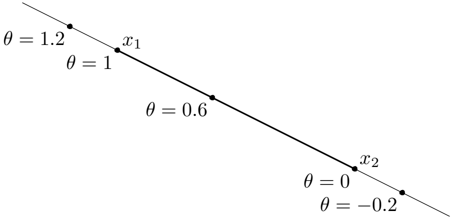
$$x - 2z - 3 = 0 \tag{12}$$

From (9) $(1 + \lambda_1)y = 0$, so $y = 0 \ \vee \ \lambda_1 = -1$. If $y = 0$, from (11) $x^2 = z^2 \implies x = \pm z$. If $x = z$, from (12) $x = z = -3$. If $x = -z$, from (12) $x = 1$, $z = -1$; if $\lambda_1 = -1$, from (8) $\lambda_2 = 0$, from (10) $z = 0$, substituting into (11) gives $x = y = 0$, which contradicts (12). Therefore, the extremum points are $(x, y, z) = (-3, 0, -3)$, $(1, 0, -1)$; optimizer: $(1, 0, -1)$.

# Introduction to Convex Programming

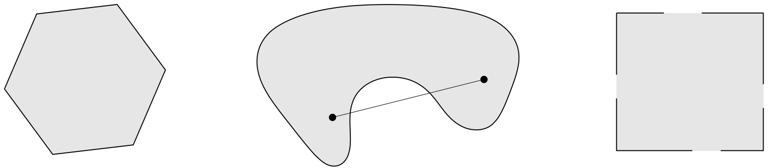
# Affine Set

- **line** through $x_1$, $x_2$: all points of the form $x = \theta x_1 + (1-\theta)x_2$, $\theta \in \mathbb{R}$
- **affine set** contains the line through any two distinct points in the set
- e.g. solution set of linear equations $\{x \mid Ax = b\}$; every affine set can be expressed as solution set of system of linear equations



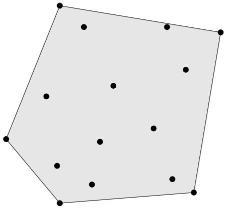
# Convex Set

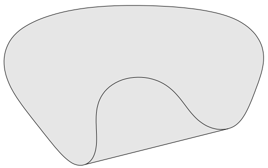
- **line segment** through $x_1$, $x_2$: all points of the form
  $x = \theta x_1 + (1-\theta)x_2$, $0 \leqslant \theta \leqslant 1$
- **convex set** contains the line segment between any two distinct points in the set:
  $x_1, x_2 \in S \implies \forall\, 0 \leqslant \theta \leqslant 1,\ \theta x_1 + (1-\theta)x_2 \in S$



# Convex Combination, Convex Hull

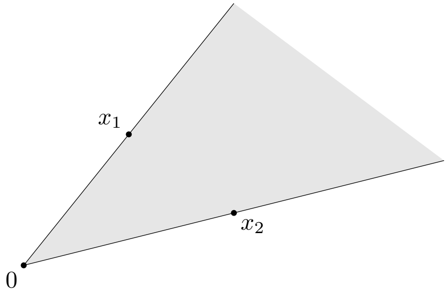
- **convex combination** of $x_1$, $x_2$, ..., $x_k$: any point $x$ of the form $x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ with $\theta_1 + \theta_2 + \cdots + \theta_k = 1$, $\theta_i \geqslant 0$
- **convex hull** $\operatorname{conv} S$: sets of all convex combinations of points in $S$





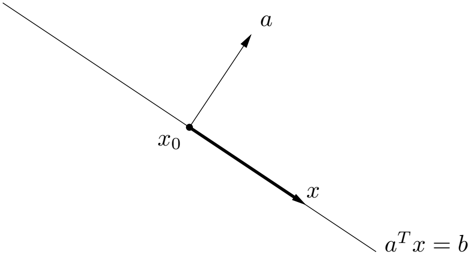
# Convex Cone

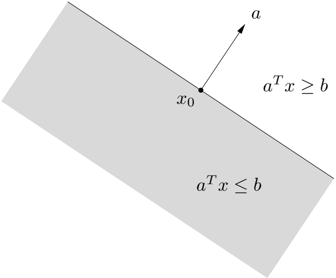
- **conic (nonnegative) combination** of $x_1$ and $x_2$: any point $x$ of the form $x = \theta_1 x_1 + \theta_2 x_2$ with $\theta_i \geqslant 0$
- **convex cone** set that contains all conic combinations of points in the set



# Hyperplane, Halfspace

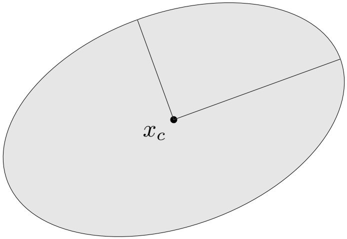
- **hyperplane**: set of the form $\{x \mid a^\top x = b\}$ with $a \neq 0$
- **halfspace**: set of the form $\{x \mid a^\top x \leqslant b\}$ with $a \neq 0$
- $a$: normal vector
  hyperplanes are affine and convex, halfspaces are convex



# Euclidean Ball, Ellipsoid

- **(Euclidean) ball** with center $x_c$ and radius $r$:
  $B(x_c, r) = \{x \mid \|x - x_c\|_2 \leqslant r\} = \{x_c + r\,u \mid \|u\|_2 \leqslant 1\}$
- **ellipsoid**: set of the form $\{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leqslant 1\}$ with $P \in \mathsf{S}^n_{++}$ ($P$ symmetric positive definite), or $\{x_c + A\,u \mid \|u\|_2 \leqslant 1\}$ with nonsingular $A$



# Norm Ball, Norm Cone

- **norm**: a function $\|\cdot\|$ that satisfies
  - $\|x\| \geqslant 0$; $\|x\| = 0 \iff x = 0$
  - $\|tx\| = |t|\|x\|$, $\forall\, t \in \mathbb{R}$
  - $\|x + y\| \leqslant \|x\| + \|y\|$
- **norm ball** with center $x_c$ and radius $r$: $\{x \mid \|x - x_c\| \leqslant r\}$
- **norm cone**: $\{(x, t) \mid \|x\| \leqslant t\}$
- norm balls and norm cones are convex
- notation for different norms: $\|\cdot\|_2$, $\|\cdot\|_{\text{symb}}$

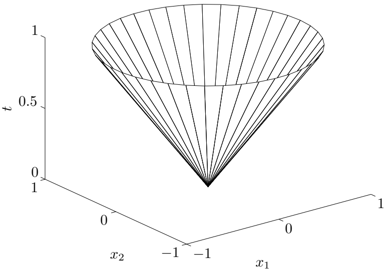
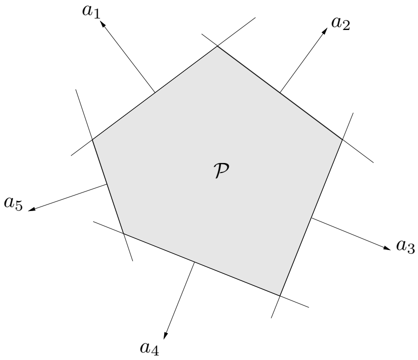



Figure: Boundary of second-order cone in $\mathbb{R}^3$, $\{(x_1, x_2, t) \mid (x_1^2 + x_2^2)^{\frac{1}{2}} \leqslant t\}$.

# Polyhedra

- **polyhedron**: solution set of finitely many linear equalities and inequalities $\{x \mid A\,x \preccurlyeq b,\, C\,x = d\}$, where $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$, $\preccurlyeq$ is componentwise inequality
- intersection of finite number of halfspaces and hyperplanes



# Positive Semidefinite Cone

- $\mathsf{S}^n$: set of symmetric $n \times n$ matrices
- $\mathsf{S}_+^n = \{X \in \mathsf{S}^n \mid X \succcurlyeq 0\}$: set of positive semidefinite (symmetric) $n \times n$ matrices; $X \in \mathsf{S}_+^n \iff z^\top X z \geqslant 0 \; \forall\, z$; a convex cone, the **positive semidefinite cone**; Below: $\begin{pmatrix} x & y \\ y & z \end{pmatrix} \in \mathsf{S}_+^2$
- $\mathsf{S}_{++}^n = \{X \in \mathsf{S}^n \mid X \succ 0\}$: set of positive definite (symmetric) $n \times n$ matrices

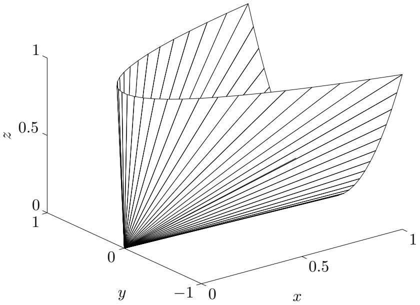



Figure: $x \geqslant 0$, $z \geqslant 0$, $xz \geqslant y^2$.

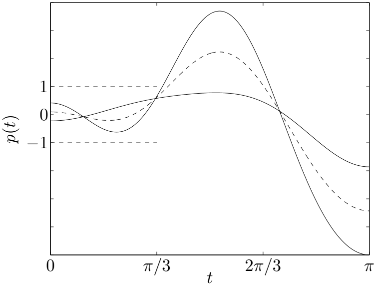
# Showing a Set is Convex

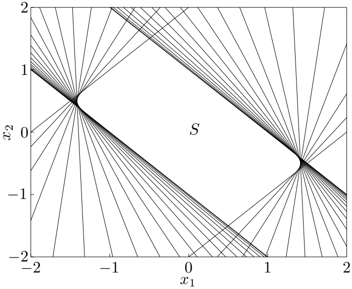
- apply definition: $x_1$, $x_2 \in S \implies \theta x_1 + (1-\theta)x_2 \in S$, $\forall\, 0 \leqslant \theta \leqslant 1$
  recommended only for simple sets
- use convex functions (later)
- show that the set is obtained from other simple convex sets (e.g. hyperplanes, halfspaces, norm balls) by operations that preserve convexity:
  - intersection
  - affine mapping
  - perspective mapping
  - linear-fractional mapping
- mostly using last two

# Intersection

- intersection of (any number of) convex sets is convex
- e.g. $S = \left\{ x \in \mathbb{R}^m \;\middle|\; |p(t)| \leqslant 1 \ \forall\, |t| \leqslant \frac{\pi}{3} \right\}$, $p(t) = \sum_{k=1}^{m} x_k \cos kt$

  is convex by $S = \bigcap_{|t| \leqslant \frac{\pi}{3}} \{x \mid |p(t)| \leqslant 1\}$; intersection of convex slabs. Below: $m = 2$.



# Affine Mappings

- suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is **affine**, i.e.

$$f(x) = A\,x + b \quad \text{with } A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$$

- the **image** of a convex set under $f$ is convex:

$$S \subseteq \mathbb{R}^n \text{ convex} \implies f(S) = \{f(x) \,|\, x \in S\} \text{ convex}$$

- the **inverse image** of a convex set under $f$ is convex:

$$C \subseteq \mathbb{R}^m \text{ convex} \implies f^{-1}(C) = \{x \in \mathbb{R}^n \,|\, f(x) \in C\} \text{ convex}$$

- e.g. scaling $aS + b = \{ax + b \,|\, x \in S\}$, $a$, $b \in \mathbb{R}$ is convex
- e.g. projection $\underset{x}{\text{proj}}(S) = \{x \,|\, (x, y) \in S\}$ is convex

# Perspective and Linear-Fractional Function

- **perspective function** $p : \mathbb{R}^{n+1} \to \mathbb{R}^n$:

$$p(x,t) = \frac{x}{t} \quad \mathrm{dom}\, p = \{(x,t) \mid t > 0\}$$

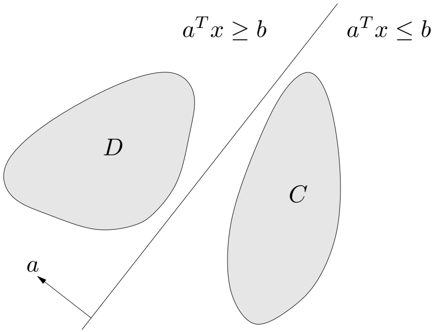
- **linear-fractional function** $f : \mathbb{R}^n \to \mathbb{R}^m$:

$$f(x) = \frac{A\,x + b}{c^\top x + d} \quad \mathrm{dom}\, f = \{x \mid c^\top x + d > 0\}$$

- images and inverse images of convex sets under perspective and linear-fractional functions are all convex

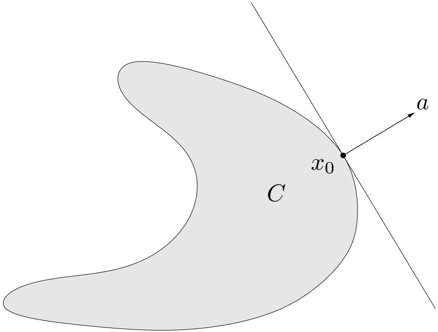
# Separating Hyperplane Theorem

- if $C$, $D$ are nonempty disjoint ($C \cap D = \emptyset$) convex sets, $\exists\, a \neq 0$, $b$ s.t. $a^\top x \leqslant b$ for $x \in C$, $a^\top x \geqslant b$ for $x \in D$
- the hyperplane $\{x \mid a^\top x = b\}$ **separates** $C$ and $D$
- strict separating requires additional assumptions (e.g. $C$ is closed; $D$ is a singleton)



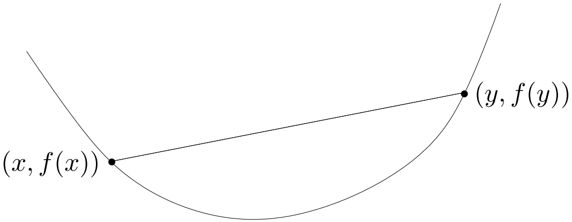
# Supporting Hyperplane Theorem

- suppose $x_0$ is a boundary point of $C \subseteq \mathbb{R}^n$
- **supporting hyperplane** to $C$ at $x_0$: $\{x \mid a^\top x = a^\top x_0\}$, where $a \neq 0$ and $a^\top x \leqslant a^\top x_0 \ \forall\, x \in C$.
- if $C$ is convex, then there exists a supporting hyperplane at every boundary point of $C$



# Convex Function

- $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if $\operatorname{dom} f$ is convex and $\forall\, x,\, y \in \operatorname{dom} f$, $0 \leqslant \theta \leqslant 1$, $f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y)$
- $f : \mathbb{R}^n \to \mathbb{R}$ is **strictly convex** if $\operatorname{dom} f$ is convex and $\forall\, x,\, y \in \operatorname{dom} f$, $x \neq y$, $0 < \theta < 1$, $f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$
- $f$ is **concave** if $-f$ is convex



# Example Functions on $\mathbb{R}$

- convex functions
  - affine: $ax + b$, $\forall\, a,\ b \in \mathbb{R}$
  - exponential: $e^{ax}$, $\forall\, a \in \mathbb{R}$
  - power: $x^\alpha$ on $x > 0$, $\forall\, \alpha \geqslant 1 \ \vee\ \alpha \leqslant 0$
  - power of absolute value: $|x|^\alpha$, $\forall\, \alpha \geqslant 1$
  - positive part (relu): $\max\{x, 0\}$
- concave functions
  - affine: $ax + b$, $\forall\, a,\ b \in \mathbb{R}$
  - power: $x^\alpha$ on $x > 0$, $\forall\, 0 \leqslant \alpha \leqslant 1$
  - logarithm: $\log x$ on $x > 0$
  - entropy: $-x \log x$ on $x > 0$
  - negative part: $\min\{x, 0\}$

# Example Convex Functions on $\mathbb{R}^n$

- affine: $a^\top x + b$
- any norm
  - $\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}$, $\forall\, p > 1$
  - $\|x\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$
- sum of squares: $\|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_n^2$
- max function: $\max(x) = \max\{x_1, x_2, \ldots, x_n\}$
- softmax / log-sum-exp: $\log\left(e^{x_1} + e^{x_2} + \cdots + e^{x_n}\right)$

# Example Functions on $\mathbb{R}^{m \times n}$

- Let $X \in \mathbb{R}^{m \times n}$ be the variable
- general affine function

$$f(X) = \operatorname{tr}(A^\top X) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b, \quad A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}$$

- spectral norm (maximum singular value) is convex:

$$f(X) = \|X\|_2 = \sigma_{\mathsf{max}}(X) = \sqrt{\lambda_{\mathsf{max}}(X^\top X)}$$

- log determinant is concave:

$$f(X) = \log \det X, \quad X \in \mathsf{S}^n_{++}$$

# Extended-Value Extension

- suppose $f$ is convex on $\mathbb{R}^n$
- its extended-value extension $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined as

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \operatorname{dom} f \\ \infty & x \notin \operatorname{dom} f \end{cases}$$

- this often simplifies notation; e.g. the condition

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

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

  - $\operatorname{dom} f$ is convex
  - $x,\, y \in \operatorname{dom} f,\ 0 \leqslant \theta \leqslant 1 \implies f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta)f(y)$

# Restriction of a Convex Function to a Line

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex (concave) $\iff$ $g : \mathbb{R} \to R$,

$$g(t) = f(x + t\,v), \quad \mathrm{dom}\, g = \{t \,|\, x + t\,v \in \mathrm{dom}\, f\}$$

  is convex (concave) in $t$ for all $x \in \mathrm{dom}\, f$ and $v \in \mathbb{R}^n$

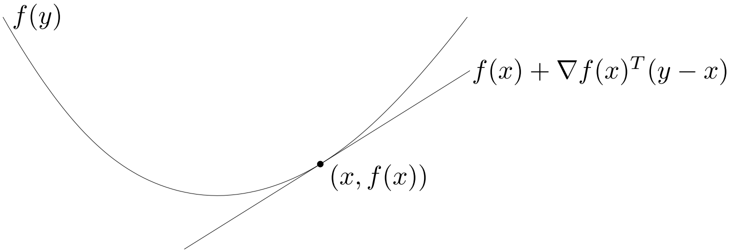
- useful for checking convexity / concavity of multivariate $f$; e.g. to check the concavity of log determinant: Let $X \in \mathsf{S}^n_{++}$, $V \in \mathsf{S}^n$,

$$\begin{aligned} g(t) = f(X + t\,V) &= \log\det(X + t\,V) \\ &= \log\det\left(X^{\frac{1}{2}}(I + t\,X^{-\frac{1}{2}}VX^{-\frac{1}{2}})X^{\frac{1}{2}}\right) \\ &= \log\det X + \log\det\left(I + t\,X^{-\frac{1}{2}}VX^{-\frac{1}{2}}\right) \\ &= \log\det X + \sum_{i=1}^{n} \log(1 + t\lambda_i) \end{aligned}$$

  where $\lambda_i$ are the eigenvalues of $X^{-\frac{1}{2}}VX^{-\frac{1}{2}}$; $g$ is concave in $t$

# First-Order Condition

- $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable if $\operatorname{dom} f$ is open and the gradient $\nabla f$ exists at each $x \in \operatorname{dom} f$.
- **first-order condition** differentiable $f$ with convex domain is convex $\iff$ $f(y) \geqslant f(x) + \nabla f(x)^\top (y - x), \ \forall\, x, \, y \in \operatorname{dom} f$
- first order Taylor approximation of convex $f$ is a **global underestimator** of $f$



# Second-Order Condition

- $f : \mathbb{R}^n \to \mathbb{R}$ is **differentiable** if $\operatorname{dom} f$ is open and the Hessian matrix $\nabla^2 f \in \mathsf{S}^n$ exists at each $x \in \operatorname{dom} f$:

$$\left\{\nabla^2 f(x)\right\}_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

- **second-order condition** for twice differentiable $f$ with convex domain is convex:
  - $f$ is convex $\iff \nabla^2 f \succeq 0$, $\forall\, x \in \operatorname{dom} f$
  - $\nabla^2 f \succ 0$, $\forall\, x \in \operatorname{dom} f \implies f$ is strictly convex

# Examples I

- **quadratic function**: $f(x) = \dfrac{1}{2}\, x^\top P x + q^\top x + r$ with $P \in \mathsf{S}^n$

$$\nabla f(x) = Px + q, \quad \nabla^2 f(x) = P$$

  convex if $P \succcurlyeq 0$ (concave if $P \preccurlyeq 0$)

- **least-squares objective**: $f(x) = \|A\,x - b\|^2$

$$\nabla f(x) = 2A^\top (A\,x - b), \quad \nabla^2 f(x) = 2A^\top A$$

  convex for any $A$

- **quadratic-over-linear function**: $f(x, y) = \dfrac{x^2}{y}$, $y > 0$

$$\nabla f(x, y) = \begin{pmatrix} \frac{2x}{y} & -\frac{x^2}{y^2} \end{pmatrix}, \quad \nabla^2 f(x, y) = \frac{2}{y^3} \begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix}$$

  convex for $y > 0$

# Examples II

- **log-sum-exp function**: $f(x) = \log\left(\sum_{k=1}^{n} e^{x_k}\right)$ is convex:

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^\top z}\,\mathrm{diag}(z) - \frac{1}{(\mathbf{1}^\top z)^2} z z^\top, \quad z_k = e^{x_k}$$

  - to show that $\nabla^2 f(x) \succcurlyeq 0$, one must verify $v^\top \nabla^2 f(x)\, v \geqslant 0\ \forall\, v$:

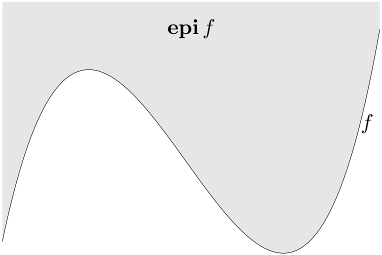
$$v^\top \nabla^2 f(x)\, v = \frac{\left(\sum_k z_k v_k^2\right)\left(\sum_k z_k\right) - \left(\sum_k v_k z_k\right)^2}{\left(\sum_k z_k\right)^2} \geqslant 0$$

  by Cauchy-Schwarz inequality $\left(\sum_k z_k v_k^2\right)\left(\sum_k z_k\right) \geqslant \left(\sum_k v_k z_k\right)^2$

- **geometric-mean function**: $f(x) = \left(\prod_{k=1}^{n} x_k\right)^{\frac{1}{n}}$ on $x \succ 0$ is concave

# Epigraph, Sublevel Set

- $\alpha$-**sublevel set** of $f : \mathbb{R}^n \to \mathbb{R}$: $C_\alpha = \{x \in \operatorname{dom} f \mid f(x) \leqslant \alpha\}$
- sublevel sets of convex functions are convex sets
- **epigraph** of $f : \mathbb{R}^n \to \mathbb{R}$: $\operatorname{epi} f = \{(x, t) \in \mathbb{R}^{n+1} \mid x \in \operatorname{dom} f,\ f(x) \leqslant t\}$
- $f$ is convex $\iff$ $\operatorname{epi} f$ is a convex set



# Jensen's Inequality

- **basic form**: if $f$ is convex, then for $x$, $y \in \operatorname{dom} f$, $0 \leqslant \theta \leqslant 1$

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

- **extension**: if $f$ is convex and $z$ is a random variable on $\operatorname{dom} f$,

$$f(\mathsf{E}\, z) \leqslant \mathsf{E}\, f(z)$$

- basic form is special case with discrete distribution

$$\mathsf{P}\{z = x\} = \theta, \quad \mathsf{P}\{z = y\} = 1 - \theta$$

- e.g. for $z \sim \mathsf{N}(\mu, \sigma^2)$, let $f(x) = e^x$, then

$$f(\mathsf{E}\, z) = f(\mu) = e^{\mu} \leqslant e^{\mu + \frac{\sigma^2}{2}} = \mathsf{E}\, f(z)$$

# Showing Convexity of a Function

- apply definition (often simplified by restricting to a line)
- for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$
- show that $f$ is obtained from simple convex functions by operations that preserve convexity
  - nonnegative multiple, sum, integral
  - composition with affine function
  - pointwise maximum and supremum
  - partial minimization
  - composition
  - perspective

# Nonnegative Multiple, Sum, Integral

- **nonnegative multiple**: $\alpha f$ is convex if $f$ is convex and $\alpha \geqslant 0$
- **sum**: $f_1 + f_2$ is convex if $f_1$, $f_2$ is convex
- **infinite sum**: if each of $f_i$ is convex, then $\sum_{i=1}^{\infty} f_i$ is convex
- **integral**: if $f(x, \alpha)$ is convex in $x$ for each $\alpha \in \mathcal{A}$, then

$$\int_{\alpha \in \mathcal{A}} f(x, \alpha) \, \mathsf{d}\alpha$$

  is convex
- analogous rules for concave functions

# Composition with Affine Function

- $f(A\,x + b)$ is convex if $f$ is convex
- e.g.
  - log barrier for linear inequalities

$$f(x) = -\sum_{i=1}^{m} \log\left(b_i - a_i^\top x\right)$$

$$\operatorname{dom} f = \{x \mid a_i^\top x < b_i,\ i = 1, 2, \ldots, m\}$$

  - norm approximation error (any norm)

$$f(x) = \|A\,x - b\|$$

# Pointwise Maximum

- $f(x) = \max\{f_1(x), f_2(x), \dots, f_m(x)\}$ is convex if each $f_i$ is convex
- e.g.
  - piecewise linear function

$$f(x) = \max_i \left(a_i^\top x + b_i\right)$$

  - sum of $r$ largest components of $x \in \mathbb{R}^n$

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

  where $x_{[i]}$ is $i$-th largest component of $x$. Note that

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n\}$$

# Pointwise Supremum

- $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ is convex if $f(x, y)$ is convex in $x$ for each $y \in \mathcal{A}$
- e.g.
  - distance to farthest point in a set $C$

$$f(x) = \sup_{y \in C} \|x - y\|$$

  - maximum eigenvalue of symmetric matrix

$$\lambda_{\mathsf{max}}(X) = \sup_{\|y\|_2 = 1} y^\top X\, y, \quad X \in \mathsf{S}^n$$

  - support function of a set $C$

$$S_C(x) = \sup_{y \in C} y^\top x$$

# Partial Minimization

- the function $g(x) = \inf_{y \in C} f(x, y)$ is called the **partial minimization** of $f$ w.r.t. $y$
- if $f(x, y)$ is convex in $(x, y)$ and $C$ is a convex set, then partial minimization $g$ is convex
- e.g.
  - let $f(x, y) = x^\top A\, x + 2x^\top B\, y + y^\top C\, y$ with $\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} \succcurlyeq 0$, $C \succ 0$; minimizing over $y$ gives

$$g(x) = \inf_{y \in C} f(x, y) = x^\top (A - BC^{-1} B^\top)\, x$$

    $g$ is convex, hence Schur complement $A - BC^{-1}B^\top \succcurlyeq 0$
  - distance to a convex set $S$

$$\mathrm{dist}(x, S) = \inf_{y \in S} \|x - y\|$$

# Composition with Scalar Functions

- composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$ is $f(x) = h(g(x))$ ($f = h \circ g$)
- composition $f$ is convex if
  - $g$ convex, $h$ convex, $\tilde{h}$ nondecreasing; or
  - $g$ concave, $h$ convex, $\tilde{h}$ nonincreasing
- proof for $n = 1$, differentiable $g$, $h$

$$f''(x) = h''(g(x))\, g'(x)^2 + h'(g(x))\, g''(x)$$

- e.g.
  - $f(x) = e^{g(x)}$ is convex if $g$ is convex
  - $f(x) = \dfrac{1}{g(x)}$ is convex if $g$ is concave and positive

# Composition: General

- composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^n \to \mathbb{R}$ is $f(x) = h(g_1(x), g_2(x), \ldots, g_k(x))$
- composition $f$ is convex if $h$ is convex and for each $i$, one of the following holds:
  - $g_i$ convex, $\tilde{h}$ nondecreasing in its $i$-th argument
  - $g_i$ concave, $\tilde{h}$ nonincreasing in its $i$-th argument
  - $g_i$ affine
- e.g.
  - $\log\left(\sum_{i=1}^{m} e^{g_i(x)}\right)$ is convex if each $g_i$ is convex
  - $\dfrac{p(x)^2}{q(x)}$ is convex if $p$ is nonnegative and convex and $q$ is positive and concave

# Perspective

- perspective of $f : \mathbb{R}^n \to \mathbb{R}$ is the function $g(x,t) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ defined as

$$g(x,t) = t\, f\Big(\frac{x}{t}\Big), \quad \operatorname{dom} g = \Big\{(x,t) \,\Big|\, \frac{x}{t} \in \operatorname{dom} f,\ t > 0\Big\}$$

- $g$ is convex if $f$ is convex
- e.g.
  - $f(x) = x^\top x$ is convex, so $g(x,t) = \dfrac{x^\top x}{t}$ is convex if $t > 0$
  - $f(x) = -\log x$ is convex, so the **relative entropy**

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

  is convex on $x > 0$, $t > 0$

# Convexity Verification: An Example

- test the convexity of $f(x,y) = \dfrac{(x-y)^2}{1-\max(x,y)},\ x<1,\ y<1$
- $x$, $y$, and $1$ are affine
- $\max(x,y)$ is convex; $x-y$ is affine
- $1-\max(x,y)$ is concave
- $\dfrac{u^2}{v}$ is convex, monontone decreasing in $v$ for $v>0$
- $f$ is composition of $\dfrac{u^2}{v}$ with $u=x-y$, $v=1-\max(x,y)$, hence convex

# Convexity Verification: A Caveat

- test the convexity of $f(x) = \sqrt{1+x^2}$
- $\sqrt{\cdot}$ is concave
- $1$, $x^2$ are convex
- $\sqrt{1+x^2}$ is ... indefinite ?
- but, note that $\|\cdot\|_2$ is convex
- $\sqrt{1+x^2}$ can be represented as the 2-norm of vector $(1, x)$ — $\|(1, x)\|_2$, hence is convex
- The general composition rules are only sufficient, not necessary

## Standard Form of General Optimization

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leqslant 0, \quad i = 1, 2, \dots, m \\ & h_i(x) = 0, \quad i = 1, 2, \dots, p \end{array}$$

- $x \in \mathbb{R}^n$ is the optimization variable
- $f_0(x) : \mathbb{R}^n \to \mathbb{R}$ is the objective / cost
- $f_i(x) : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \dots, m$ are the inequality constraints
- $h_i(x) : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \dots, p$ are the equality constraints

## Standard Form of Convex Optimization

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leqslant 0, \quad i = 1, 2, \cdots, m \\ & a_i^\top x = b_i, \quad i = 1, 2, \dots, p \end{array}$$

- objective and inequality constraints $f_0, f_1, \dots, f_m$ are convex
- equality constraints are affine, often written as $A\,x = b$
- feasible and optimal sets of a convex optimization problem are convex

# Local and Global Optima

## Theorem

Locally optimal point of a convex optimization problem is (globally) optimal.

## Proof

- suppose $x$ is locally optimal, but $\exists\, y$ with $f_0(y) < f_0(x)$
- $x$ locally optimal means $\exists\, R > 0$ such that if $x'$ is feasible and $\|x' - x\| \leqslant R$, then $f_0(x') \geqslant f_0(x)$
- set $z = \theta y + (1-\theta)x$ with $\theta = \dfrac{R}{2\|y - x\|_2}$
- $\|y - x\|_2 > R$, so $0 < \theta < \dfrac{1}{2}$
- $z$ is a convex combination of two feasible points, hence also feasible
- $\|z - x\|_2 = \dfrac{R}{2}$ and $f_0(z) \leqslant \theta f_0(y) + (1-\theta) f_0(x) < f_0(x)$, which contradicts that $x$ is locally optimal