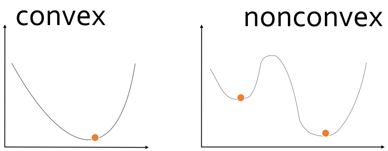
Lecture 2 Convex Sets and Functions

Can Li

ChE 597: Computational Optimization
Purdue University

Why Convexity?

For convex functions, local minima are global minima.



Global Optimum: A point x^* is a global minimum of $f(x)$ if for all x in the domain of f,

$$f(x^*) \leq f(x).$$

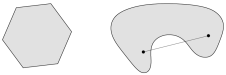
Local Minimum: A point x_0 is a local minimum of $f(x)$ if there exists $\delta > 0$ such that for all x within $d_X(|x - x_0|) < \delta$,

$$f(x_0) \leq f(x)$$

Convex sets and functions

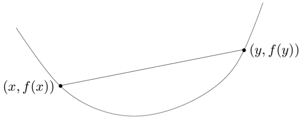
Convex set: A set $C \subseteq \mathbb{R}^n$ is convex if, for any $x, y \in C$, the line segment between x and y is contained in C. That is,

$$\forall x, y \in C \implies \lambda x + (1-\lambda)y \in C, \; \forall\, 0 \leq \lambda \leq 1$$



Convex function: A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain $\mathrm{dom}(f)$ is a convex set and if, for any $x, y \in \mathrm{dom}(f)$, the following inequality holds:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \text{ for all } 0 \leq \lambda \leq 1$$



Convex Optimization Problems

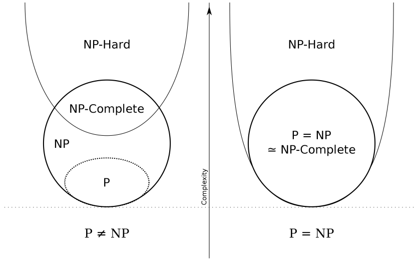
Optimization problem:

$$
\begin{aligned}
\min_{x \in D} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m \\
& h_j(x) = 0, \quad j = 1, \dots, r
\end{aligned}
$$

- Here $D = \text{dom}(f) \cap \bigcap_{i=1}^{m} \text{dom}(g_i) \cap \bigcap_{j=1}^{p} \text{dom}(h_j)$, common domain of all the functions
- This is a *convex optimization problem* provided the functions f and $g_i, i = 1, \dots, m$ are convex, and $h_j, j = 1, \dots, p$ are affine: $h_j(x) = a_j^T x + b_j, \quad j = 1, \dots, p$
- Not the focus of this class. Take AAE 561/IE 561 if interested in more details.

Convex Optimization and Polynomial Solvability

- Convex optimization problems are in **P (polynomial time solvable)**.
- **NP (Nondeterministic Polynomial time):** A complexity class that includes decision problems for which a given solution can be verified in polynomial time.
- **NP-hard:** A classification of problems to which all problems in NP can be reduced in polynomial time, and they are at least as hard as the hardest problems in NP.
- MILP/ nonconvex QCQP/ MINLP are NP-hard in general



Combinations of Points(vectors)

Given points (vectors) $x_1, x_2, \ldots, x_n \in \mathbb{R}^n$ and weights $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}^1$, we define:

- **Convex Combination:** A combination $\sum_{i=1}^{n} \lambda_i x_i$ where $\lambda_i \geq 0$ for all i and $\sum_{i=1}^{n} \lambda_i = 1$. It represents a point inside the polytope formed by $x_1, \ldots, x_n$.
- **Affine Combination:** A combination $\sum_{i=1}^{n} \lambda_i x_i$ where $\sum_{i=1}^{n} \lambda_i = 1$ but λ_i are not restricted to be non-negative. It represents any point on the affine hull of the points, extending beyond the polytope.
- **Conic Combination:** A combination $\sum_{i=1}^{n} \lambda_i x_i$ where $\lambda_i \geq 0$ for all i, without the requirement that they sum to one. It represents a point in the cone spanned by the points.
- **Linear Combination:** A combination $\sum_{i=1}^{n} \lambda_i x_i$ with no restrictions on λ_i. It represents any point in the space spanned by the vectors.

Convex Combinations in Convex Sets

Claim: If C is a convex set and $x_1, x_2, \ldots, x_n$ are points in C, then any convex combination of these points also lies in C. **Proof by induction:** *Base case* ($n = 2$): For two points $x_1, x_2 \in C$, the convex combination $\lambda x_1 + (1 - \lambda)x_2$ is in C by the definition of convexity, for any λ such that $0 \leq \lambda \leq 1$.

Inductive step: Assume the statement is true for any $n - 1$ points in C. Now consider n points $x_1, x_2, \ldots, x_n \in C$ and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be non-negative numbers that sum to 1. Consider the convex combination $y = \sum_{i=1}^{n} \lambda_i x_i$. Without loss of generality, assume $\lambda_n \neq 1$. We can write y as:

$$y = \lambda_n x_n + (1 - \lambda_n) \left(\sum_{i=1}^{n-1} \frac{\lambda_i}{1 - \lambda_n} x_i \right)$$

By the inductive hypothesis, the term in the parentheses is a convex combination of the $n - 1$ points and thus lies in C. Since C is convex, the entire expression for y also lies in C.

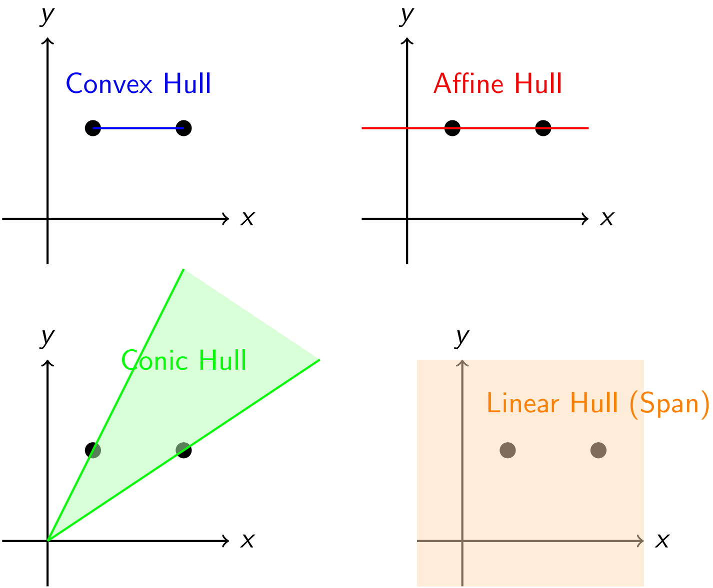
Conclusion: By induction, any convex combination of n points in C lies in C.

Hulls and Spans of a Set C

Given a set C in a vector space, we define:

- **Convex Hull:** The smallest convex set that contains all the elements of C. It is the set of all convex combinations of finite subsets of C.
- **Affine Hull:** The smallest affine set that contains all the elements of C. It is the set of all affine combinations of finite subsets of C, forming an affine subspace.
- **Conic Hull:** The smallest cone that contains all the elements of C. It consists of all conic combinations of finite subsets of C, forming a cone with its vertex at the origin.
- **Linear Hull (Linear Span):** The set of all linear combinations of elements in C. This set forms the smallest subspace that contains all the elements of C.

Geometric Interpretation



Examples of Convex Sets

- **Trivial ones**: empty set, point, line
- **Norm ball**: $\{x : \|x\| \leq r\}$, for given norm $\|\cdot\|$, radius r
- **Hyperplane**: $\{x : a^T x = b\}$, for given a, b
- **Halfspace**: $\{x : a^T x \leq b\}$
- **Affine space**: $\{x : Ax = b\}$, for given A, b
- **Polyhedron**: $\{x : Ax \leq b\}$, where inequality $\leq$ is interpreted componentwise. Note: the set $\{x : Ax \leq b, Cx = d\}$ is also a polyhedron.

Operations Preserving Convexity

- **Intersection**: the intersection of convex sets is convex.
- **Scaling and translation**: if C is convex, then

$$aC + b = \{ax + b : x \in C\}$$

  is convex for any a, b.
- **Affine images and preimages**: if $f(x) = Ax + b$ and C is convex then

$$f(C) = \{f(x) : x \in C\}$$

  is convex, and if D is convex then

$$f^{-1}(D) = \{x : f(x) \in D\}$$

  is convex.

More Operations Preserving Convexity

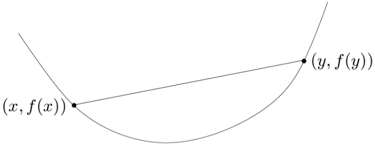
- **Perspective images and preimages:** the perspective function is $\mathbf{P} : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ (where $\mathbb{R}_{++}$ denotes positive reals),
$$\mathbf{P}(x, z) = \frac{x}{z}$$
for $z > 0$. If $C \subseteq \mathrm{dom}(\mathbf{P})$ is convex then so is $\mathbf{P}(C)$, and if D is convex then so is $\mathbf{P}^{-1}(D)$
- **Linear-fractional images and preimages:** the perspective map composed with an affine function,
$$f(x) = \frac{Ax + b}{c^T x + d}$$
is called a **linear-fractional function**, defined on $c^T x + d > 0$. If $C \subseteq \mathrm{dom}(f)$ is convex then so if $f(C)$, and if D is convex then so is $f^{-1}(D)$

Convex Functions

Convex function: A function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\mathrm{dom}(f) \subseteq \mathbb{R}^n$ is convex, and

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

for all $x, y \in \mathrm{dom}(f)$ and $0 \leq t \leq 1$.



In words, a function lies below the line segment joining $f(x)$, $f(y)$.
Concave function: The opposite inequality above, so that f concave $\Leftrightarrow -f$ convex.

Important Modifiers

- **Strictly convex:** A function f is strictly convex if

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y)$$

  for all $x \neq y$ and $0 < t < 1$. In words, f is convex and has greater curvature than a linear function.

- **Strongly convex** with parameter $m > 0$: A function f is strongly convex if

$$f - \frac{m}{2}\|x\|^2$$

  is convex. In words, f is at least as convex as a quadratic function.

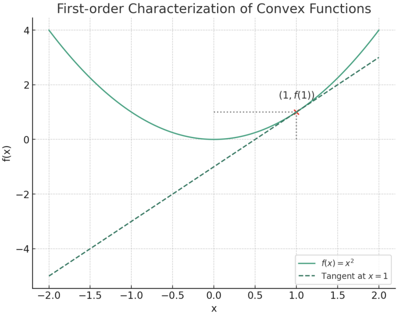
Note: Strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex.
(Analogously for concave functions)

First-order Characterization of Convex Functions

- If f is differentiable, then f is convex if and only if $\text{dom}(f)$ is convex, and

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

for all $x, y \in \text{dom}(f)$.



First-order Characterization of Convex Functions

If f is differentiable, then f is convex if and only if for all $x, y \in \text{dom}(f)$,

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

. **Proof:**

- For the forward direction, assume f is convex. By definition, for any $x, y \in \text{dom}(f)$ and $0 \leq \lambda \leq 1$,

$$f(\lambda y + (1 - \lambda)x) \leq \lambda f(y) + (1 - \lambda) f(x).$$

$$f(x + \lambda(y - x)) - f(x) \leq \lambda(f(y) - f(x)).$$

$$\frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \leq (f(y) - f(x)).$$

Taking the limit as λ approaches 0, the inequality becomes

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

where $\nabla f(x)^T (y - x)$ is the directional derivative.

First-order Characterization of Convex Functions

If f is differentiable, then f is convex if and only if for all $x, y \in \mathrm{dom}(f)$,

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

. **Proof:**

- For the reverse direction, suppose the inequality holds for all $x, y \in \mathrm{dom}(f)$. Consider any $x, y \in \mathrm{dom}(f)$ and $0 < \lambda < 1$, let $z = \lambda x + (1 - \lambda) y$ and apply the given inequality twice:

$$f(x) \geq f(z) + \nabla f(z)^T (x - z)$$

$$f(y) \geq f(z) + \nabla f(z)^T (y - z)$$

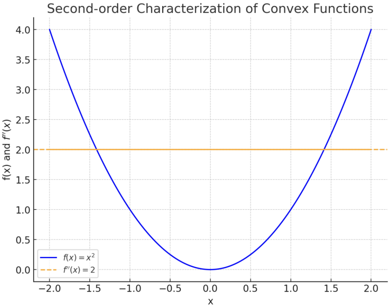
Multiplying the first inequality by λ and the second by $1 - \lambda$, and adding them yields

$$\lambda f(x) + (1 - \lambda) f(y) \geq f(z) = f(\lambda x + (1 - \lambda) y),$$

which is the definition of convexity for f.

Second-order Characterization of Convex Functions

- If f is twice differentiable, then f is convex if and only if $\mathrm{dom}(f)$ is convex, and $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}(f)$.



Second-order Characterization of Convex Functions

If f is twice differentiable, then f is convex if and only if $\mathrm{dom}(f)$ is convex, and $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}(f)$.

- **Necessity** $\Rightarrow$ For contradiction, assume that there exists x^0 such that $\nabla^2 f\left(x^0\right)$ is not positive semidefinite. Then we can choose a vector p such that $p^T \nabla^2 f\left(x^0\right) p < 0$, and because $\nabla^2 f$ is continuous near x^0, there is a scalar $\delta > 0$ such that $p^T \nabla^2 f\left(x^0 + tp\right) p < 0$ for all $t \in [-\delta, \delta]$. Using the mean value theorem from calculus at $x^0 + \delta p$ and $x^0 - \delta p$ we have

$$f\left(x^0 + \delta p\right) = f\left(x^0\right) + \delta p^T \nabla f\left(x^0\right) + \frac{1}{2}\delta^2 p^T \nabla^2 f\left(x^0 + t_1 p\right) p$$

$$f\left(x^0 - \delta p\right) = f\left(x^0\right) - \delta p^T \nabla f\left(x^0\right) + \frac{1}{2}\delta^2 p^T \nabla^2 f\left(x^0 + t_2 p\right) p$$

  for some $t_1 \in [0, \delta], t_2 \in [-\delta, 0]$. Add them up we have
  $f\left(x^0 + \delta p\right) + f\left(x^0 - \delta p\right) =$
  $2f\left(x^0\right) + \frac{1}{2}\delta^2 p^T \nabla^2 f\left(x^0 + t_1 p\right) p + \frac{1}{2}\delta^2 p^T \nabla^2 f\left(x^0 + t_2 p\right) p < 2f\left(x^0\right)$
  Note that $x^0 = \frac{1}{2}(x^0 + \delta p + x^0 - \delta p)$, which violates the definition of convexity.

Second-order Characterization of Convex Functions

- If f is twice differentiable, then f is convex if and only if $\mathrm{dom}(f)$ is convex, and $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}(f)$.

Proof:

- **Sufficiency**. $\forall x, y \in \mathrm{dom}(f)$. Mean value theorem from calculus:

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(z)(y - x)$$

where $\nabla^2 f(z)$ is the Hessian matrix of f at some point z on the line segment between x and y. since $\nabla^2 f(x) \succeq 0$ $\frac{1}{2}(y - x)^T \nabla^2 f(z)(y - x) \geq 0$. Thus, we have

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

which is equivalent to f being convex according to the first-order condition.

Examples of Convex Functions

- Univariate functions:
  - Exponential function: e^{ax} is convex for any a over $\mathbb{R}$
  - Power function: x^a is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$ (nonnegative reals)
  - Power function: x^a is concave for $0 \leq a \leq 1$ over $\mathbb{R}_+$
  - Logarithmic function: $\log x$ is concave over $\mathbb{R}_{++}$
- Affine function: $a^T x + b$ is both convex and concave
- Quadratic function: $\frac{1}{2}x^T Qx + b^T x + c$ is convex provided that $Q \succeq 0$ (positive semidefinite)
- Least squares loss: $\|y - Ax\|_2^2$ is always convex (since $A^T A$ is always positive semidefinite)

- **Norm:** $\|x\|$ is convex for any norm; e.g., ℓ_p norms,

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \quad \text{for } p \geq 1, \qquad \|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$$

- **Support function:** for any set C (convex or not), its support function

$$I_C^*(x) = \max_{y \in C} x^T y$$

is convex.

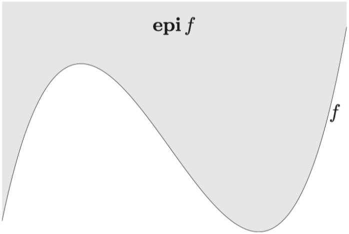
- **Max function:** $f(x) = \max\{x_1, \ldots, x_n\}$ is convex.

Key Properties of Convex Functions

- A function is convex if and only if it is convex on all lines, i.e., the function $g(t) = f(x_0 + tv)$ is convex in t for all $x_0 \in \operatorname{dom} f$ and all vector v.
- **Epigraph characterization:** a function f is convex if and only if its epigraph

$$\operatorname{epi}(f) = \{(x, t) \in \operatorname{dom}(f) \times \mathbb{R} : f(x) \leq t\}$$

is a convex set.



Operations preserving convexity

- **Nonnegative linear combination:** $f_1, \ldots, f_m$ convex implies $a_1 f_1 + \ldots + a_m f_m$ convex for any $a_1, \ldots, a_m \geq 0$
- **Pointwise maximization:** if f_s is convex for any $s \in S$, then $f(x) = \max_{s \in S} f_s(x)$ is convex. Note that the set S here (number of functions f_s) can be infinite
- **Partial minimization:** if $g(x, y)$ is convex in x, y, and C is convex, then $f(x) = \min_{y \in C} g(x, y)$ is convex

Example: distances to a set

Let C be an arbitrary set, and consider the **maximum distance** to C under an arbitrary norm $\|\cdot\|$:

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

Let's check convexity: $f_y(x) = \|x - y\|$ is convex in x for any fixed y, so by pointwise maximization rule, f is convex.
Now let C be convex, and consider the **minimum distance** to C:

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

Let's check convexity: $g(x, y) = \|x - y\|$ is convex in x, y jointly, and C is assumed convex, so apply partial minimization rule.