FOUNDATIONS OF OPTIMIZATION: IE6001

Convex Functions

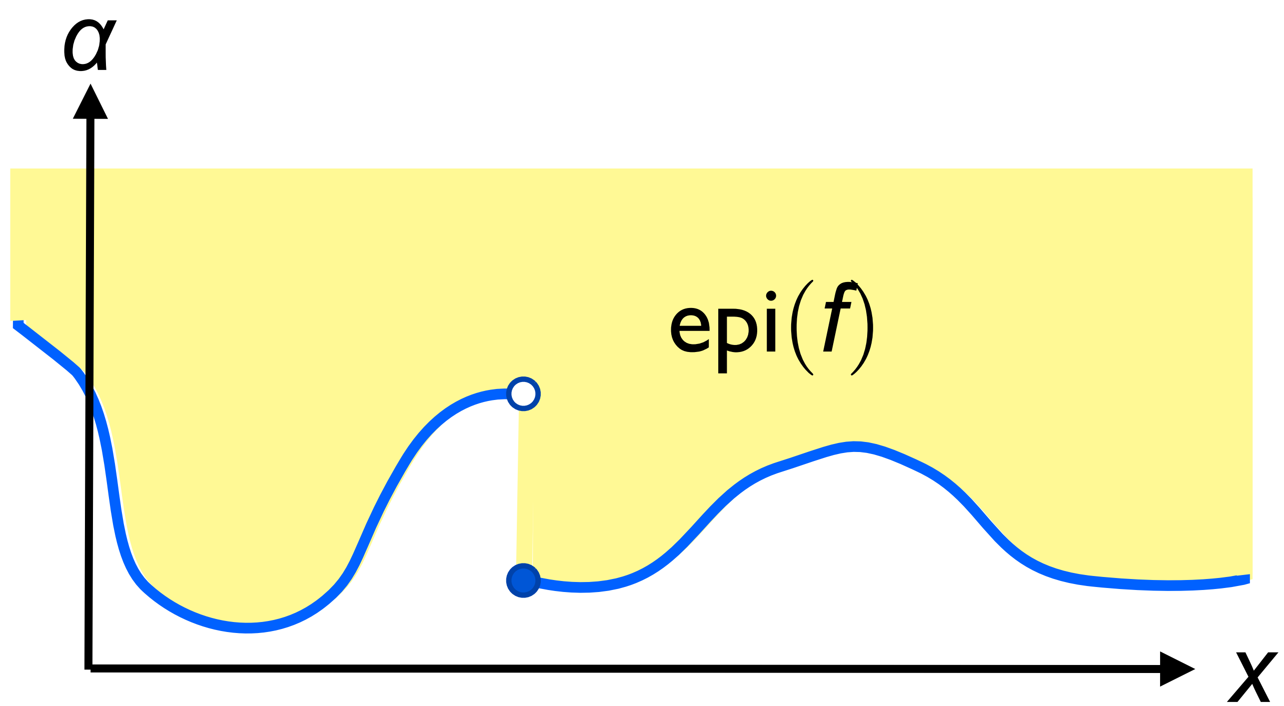
Napat Rujeerapaiboon

Semester I, AY2022/2023

Epigraph and Domain

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

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



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

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}$$

Definition: $f : \mathbb{R}^n \to (-\infty, +\infty]$ is called *proper* if $\mathrm{dom}(f) \neq \emptyset$.

Convex Functions

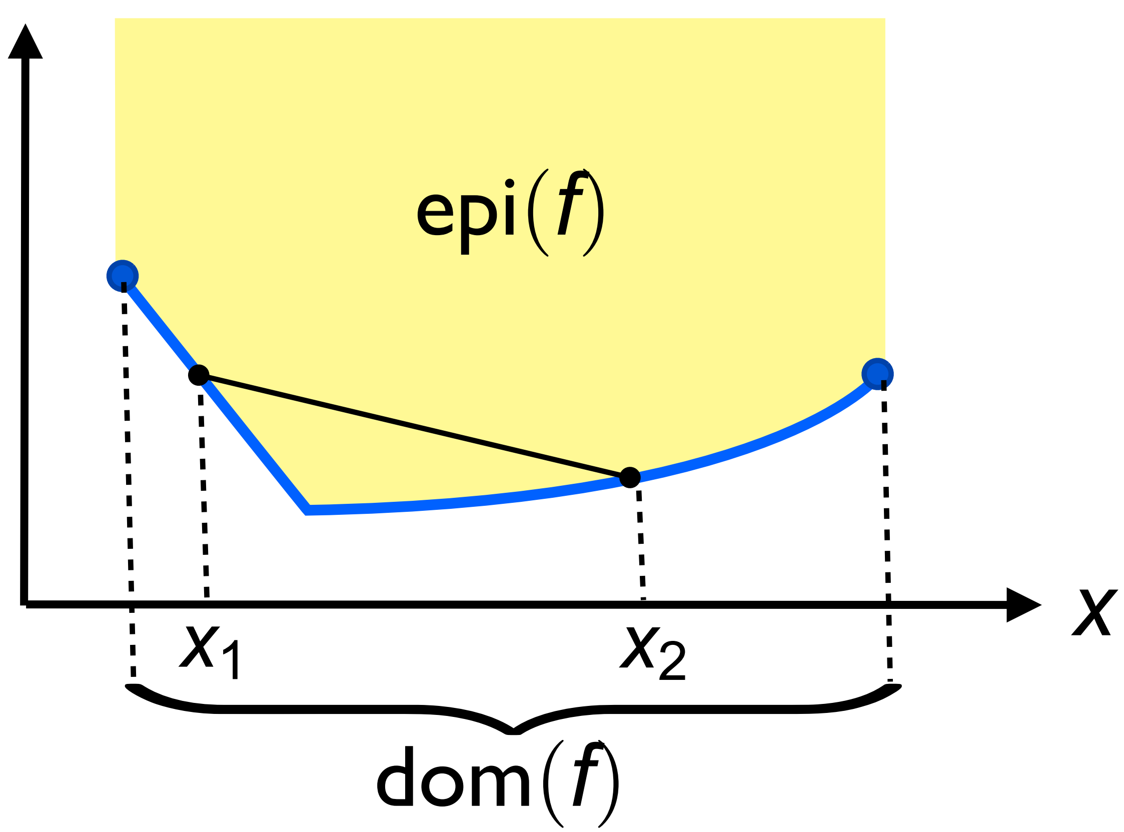
Definition: A function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is *convex* if its epigraph is a convex set.

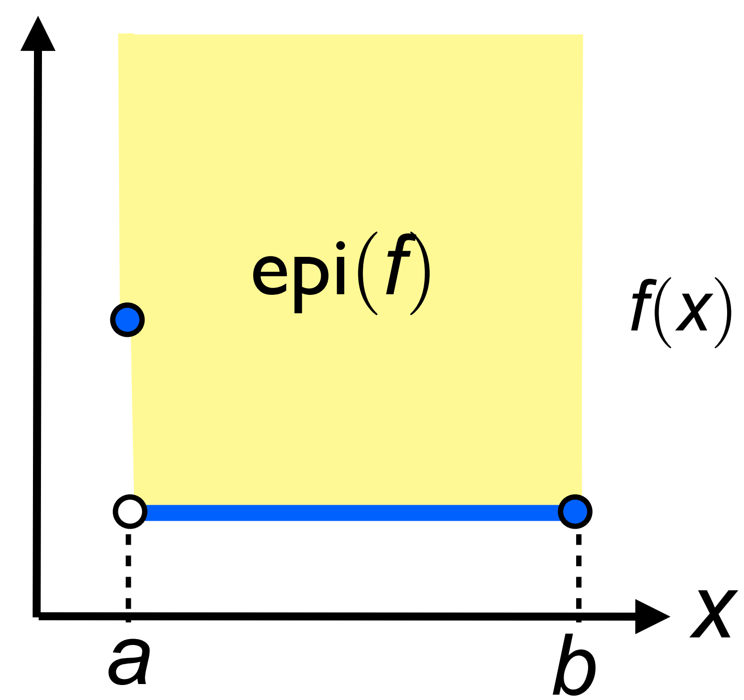
f is convex if and only if its *domain* is a *convex* set and

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta) f(x_2)$$

for all $x_1, x_2 \in \mathrm{dom}(f),\ \theta \in [0,1]$.

Examples:



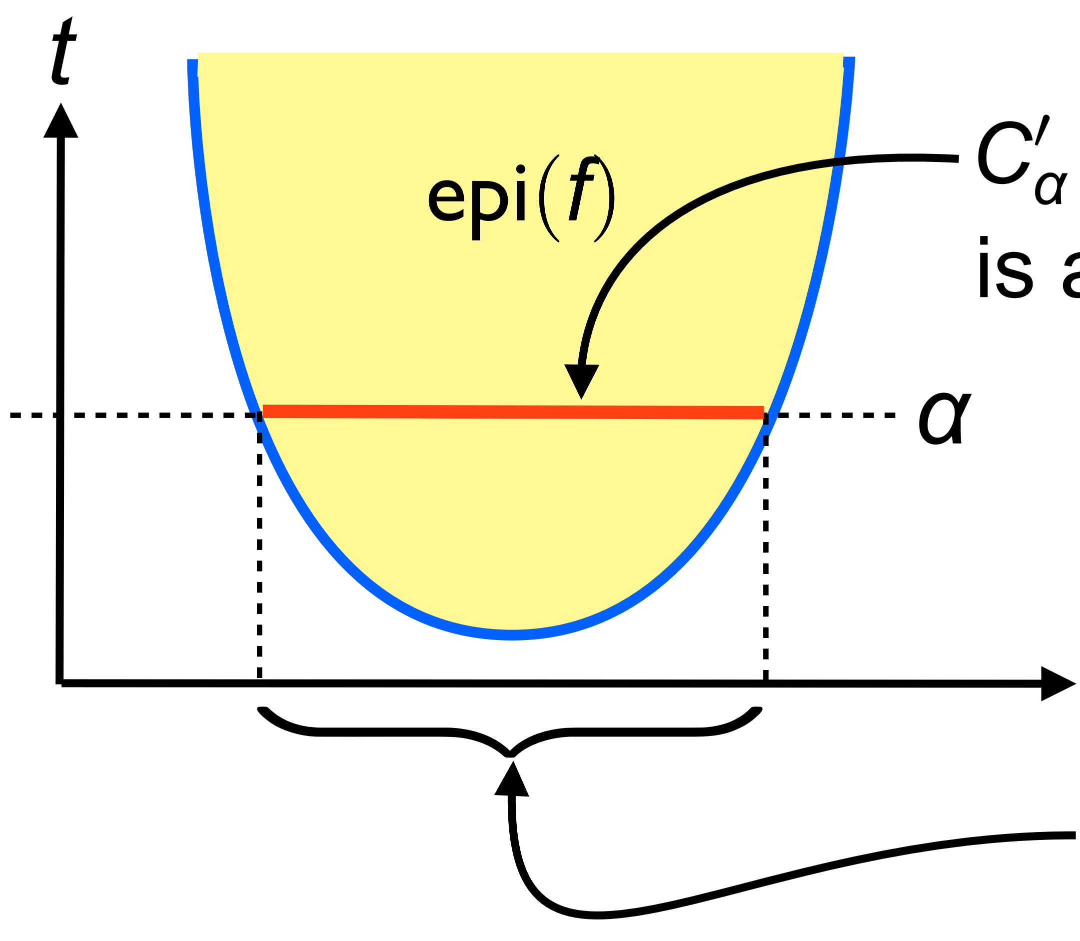


$$f(x) = \begin{cases} 1 & \text{if } a < x \leq b \\ 3 & \text{if } x = a \\ \infty & \text{else} \end{cases}$$

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: If $f : \mathbb{R}^n \to (-\infty, +\infty]$ is convex, then all of its sublevel sets are convex. The reverse implication is *not* true



$C'_\alpha = \{(x, t) \in \mathrm{epi}(f)\} \cap \{(x, t) : t = \alpha\}$
is an intersection of convex sets

$C_\alpha = \{x : (x, t) \in C'_\alpha\}$
is convex as a projection of a convex set

Examples of Convex Functions

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

*Convention: $\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n : x_i > 0 \ \forall i = 1, \ldots, n\}$.

Examples of Convex Functions

Matrix functions: Domain

- Trace functions (i.e., linear functions)

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

- Maximum eigenvalue $f(X) = \lambda_{\max}(X)$ $\qquad \mathbb{S}^n$
- Spectral norm

$$f(X) = \|X\|_2 = \sup_{v \neq 0} \|Xv\|_2 / \|v\|_2 \qquad \mathbb{R}^{m \times n}$$

(maximum scale by which X can stretch a vector)

Checking Convexity along Lines

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

$$g(t) = f(a + tb) \text{ for } a, b \in \mathbb{R}^n$$

is convex in t.

$\Longrightarrow$ One can check convexity of a multivariate function by restricting it to a line.

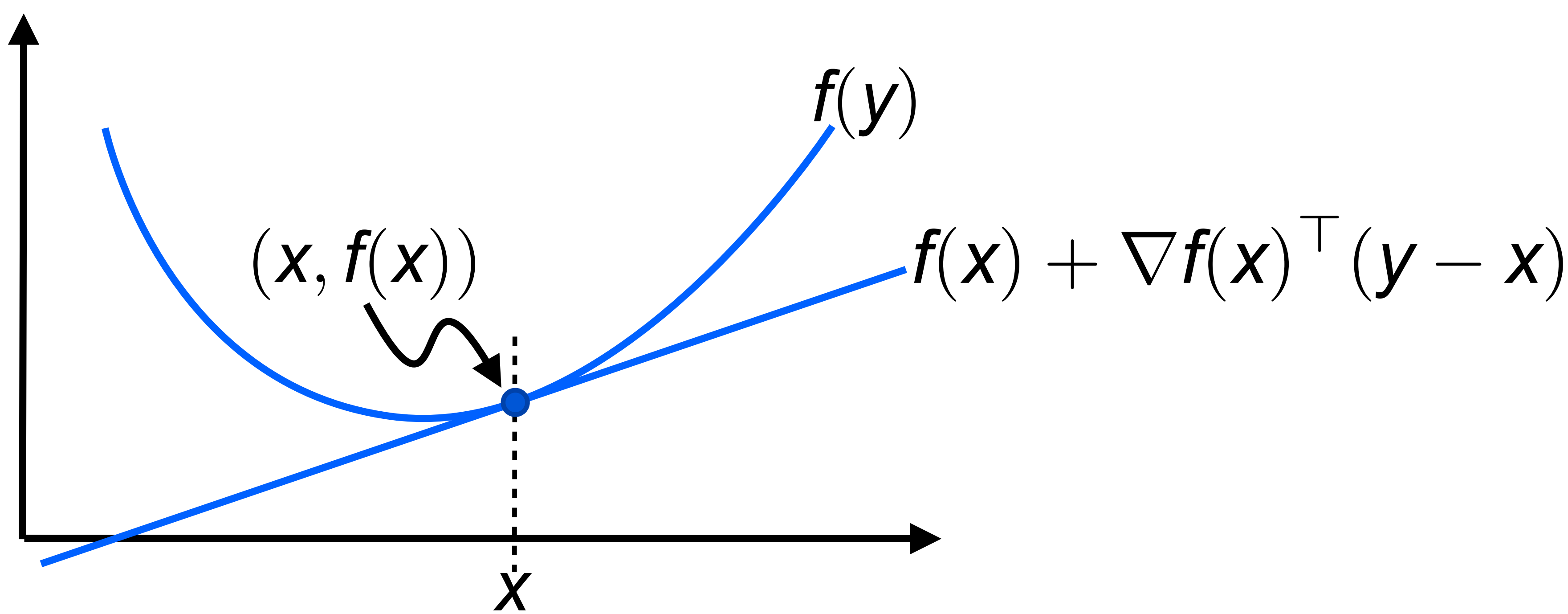
1st-Order Conditions

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

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

⇒ *1st-order Taylor approximation* underestimates f *globally*.



⇒ From *local information* about a convex function we can obtain *global information*.

Univariate Functions

Proposition: A *differentiable function* $f : \mathbb{R} \to \mathbb{R}$ is *convex* if and only if $f(y) \geq f(x) + f'(x)(y - x)\ \forall x, y \in \mathbb{R}$.

Proof:

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

$$
\begin{array}{rcll}
f(x + t(y - x)) & \leq & (1 - t)f(x) + tf(y) & \text{(convexity)} \\
f(y) - f(x) & \geq & [f(x + t(y - x)) - f(x)]/t & \text{(divide by } t\text{)} \\
f(y) - f(x) & \geq & f'(x)(y - x) & \text{(limit } t \downarrow 0\text{)}
\end{array}
$$

$\Longleftarrow:$ For $x, y \in \mathbb{R},\ 0 \leq t \leq 1,$ set $z = tx + (1 - t)y$.

$$
\begin{array}{rcll}
t[f(x) - f(z)] & \geq & tf'(z)(x - z) & \text{(by assumption)} \\
(1 - t)[f(y) - f(z)] & \geq & (1 - t)f'(z)(y - z) & \text{(by assumption)} \\
tf(x) + (1 - t)f(y) & \geq & f(z) & \text{(sum of above)}
\end{array}
$$

1st-Order Conditions

Proposition: A *differentiable function* $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if and only if $f(y) \geq f(x) + \nabla f(x)^\top (y - x) \;\; \forall x, y \in \mathbb{R}^n$.

Proof:

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

$$\begin{array}{rcll} g'(t) & = & \nabla f(ty + (1-t)x)^\top (y - x) & \text{(definition of } g\text{)} \\ g(1) & \geq & g(0) + g'(0) & \text{(univariate case)} \\ f(y) & \geq & f(x) + \nabla f(x)^\top (y - x) & \text{(substitution)} \end{array}$$

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

$$\begin{array}{rcll} f(z) & \geq & f(\tilde{z}) + \nabla f(\tilde{z})^\top (z - \tilde{z}) & \text{(by assumption)} \\ g(t) & \geq & g(\tilde{t}) + g'(\tilde{t})(t - \tilde{t}) & \text{(definition of } g, z, \tilde{z}\text{)} \end{array}$$

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

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

exists at each point in dom(f), and dom(f) is open.

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

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

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 differentiable function* $f : \mathbb{R} \to \mathbb{R}$ is *convex* if and only if $f''(x) \geq 0 \ \forall x \in \mathbb{R}$.

Proof:

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

$$\begin{array}{rcll}
f(y) & \geq & f(x) + f'(x)(y-x) & \text{(1st-order cond.)} \\
f(x) & \geq & f(y) + f'(y)(x-y) & \text{(1st-order cond.)} \\
0 & \geq & [f'(x) - f'(y)]/(y-x) & \text{(sum of above } \times (y-x)^{-2}) \\
0 & \leq & f''(x) & \text{(limit } y \downarrow x)
\end{array}$$

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

$$\begin{array}{rcl}
f(y) & = & f(x) + \int_x^y f'(u) \mathrm{d}u \\
 & = & f(x) + \int_x^y f'(x) + \int_x^u f''(v) \mathrm{d}v \, \mathrm{d}u \\
 & \geq & f(x) + \int_x^y f'(x) \mathrm{d}u \\
 & = & f(x) + f'(x)(y-x)
\end{array}$$

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

2nd-Order Conditions

Proposition: A *twice differentiable function* $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if and only if $\nabla^2 f(x) \succeq 0 \ \forall x \in \mathbb{R}^n$.

Proof:

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

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

$\Longleftarrow:$ Define g as above. Then, we have for any t

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

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

Examples

- *Quadratic functions* $f(x) = \frac{1}{2}x^\top Px + q^\top x + r$

  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* functions of the form $f(x, y) = x^2/y$ are convex as long as $y > 0$ because

$$\nabla^2 f(x, y) = \frac{2}{y^3}\begin{pmatrix} y \\ -x \end{pmatrix}\begin{pmatrix} y \\ -x \end{pmatrix}^\top \succeq 0 \text{ for } y > 0.$$

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 weighted sum
- composition with affine function
- pointwise maximum and supremum
- composition
- minimization
- perspective

Affine Transformations

Affine transformation of *inputs*: If f is convex, then $g(x) = f(Ax + b)$ is also convex.

Non-negative affine transformation of *outputs*: If $f_1, \ldots, f_K$ are convex functions and $\rho_1, \ldots, \rho_K$ are non-negative numbers, then the conic combination $g(x) = \rho_1 f_1(x) + \cdots + \rho_K f_K(x)$ is convex.

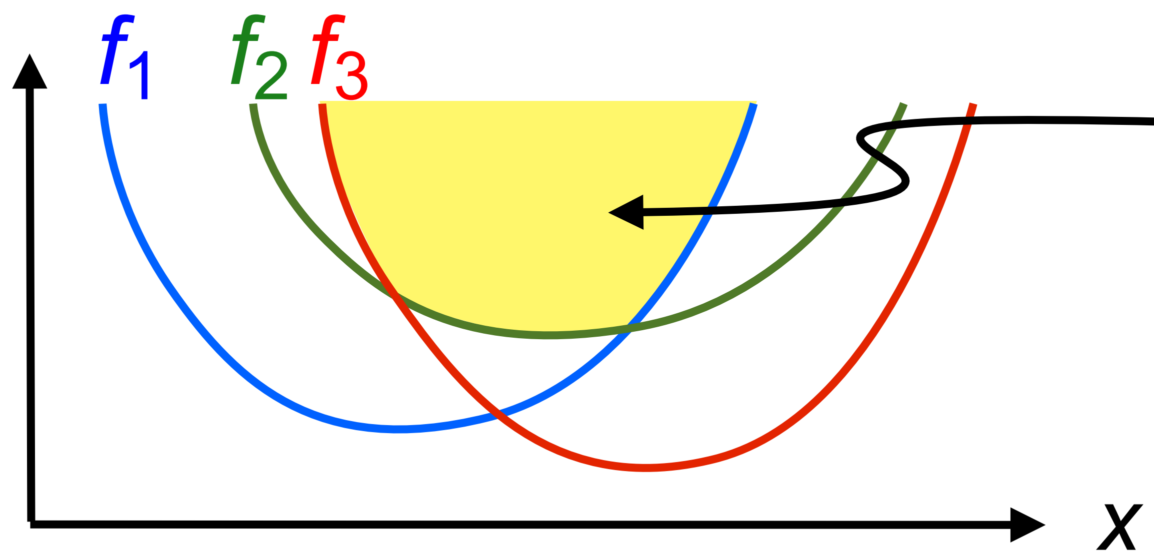
Generalization to *integrals*: If $f(x,y)$ is convex in x for each fixed $y \in \mathcal{Y}$ and $\rho(y)$ is a non-negative function of y, then

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

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.



$\mathrm{epi}(\max\{f_1, f_2, f_3\})$
$= \mathrm{epi}(f_1) \cap \mathrm{epi}(f_2) \cap \mathrm{epi}(f_3)$

(this is convex because *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

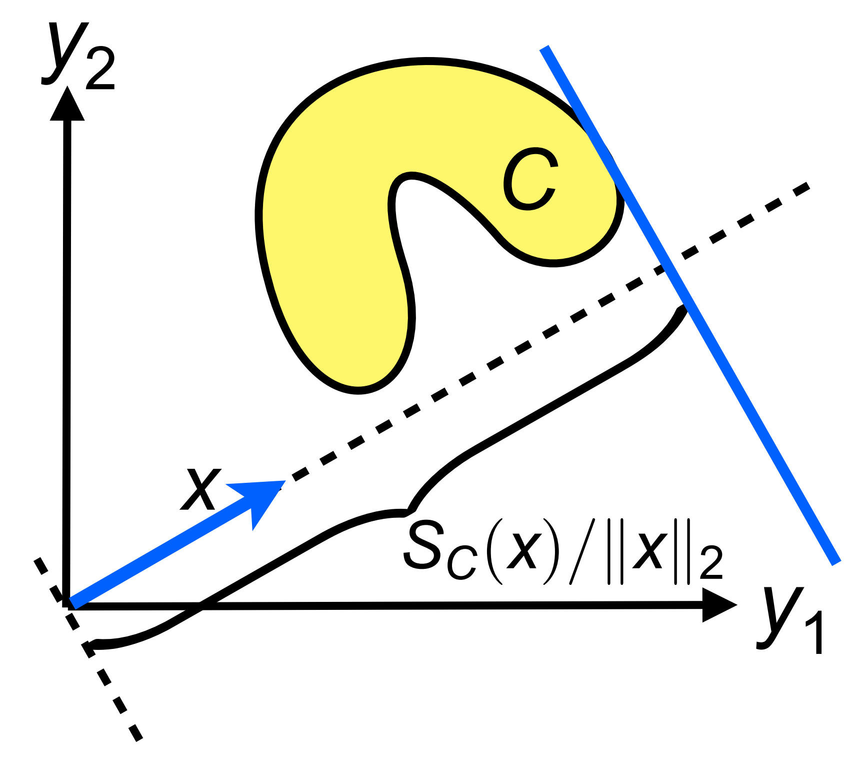
- Piecewise linear functions $f(x) = \max_{k=1,\ldots,K} (a_k^\top x + b_k)$ are convex.
- The sum of the r largest components of $x \in \mathbb{R}^n$ is convex as it can be written as a maximum of linear functions.

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

- The support function of a (possibly nonconvex) set C is convex.

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

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



Examples

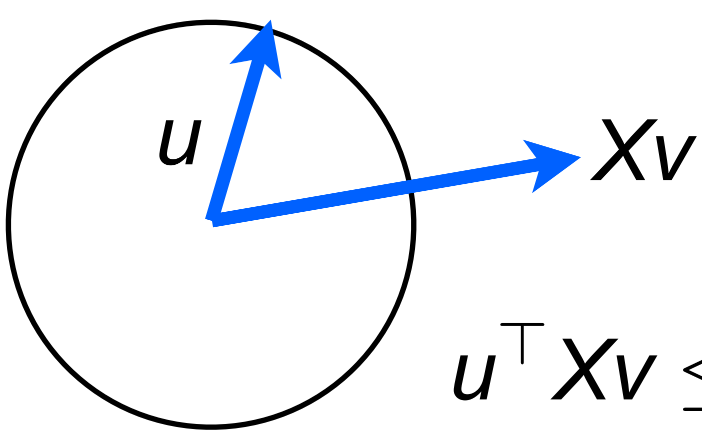
- Maximum eigenvalue $f(X) = \lambda_{\max}(X) \quad (X \in \mathbb{S}^N)$

  Write $X = RDR^\top$ with R orthogonal and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

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

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

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



$$u^\top X v \leq \|u\|_2 \|Xv\|_2 = \|Xv\|_2$$

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*, then $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = h(g(x))$ is convex.

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

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

Thus, f is convex.

Example: $\exp(g(x))$ 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 $g : \mathbb{R}^n \to \mathbb{R}$ is *concave* and $h : \mathbb{R} \to \mathbb{R}$ is *convex* and *non-increasing*, then $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = h(g(x))$ 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

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

is convex.

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

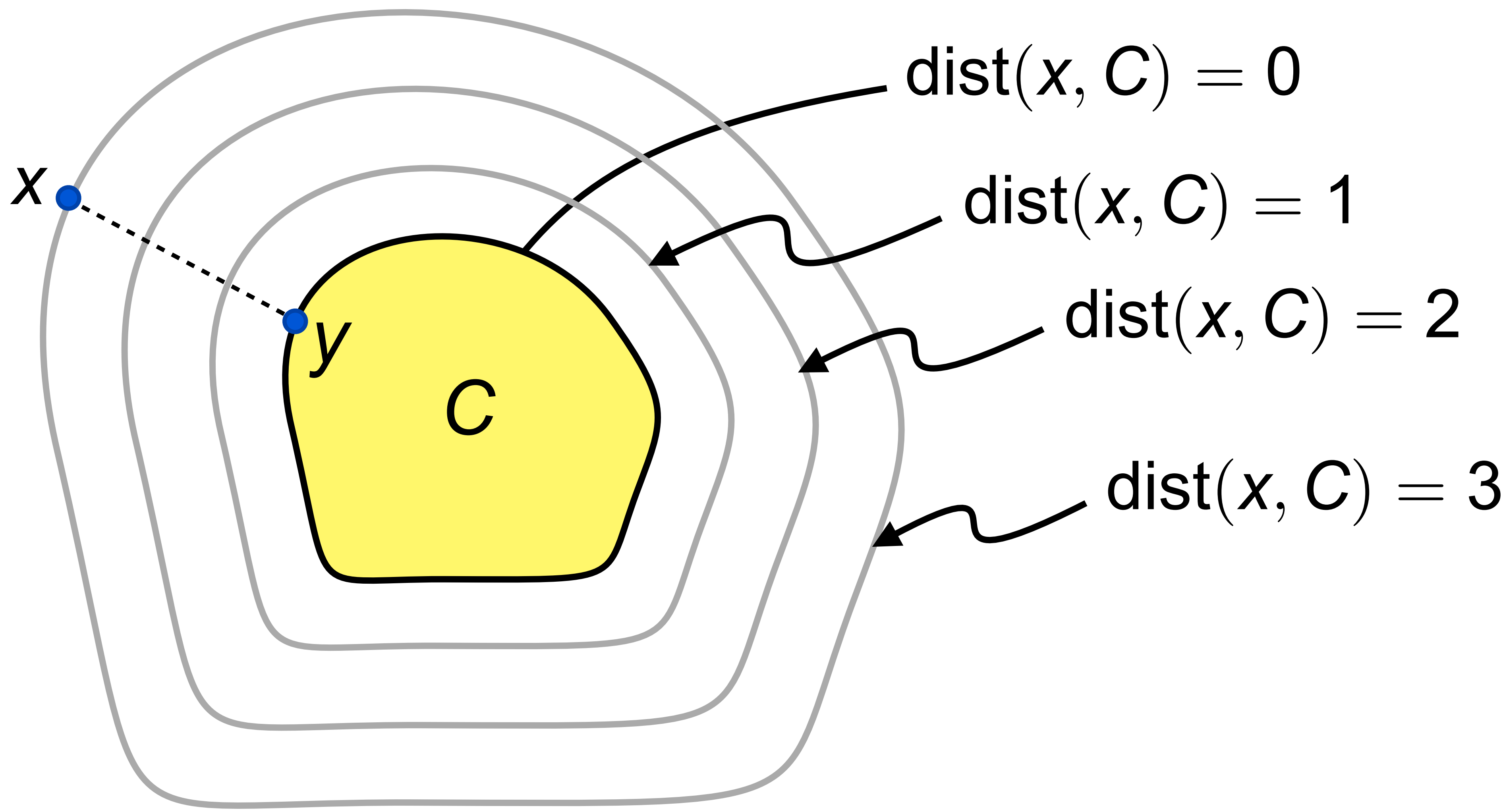
$$
\begin{aligned}
h(\theta x_1 + (1-\theta)x_2) &= \inf_{y \in C}\{f(\theta x_1 + (1-\theta)x_2, y) : g(\theta x_1 + (1-\theta)x_2, y) \leq 0\} \\
&\leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\
&\leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\
&= \theta h(x_1) + (1-\theta) h(x_2)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

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

Distance Function

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

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



Perspective Function

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

$$g(x,t) = tf(x/t), \ \mathrm{dom}(g) = \{(x,t) : (x/t) \in \mathrm{dom}(f), \ t > 0\},$$

is convex in (x,t).

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

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

Thus, g is convex in (x,t).

Relative Entropy

As the negative logarithm $f(x) = -\log(x)$ is convex on $\mathbb{R}_{++}$, we conclude that its perspective

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

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

The *relative entropy* of two vectors $u, v \in \mathbb{R}^n_{++}$ is defined as

$$\sum_{i=1}^{n} u_i \log(u_i/v_i).$$

It is convex because it is a sum of n convex functions.

Convexity w.r.t. Generalized Inequalities

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

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

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,

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

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.

Main Take-Away Points

- **Definitions:** 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
- **Generalized inequalities:** constructing convex sets using K-convex constraint functions and conic inequalities