Chapter 2: Optimization for Data Science
Convex Functions

TANN Chantara

Department of Applied Mathematics and Statistics
Institute of Technology of Cambodia

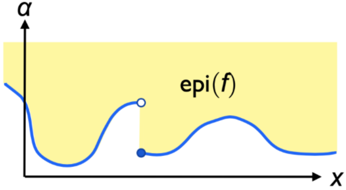
October 9, 2022

Table of Contents

Epigraph and Domain

Definition: The epigraph of $f : \mathbb{R}^n \to (-\infty, \infty]$ is the set

$$\text{epi}(f) = \{(x, \alpha) \in \mathbb{R}^{n+1} : f(x) \leq \alpha\}$$



Definition: The domain of $f : \mathbb{R}^n \to (-\infty, \infty]$ is the set

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

Definition: A function $f : \mathbb{R}^n \to (-\infty, \infty]$ is called proper if $\text{dom}(f) \neq \emptyset$

Convex Functions

Definition: A function $f : \mathbb{R}^n \to (-\infty, \infty]$ is called convex if its epigraph is a convex set.

Proposition: f is conve if and only if its domain is a convex set and

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

for all $x, y \in \mathrm{dom}(f)$ and $\theta \in [0, 1]$

The line segment between $(x, f(x))$ and $(y, f(y))$, which is the chord from x to y, lies above the graph of f

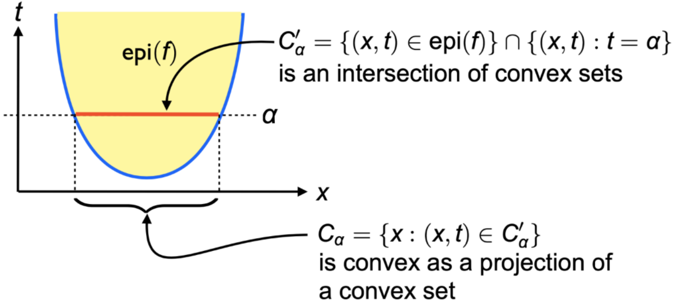
- f is called strictly convex if the in equality in (1) is strict.
- f is called concave if $-f$ is convex.

Sublevel Sets

Definition: The α-sublevel set of a function $f : \mathbb{R}^n \to (-\infty, \infty]$ is defined as $C_\alpha = \{x : f(x) \leq \alpha\}$

Proposition: f is convex then all of its sublevel sets are convex.

- Reverse implication is not true.
- **Exercise:** Find a non-convex function whose sublevel sets are all convex.



Examples of Convex Functions

Univariate functions		Domain
• Exponential functions	$f(x) = e^{ax}$	$\mathbb{R}$
• Powers	$f(x) = x^{\alpha} (\alpha \geq 1, \alpha \leq 0)$	$\mathbb{R}_{++}$
• Negative logarithm	$f(x) = -\log x$	$\mathbb{R}_{++}$
• Negative entropy	$f(x) = x \log x$	$\mathbb{R}_{++}$
Multivariate functions		Domain
• Negative entropy	$f(x) = a^{\top} x + b$	$\mathbb{R}^n$
• p-Norms $(p \geq 1)$	$f(x) = \|\|x\|\|_p = (\sum_{i=1}^{n} \|\|x\|\|^p)^{1/p}$	$\mathbb{R}^n$
• ∞-Norm	$f(x) = \|\|x\|\|_p = max_i \|x_i\|$	$\mathbb{R}^n$
• Indicator function of convex set C	$f(x) = \begin{cases} 0, & x \in C \\ \infty, & \text{else} \end{cases}$	C

Convention: $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$

Example of Convex Functions (cont'd)

Univariate functions — Domain

- Trace functions (linear functions) — $\mathbb{R}^{m\times n}$

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

- maximum eigenvalue $f(X) = \lambda_{max}(X)$ — $\mathbb{S}^n$
- Spectral norm $f(X) = ||X||_2 = \sup_{v\neq 0}||Xv||_2/||v||_2$ — $\mathbb{R}^{m\times n}$

Checking Convexity Along Line

Proposition: A function $f : \mathbb{R}^n \to (-\infty, \infty]$ is convex if and only if each univariate function $g : \mathbb{R} \to (-\infty, \infty]$ of the form

$$g(t) = f(x+ty), \qquad \text{for } x,\ y \in \mathbb{R}^n$$

is convex in t.

Proof:

$\Rightarrow$: For any $x, y \in \mathbb{R}^n$ and $\theta \in (0, 1)$, consider $t = \theta a + (1-\theta)b$ for arbitrary $a, b \in \mathbb{R}$

$$\begin{aligned} g(t) &= g(\theta a + (1-\theta)b) = f(x + (\theta a + (1-\theta)b)y) \\ &= f(\theta(x + ay) + (1-\theta)(x + by)) \\ &\le \theta f(x + ay) + (1-\theta)f(x + by) = \theta g(a) + (1-\theta)g(b) \end{aligned}$$

Checking Convexity Along Line

Proof: Cont'd

$\Leftarrow$: For any $x, y \in \mathbb{R}^n, \theta \in (0,1)$ and $t_1, t_2 \in \mathbb{R}$

$$f(x + (\theta t_1 + (1-\theta)t_2)y) = g(\theta t_1 + (1-\theta)t_2)$$
$$\leq \theta g(t_1) + (1-\theta)g(t_2) = \theta f(x + t_1 y) + (1-\theta)f(x + t_2 y)$$

For $t_1 = 0, t_2 = 1, x = x', y = y' - x'$ we get

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

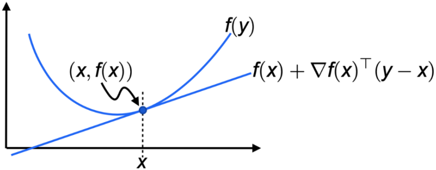
1st Order Conditions

Definition: A function $f : \mathbb{R}^n \to (-\infty, \infty]$ is differentiable if its gradient $\nabla f = (\partial f/\partial x_1, \dots, \partial f/\partial x_n)$ exists at each point in dom(f) and if dom(f) is open

Proposition: A differentiable function $f : \mathbb{R}^n \to (-\infty, \infty]$ is convex if and only if dom(f) is convex and

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \quad \forall x, y \in \mathrm{dom}(f)$$

$\Rightarrow$ 1st-order Taylor approximation underestimates f globally.
$\Rightarrow$ From local information about convex function we can obtain global information.



Univariate Functions

Proposition: A differential function $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if

$$f(y) \geq f(x) + f'(x)(y - x) \qquad \forall x, y \in \mathbb{R}$$

Proof:

$\Rightarrow$: If $x, y \in \mathbb{R}, 0 < t \leq 1$, then

$$f(x + t(y - x)) \leq (1 - t)f(x) + tf(y) \qquad \text{(convexity)}$$
$$f(y) - f(x) \geq [f(x + t(y - x))]/t \qquad \text{(divide by t)}$$
$$f(y) - f(x) \geq f'(x)(y - x) \qquad \text{(limit } t \to \infty)$$

$\Leftarrow$: For any $x, y \in \mathbb{R}, 0 < t \leq 1$, Let $z = tx + (1 - t)y$

$$t(f(x) - f(z)) \geq tf'(z)(x - z) \qquad \text{(by assumption)}$$
$$(1 - t)(f(y) - f(z)) \geq (1 - t)t'(z)(y - z) \qquad \text{(by assumption)}$$
$$tf(x) + (1 - t)f(y) \geq f(z) \qquad \text{(sum of above)}$$

1st Order Conditions*

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

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \qquad \forall x, y \in \mathbb{R}^n$$

Proof:

$\Rightarrow$: $g(t) = f(tx + 1(1-t)y)$ is convex in t for any $x, y \in \mathbb{R}^n$

$$g'(t) = \nabla f(tx + (1-t)y)^\top (y - x) \qquad \text{(definition of g)}$$

$$g(1) \geq g(0) + g'(0) \qquad \text{(convexity of g)}$$

$$f(x) \geq f(y) + \nabla f(y)^\top (y - x) \qquad \text{(substition)}$$

$\Leftarrow$: $x, y \in \mathbb{R}^n, t, \tilde{t} \in \mathbb{R}, z = ty + (1-t)x, \tilde{z} = \tilde{t}y + (1-\tilde{t})x$

$$f(z) \geq f(\tilde{z}) + \nabla f(\tilde{z})^\top (z - \tilde{z}) \qquad \text{(by assumption)}$$

$$g(t) \geq g(\tilde{t}) + \nabla g(\tilde{t})^\top (t - \tilde{t}) \qquad \text{(definition of } g, z, \tilde{z})$$

By the 1st-order condition for univariate functions g is convex. Thus f is also convex.

2nd Order Conditions

Definition: A function $f : \mathbb{R}^n \to (-\infty, \infty]$ is twice differentiable if its Hessian

$$\nabla^2 f(x) = \begin{pmatrix} \partial^2 f/\partial x_1 \partial x_1 & \cdots & \partial^2 f/\partial x_1 \partial x_n \\ \vdots & \ddots & \vdots \\ \partial^2 f/\partial x_n \partial x_1 & \cdots & \partial^2 f/\partial x_n \partial x_n \end{pmatrix}$$

exists at each point in $\mathrm{dom}(f)$, and $\mathrm{dom}(f)$ is open.

Proposition: A twice differential function $f : \mathbb{R}^n \to (-\infty, \infty]$ is convex if and only if $\mathrm{dom}(f)$ is convex and

$$\nabla^2 f(x) \succeq 0 \qquad \forall x \in \mathrm{dom}(f)$$

The condition $\nabla^2 f(x) \succeq 0$ can be interpreted geometrically as the requirement that f has upward curvature at x.

Univariate Functions*

Proposition: A twice differential function $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if $f''(x) \geq 0 \ \ \forall x \in \mathbb{R}$

Proof:

$\Rightarrow$: If $x, y \in \mathbb{R}, y > x,$ then

$$\begin{aligned} f(y) &\geq f(x) + f'(x)(y-x) && \text{(1st order conditions)} \\ f(x) &\geq f(y) + f'(y)(x-y) && \text{(1st order conditions)} \\ 0 &\geq (f'(x) - f'(y))/(y-x) && \text{(sum of above } \times (y-x)^{-2}) \\ 0 &\geq f''(x) && \text{(limit } y \to x) \end{aligned}$$

$\Leftarrow$: For $x, y \in \mathbb{R},$ we have

$$\begin{aligned} f(y) &= f(x) + \textstyle\int_x^y f'(u)du = f(x) + \int_x^y f'(x) + \int_x^u f''(v)dvdu \\ &\geq f(x) + \textstyle\int_x^y f'(x)du = f(x) + f'(x)(y-x) \end{aligned}$$

Thus, f is convex as it satisfies the 1st-order condition.

2nd-Order Conditions*

Proposition: A twice differential function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if $f''(x) \succeq 0 \ \ \forall x \in \mathbb{R}^n$

Proof:

$\Rightarrow$: $g(t) = f(x + ty)$ is convex in t for any$x, y \in \mathbb{R}^n$

$$
\begin{aligned}
g''(t) &= y^\top \nabla^2 f(x + ty)y && \text{(definition of g)}\\
g''(t) &\geq 0 && \text{(univariate case)}\\
\nabla^2 f(x) &\succeq 0 && \text{(as y is arbitrary)}
\end{aligned}
$$

$\Leftarrow$: Define g as above. For any t we have

$$
\begin{aligned}
\nabla^2 f(x + ty) &\succeq 0 && \text{(by assumption)}\\
g''(t) &\geq 0 && \text{(definition of g)}
\end{aligned}
$$

By the 2nd-order condition for univariate functions, g is convex. Thus, f is also convex.

Examples

- Quadratic functions $f(x) = x^\top Px + q^\top x + r$ are convex if $\nabla^2 f(x) = P \succeq 0$

- The least-squares objective $f(x) = ||Ax - b||_2^2$ is convex because $\nabla^2 f(x) = 2A^\top A \succeq 0$ for all $A \in \mathbb{R}^{m\times n}$

- Quadratic-over-linear function of the type $f(x, y) = x^2/y$ are convex as long as $y > 0$ because

$$\nabla^2 f(x, y) = \frac{2}{y^3}\begin{pmatrix} y \\ -x \end{pmatrix}(y - x) \succeq 0 \qquad \forall y > 0$$

Negative Log-Determinant

Proposition: The log-determinant function $f(X) = -\log\det(X)$ is convex on the set of positive definite matrices $\mathbb{S}^n_{++}$.

Proof: Homework.

Convexity Preserving Transformations

Sometimes one can establish convexity of f by showing that f is obtained from simple convex functions via transformations that preserve convexity:

- non-negative weight sum
- composition with affine function
- pointwise maximum and supremum
- composition
- minimization
- perspective

Affine Transformations

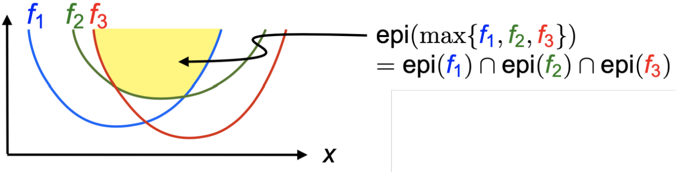
- Affine transformation of input: if f is convex, then $g(x) = f(Ax + b)$ is also convex
- Non-negative affine transformation of output: If $f_1, \dots, f_K$ are convex functions and $\rho_1, \dots, \rho_K$ are non-negative numbers, then the conic combination $g(x) = \rho_1 f_1(x) + \cdots + \rho_K f_K(x)$ is convex
- Generalization to integrals: if $f(x, y)$ is convex in x for each fixed $y \in \mathcal{Y}$ and $\rho(y)$ is a non-negative function of y, then

  $$g(x) = \int_{\mathcal{Y}} \rho(y) f(x, y) dy$$

  is convex in x (provided that the integral exists)

Pointwise maximum and supremum

Maximum of convex functions: If $f_1, \ldots, f_K$ are convex, then the pointwise maximum $g(x) = \max\{f_1(x), \ldots, f_K(x)\}$ is also convex.



Recall : Intersections of convex sets are convex

Supremum of convex functions: If $f(x, y)$ is convex in x for every fixed $y \in \mathcal{Y}$, then the pointwise supremum

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

is also convex

Examples

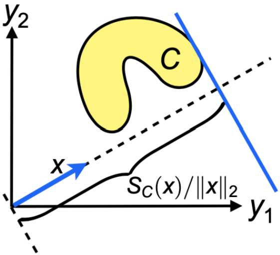
1. Piecewise linear functions $f(x) = \max_{i=1,..,K}\{a_i^\top x + b_j\}$ are convex.
2. The sum of the r largest components of $x \in \mathbb{R}^n$ is convex as it can be written as a maximum of linear functions.

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

3. The support function of a (possibly noncovex) set C is convex.

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

The hyperplane $\{y : y^\top x = S_C(x)\}$ is orthogonal to x and "supports" C



Examples Cont'd

4 Maximum eigenvalue $f(X) = \lambda_{\max}(X)$ for $X \in \mathbb{S}^n$
Write $X = RDR^\top$, with R orthogonal and $D = \mathrm{diag}(\lambda_1, ..., \lambda_n)$

$$f(X) = \sup_{||v||_2=1} v_1^2\lambda_1 + \cdots + v_n^2\lambda_n = \sup_{||v||_2=1} v^\top Dv = \sup_{||v||_2=1} v^\top Xv$$

5 Spectral norm $f(X) = ||X||_2 = \sup_{v\neq 0} ||Xv||_2/||v||_2$ for $X \in \mathbb{R}^{m\times n}$

$$f(X) = \sup_{||v||_2=1} ||Xv||_2 = \sup_{||V||_2=1} \sup_{||u||_2=1} u^\top Xv$$

Recall: $u^\top Xv \leq ||u||_2||Xv||_2 = ||Xv||_2$

In both cases $f(X)$ is the supremum of linear functions in X and thus convex.

Composition

Proposition: If $g : \mathbb{R}^n \to \mathbb{R}$ is convex and $h : \mathbb{R} \to \mathbb{R}$ is convex and non-decreasing, the $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x) = h(g(x))$ is convex.

Proof: For any $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$

$$\begin{aligned}
f(\theta x + (1-\theta)y) &= h(g(\theta x + (1-\theta)y)) && \text{(definition of f)} \\
&\leq h(\theta g(x) + (1-\theta)g(y)) && \text{(conv. of g, mono. of h)} \\
&\leq \theta h(g(x)) + (1-\theta)h(g(y)) && \text{(convexity of h)} \\
&= \theta f(x) + (1-\theta)f(y) && \text{(definition of f)}
\end{aligned}$$

Thus f is convex.

Example: $f(x) = \exp(g(x))$ is convex if g is convex.

Generalizations

Definition: A function $f : \mathbb{R}^n \to [-\infty, \infty)$ is concave if $-f$ is convex.

Proposition: If $g : \mathbb{R}^2 \to \mathbb{R}$ is concave and $h : \mathbb{R} \to \mathbb{R}$ is convex and non-increasing, then $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x) = h(g(x))$ is convex.

Proposition: If $g : \mathbb{R}^n \to \mathbb{R}^k$ is convex in each component, while $h : \mathbb{R}^k \to \mathbb{R}$ is non-decreasing in each argument and convex, then $f : \mathbb{R}^n \to \mathbb{R}$ defined via $f(x) = h(g(x))$ is convex.

Proposition: If $g : \mathbb{R}^n \to \mathbb{R}^k$ is concave in each component, while $h : \mathbb{R}^k \to \mathbb{R}$ is non-increasing in each argument and convex, then $f : \mathbb{R}^n \to \mathbb{R}$ defined via $f(x) = h(g(x))$ is convex.

Minimization

Proposition: If $f(x, y)$ and $g(x, y)$ are convex in (x, y) and C is a convex set, then the optimal value function

$$h(x) = \begin{cases} \inf_{y \in C} & f(x, y) \\ \text{s.t.} & g(x, y) \leq 0 \end{cases}$$

is convex

Proof: Assume that the inner problem is solvable, i.e., for every $x \in$dom(h). Choose $x_1, x_2 \in$dom(h) and let $y_1, y_2 \in C$ be the corresponding minimizers, i.e., $h(x_i) = f(x_i, y_i)$ for $i = 1, 2$, for any $\theta \in [0, 1]$

Minimization Cont'd

$$
\begin{aligned}
h(\theta x_1 + (1-\theta)x_2) = \inf_{y \in C} \{ & f(\theta x_1 + 1(1-\theta)x_2, y) : \\
& g(\theta x_1 + (1-\theta)x_2, y) \leq 0\} \\
& \leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\
& \leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\
& = \theta h(x_1) + (1-\theta) h(x_2)
\end{aligned}
$$

Thus, h is convex. If the problem is not solvable, once can use a similar argument using ε-optimal solution for $\varepsilon \to 0$

Schur Lemma

Lemma (Schur) Consider $X \in \mathbb{S}^n$ partitioned as $X = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$, where $C \succ 0$. Then

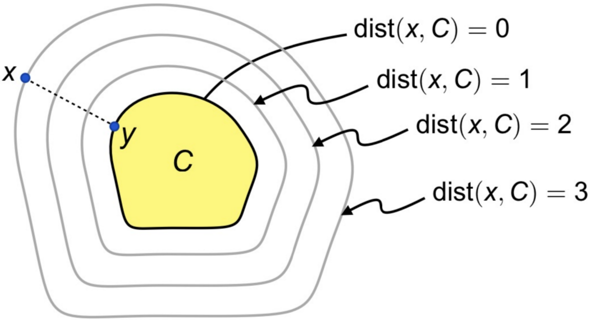
$$X \succeq 0 \iff A - BC^{-1}B^\top \succeq 0$$

The matrix $A - BC^{-1}B^\top$ is called the Schur complement of C
Proof: Consider the functions $f(x, y) = x^\top Ax + 2x^\top By + y^\top Cy$ and $h(x) = \inf_y f(x, y) = x^\top(A - BC^{-1}B^\top)x$
$\Rightarrow X \succeq 0 \implies$ f convex in $(x, y) \implies$ h convex in x
$\implies A - BC^{-1}B^\top \succeq 0$
$\Leftarrow$ We have $A - BC^{-1}B^\top$. Assume $X \not\succeq 0$
$\implies \exists (x_0, y_0) \neq 0$ with $f(x_0, y_0) < 0$
$\implies h(x_0)x_0^\top(A - BC^{-1}B^\top)x_0 < 0$, which contradicts the positive definiteness of the Schur complement. Hence, $X \succeq 0$.

Distance function

The distance of x to a fixed convex set C is convex in x, i.e.,

$$f(x) = \mathrm{dist}(x, C) = \inf_{y \in C} ||x - y||_2$$



Perspective function

Proposition: If $f(x)$ is convex, then the perspective of f, defined as

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

is convex in (x, t)

Proof: Choose $(x_1, t_1), (x_2, t_2) \in \text{dom}(g)$ and $\theta \in [0, 1]$, then

$$\begin{aligned}
g(\theta(x_1,t_1) + (1-\theta)(x_2,t_2)) &= (\theta t_1 + (1-\theta)t_2) f\left(\frac{\theta x_1 + (1-\theta)x_2}{\theta t_1 + (1-\theta)t_2}\right) \\
&= (\theta t_1 + (1-\theta)t_2) f\left(\frac{\theta t_1 x_1/t_1 + (1-\theta)t_2 x_2/t_2}{\theta t_1 + (1-\theta)t_2}\right. \\
&\leq \theta t_1 f(x_1/t_1) + (1-\theta)t_2 f(x_2/t_2) \\
&= \theta g(x_1,t_1) + (1-\theta) g(x_2,t_2)
\end{aligned}$$

Thus g is convex in (x, t)

Relative Entropy

Proposition: The relative entropy of two vector $p, q \in \mathbb{R}^n_{++}$ defined as

$$f(p, q) = \sum_{i=1}^n p_i \log(p_i/q_i)$$

is convex.

Proof: The negative logarithm $f(x) = -\log(x)$ is convex on $\mathbb{R}_{++}$. We therefore conclude that its perspective function

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

is convex on $\mathbb{R}^2_{++}$. The relative entropy now can be seen as a sum of n convex functions and as such is convex.

Convexity w.r.t generalized inequalities

Definition: Let $K \subset \mathbb{R}^m$ be a proper convex cone. The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called K-convex if

$$f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta)f(y) \qquad \forall x, y \in \mathbb{R}^n, \theta \in [0,1]$$

Proposition: If K is a proper convex cone and f is a K-convex function, then the set $C = \{x : f(x) \preceq_K 0\}$ is convex.
Proof: Consider $x, y \in C$ and $\theta \in [0,1]$. Then

$$\begin{aligned} f(\theta x + (1-\theta)y &\preceq_K \theta f(x) + (1-\theta)f(y) && (f \text{ is K-convex}) \\ &\preceq_K 0 && (x, y \in C, K \text{ convex}) \end{aligned}$$

Thus, $\theta x + (1-\theta)y \in C$, which implies that C is convex
Example: $f : \mathbb{S}^n \to \mathbb{S}^n, f(X) = X^2$, is $\mathbb{S}^n_+$-convex.

Summary

- **Definition:** epigraph, domain and sublevel sets; proper, convex and concave functions;
- **Checking convexity:** using the basic definition; checking convexity along lines; checking the 1st- or 2nd-order conditions (only for differentiable functions).
- **Convexity-preserving transformations:** non-negative weighted sum and integral; composition with affine function; parametric maximum; composition; parametric minimum (check convexity condition!); perspective.
- **Schur's lemma:** a block matrix with a positive definite diagonal block is psd if and only if this block's Schur complement is psd.
- **Generalized inequalities:** constructing convex sets using K-convex constraint functions and conic inequalities.