Math 6880/7875: Advanced Optimization
Convexity

Akil Narayan[1]

[1] Department of Mathematics, and Scientific Computing and Imaging (SCI) Institute
University of Utah

January 18, 2022

Optimization problems

General optimization problem:

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S \subseteq \mathbb{R}^n$$
$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

We have discussed:

- Local, global solutions
- Local optimality for unconstrained optimization: first-order necessary conditions, second-order necessary + sufficient conditions.
- Local optimality for constrained optimization: KKT conditions, first-order necessary conditions

KKT example

$\{x : g_i(x) = 0\}$ – a set in $\mathbb{R}^n$
(typically nonconvex)

The KKT conditions:

$$\nabla f(x^*) + \sum_{i \in [m]} \lambda_i \nabla g_i(x^*) = 0 \qquad \text{(Stationarity)}$$

$$\lambda_i g_i(x^*) = 0, \; i \in [m] \qquad \text{(Complementary Slackness)}$$

$$g_i(x^*) \leqslant 0, \; i \in [m] \qquad \text{(Primal feasibility)}$$

$$\lambda_i \geqslant 0, \; i \in [m] \qquad \text{(Dual feasibility)}$$

Example

Let's consider the KKT conditions for the following optimization problem,

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x_j \geqslant 0, \qquad j \in [n]. \end{aligned}$$

$$\min f(x)$$
$$\text{s.t. } x_j \geq 0 \quad \forall j = 1, \dots, n$$

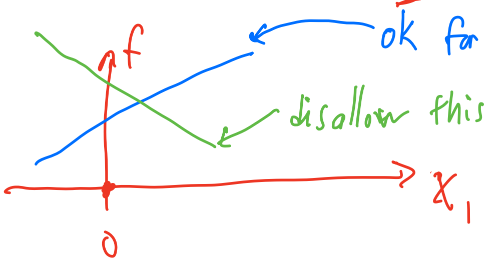
$\mathbb{R}^n$

KKT conditions: enforce "no feasible descent"

In interior of S: this means $\nabla f = 0$

on boundary, e.g. $x_1 = 0$, $x_j > 0 \quad \forall j = 2, \dots n$

Slice at constant $x_2, x_3, \dots$



allow $\frac{\partial f}{\partial x_1} > 0$ if $x_1 = 0$

ensure $\frac{\partial f}{\partial x_j} = 0$ if $j = 2, \dots n$

The KKT conditions coincide with the conditions above:

$g_i(x) \leq 0 \quad i = 1 \dots n$

$g_i(x) = -x_i \iff x_i \geq 0$

$\nabla g_i = -e_i$ cardinal unit vector in ith direction

Stationarity: $\nabla f + \sum_{j=1}^{n} \lambda_j \nabla g_j = 0 \rightarrow \nabla f = \lambda$

dual feasibility: $\lambda_j \geq 0 \quad j = 1 \dots n$

primal feasibility: $g_j(x) \leq 0 \quad j = 1 \dots n$

comp. slackness: $\underbrace{\lambda_i g_i(x) = 0}_{\downarrow} \quad i = 1 \dots n$

$$-\lambda_i x_i = 0$$

$$\left.\begin{array}{l} \lambda_i = 0 \Longrightarrow \dfrac{\partial f}{\partial x_i} = \lambda_i = 0 \\ \qquad\qquad x_i \geq 0 \end{array}\right\} x_i \text{ in interior}$$

$$\left.\begin{array}{l} x_i = 0 \Longrightarrow \lambda_i \geq 0 \\ \qquad\qquad \dfrac{\partial f}{\partial x_i} \geq 0 \end{array}\right\} x_i \text{ on boundary.}$$

I.e., KKT conditions are exactly the same as previously motivated.

KKT example summary

Example

Let's consider the KKT conditions for the following optimization problem,

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x_j \geqslant 0, & j \in [n]. \end{aligned}$$

LICQ always holds. The KKT conditions for x^* are given by,

$$\begin{aligned} \mathcal{A}(x^*) &= \left\{ i \in [n] \,\middle|\, x_i^* = 0 \right\} \\ \mathcal{I}(x^*) &= [n] \backslash \mathcal{A}(x^*) \\ \frac{\partial f}{\partial x_i} &= 0, & i \in \mathcal{I}(x^*) \\ \frac{\partial f}{\partial x_i} &\geqslant 0, & i \in \mathcal{A}(x^*). \end{aligned}$$

These conditions are highly suggestive of a projected gradient descent algorithm.

Optimization problems

General optimization problem:

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$
$$S := \left\{ x \,\middle|\, g_i(x) \leqslant 0,\ i \in [m] \right\}$$

We have discussed:

- Local, global solutions
- Local optimality for unconstrained optimization: first-order necessary conditions, second-order necessary + sufficient conditions.
- Local optimality for constrained optimization: KKT conditions, first-order necessary conditions

Generally, we cannot do better than local optimality, and the necessary vs sufficient characterizations can be complicated.

A property that makes everything a lot cleaner and clearer: **convexity**.

Optimization problems

General optimization problem:

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S$$
$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

We have discussed:

- Local, global solutions
- Local optimality for unconstrained optimization: first-order necessary conditions, second-order necessary + sufficient conditions.
- Local optimality for constrained optimization: KKT conditions, first-order necessary conditions

Generally, we cannot do better than local optimality, and the necessary vs sufficient characterizations can be complicated.

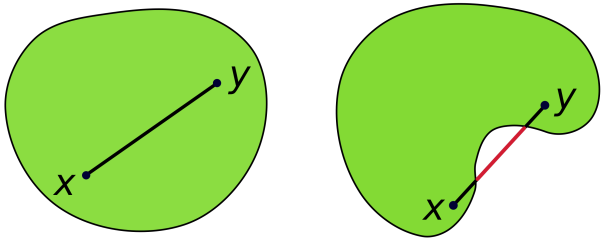
A property that makes everything a lot cleaner and clearer: **convexity**.

Convex sets

Definition

A set $S \subset \mathbb{R}^n$ is **convex** if, $\forall\ x, y \in S$ and $\forall\ \lambda \in [0, 1]$, then

$$\lambda x + (1 - \lambda) y \in S.$$



There are some generalized notions of convexity:

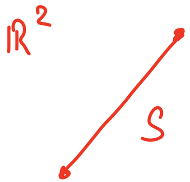
- S is *strictly* convex if $\lambda x + (1 - \lambda) y \in \mathrm{int}(S)$ for all $\lambda \in (0, 1)$ and all $x, y \in S$ with $x \neq y$.
- S is *absolutely* convex if it is convex and balanced (S is balanced if $cx \in S$ for all scalars c satisfying $|c| \leqslant 1$ and all $x \in S$)

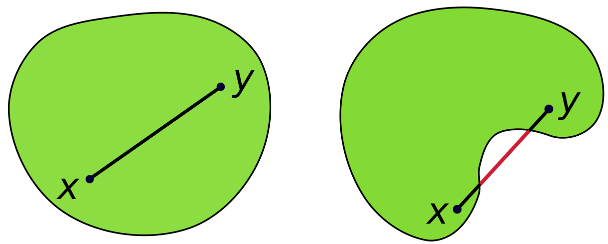
Convex sets

Definition

A set $S \subset \mathbb{R}^n$ is **convex** if, $\forall\ x, y \in S$ and $\forall\ \lambda \in [0,1]$, then

$$\lambda x + (1-\lambda) y \in S.$$





S: convex

S: not strictly convex

By CheCheDaWaff - These files were derived from: Convex polygon illustration1,2.png:, CC BY-SA 4.0,

https://commons.wikimedia.org/w/index.php?curid=49541588,

https://commons.wikimedia.org/w/index.php?curid=49543150

There are some generalized notions of convexity:

- S is *strictly* convex if $\lambda x + (1-\lambda) y \in \operatorname{int}(S)$ for all $\lambda \in (0,1)$ and all $x, y \in S$ with $x \neq y$.
- S is *absolutely* convex if it is convex and balanced (S is balanced if $cx \in S$ for all scalars c satisfying $|c| \leqslant 1$ and all $x \in S$)

Examples

Some examples of convex sets:

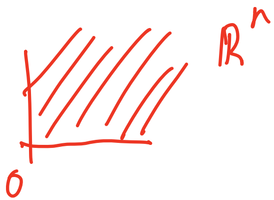
1 – All intervals (open, closed, half-open, unbounded) on $\mathbb{R}$

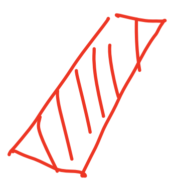
2 – The non-negative orthant: $x \in \mathbb{R}^n$ with $x_j \geqslant 0$, $j \in [n]$

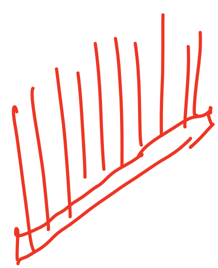
3 – Half-spaces: with $\langle \cdot, \cdot \rangle$ the Euclidean inner product and for any $y \in \mathbb{R}^n$ and $a \in \mathbb{R}$, the set $\{x \in \mathbb{R}^n \mid \langle x, y \rangle \leqslant a\}$ is convex. The set $H(y,a) = \{x \in \mathbb{R}^n \mid \langle x, y \rangle = a\}$ is a **hyperplane**.

4 – The set of positive-definite matrices in $\mathbb{R}^{n \times n}$ (or also positive semi-definite matrices)

1.) 

2.) 

3.) $\mathbb{R}^3$ $H(y,a)$ 

halfspace 

Examples

Some examples of non-convex sets:

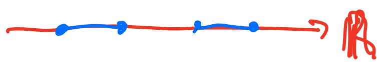
1 – Any disconnected set in $\mathbb{R}$

2 – The graph of $y = 1/x$ in $\mathbb{R}^2$

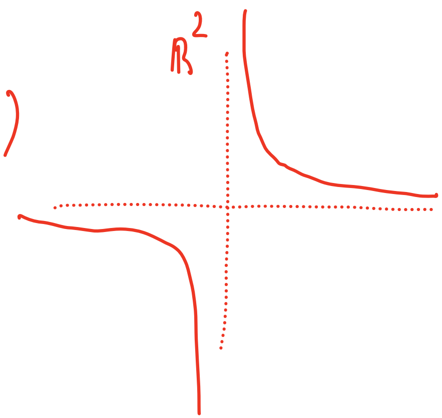
3 – The set of vectors $x \in \mathbb{R}^n$ with at most s nonzero entries $(0 < s < n)$

4 – The set of invertible/full-rank matrices in $\mathbb{R}^{n \times n}$

1.)



2.)



Convex combinations

Convexity extends to general **convex combinations**.

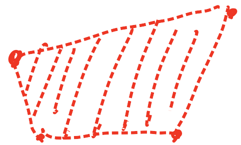
Let Δ^n be the unit simplex in n dimensions, i.e.,

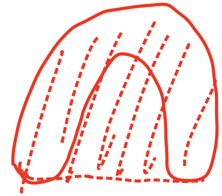
$$\Delta^n := \left\{ \lambda \in \mathbb{R}^n \;\middle|\; 0 \leqslant \lambda_j \leqslant 1 \; \forall \; j \in [n], \text{ and } \sum_{j=1}^{n} \lambda_j = 1 \right\}$$

S is convex if and only if all convex combinations of points in S lie in S.
I.e., $\forall \; \lambda \in \Delta^n$ and $x_1, \ldots, x_n \in S$, then

$$\sum_{j=1}^{n} \lambda_j x_j \in S.$$

Convex hulls





Given a set of points S, one can consider the "filled in" version of S, where the rule for filling in is convexity.

Definition

Let $S \in \mathbb{R}^n$. The **convex hull** of S is, equivalently, any of the following:

- The set of all convex combinations of points in S.
- The intersection of all convex sets containing S.
- The smallest convex set containing S (which is unique).

We write $\mathrm{Conv}(S)$ for the convex hull of S.

Convex hulls are convex sets (always).

Convex hulls are a fundamental operation in computational geometry (and statistics, and economics, and optimization, and ....).

Cool things about convex sets

Convex sets are "nice" sets in many senses.

There are numerous neat properties a convex set $S \subseteq \mathbb{R}^n$ obeys:

- All intersections of convex sets are convex.
- If S is also closed, then it can be represented as the intersection of half-spaces.

Cool things about convex sets

Convex sets are "nice" sets in many senses.

There are numerous neat properties a convex set $S \subseteq \mathbb{R}^n$ obeys:

- All intersections of convex sets are convex.
- If S is also closed, then it can be represented as the intersection of half-spaces.

Cool things about convex sets

Convex sets are "nice" sets in many senses.

There are numerous neat properties a convex set $S \subseteq \mathbb{R}^n$ obeys:

– Carathéodory's Theorem: If $x \in \mathrm{Conv}(S)$, then there exist $z_1, \ldots, z_{n+1} \in S$ and $\lambda \in \Delta^{n+1}$ such that,

$$x = \sum_{j=1}^{n+1} \lambda_j z_j .$$

"Every point in an n-dimensional convex hull can be written as a convex combination of as few as $n + 1$ points."

– **Extreme points** of S are points not lying on any line segment inside S. Krein-Milman Theorem: Every compact, convex set equals the convex hull of its extreme points.

Cool things about convex sets

Convex sets are "nice" sets in many senses.

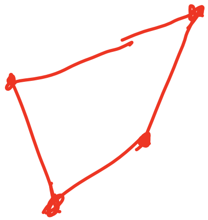
There are numerous neat properties a convex set $S \subseteq \mathbb{R}^n$ obeys:

- Carathéodory's Theorem: If $x \in \mathrm{Conv}(S)$, then there exist $z_1, \ldots, z_{n+1} \in S$ and $\lambda \in \Delta^{n+1}$ such that,

$$x = \sum_{j=1}^{n+1} \lambda_j z_j.$$

  "Every point in an n-dimensional convex hull can be written as a convex combination of as few as $n+1$ points."

- **Extreme points** of S are points not lying on any line segment inside S. Krein-Milman Theorem: Every compact, convex set equals the convex hull of its extreme points.



Cool things about convex sets

Convex sets are "nice" sets in many senses.

There are numerous neat properties a convex set $S \subseteq \mathbb{R}^n$ obeys:

- Supporting hyperplane theorem: If x is a boundary point of S, then there is a hyperplane that (i) contains x and (ii) defines a closed half-space that entirely contains S.
- Separating hyperplane theorem: If S, T are two disjoint convex sets, then there exists a hyperplane $H(y, a)$ such that

$$S \subseteq \{x \in \mathbb{R}^n \quad \langle x, y \rangle \leqslant a\}$$
$$T \subseteq \{x \in \mathbb{R}^n \quad \langle x, y \rangle \geqslant a\}.$$

Cool things about convex sets

Convex sets are “nice” sets in many senses.

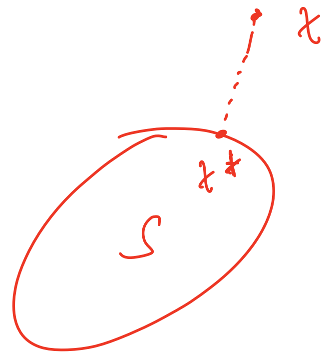
There are numerous neat properties a convex set $S \subseteq \mathbb{R}^n$ obeys:

- Supporting hyperplane theorem: If x is a boundary point of S, then there is a hyperplane that (i) contains x and (ii) defines a closed half-space that entirely contains S.
- Separating hyperplane theorem: If S, T are two disjoint convex sets, then there exists a hyperplane $H(y, a)$ such that

$$
\begin{aligned}
S &\subseteq \{x \in \mathbb{R}^n \quad \langle x, y \rangle \leqslant a\} \\
T &\subseteq \{x \in \mathbb{R}^n \quad \langle x, y \rangle \geqslant a\}.
\end{aligned}
$$

Hilbert Projection Theorem: Let S be convex and closed. Then $\forall\ x \in \mathbb{R}^n$, $\exists$ unique x^* that solves the optimization problem:

$$\underset{y \in S}{\operatorname{argmin}} \|y - x\|_2$$



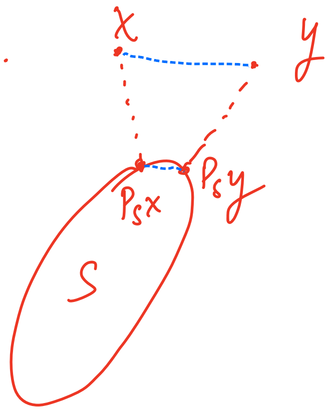
This motivates defining a projection operator:

$$P_S(x) = \underset{y \in S}{\operatorname{argmin}} \|y - x\|_2$$

<u>Lemma</u>: Projections onto convex sets are non-expansive

Given $x, y \in \mathbb{R}^n$, S closed + convex.

Then $\|P_S x - P_S y\|_2 \leq \|x - y\|_2$.



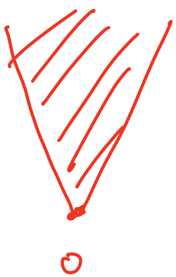
Cones and conic hulls

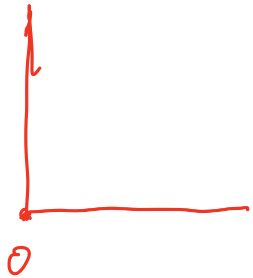
A closely related notion to convex sets is that of conic sets.

Definition

A set $S \subset \mathbb{R}^n$ is a **cone** if $ax \in S$ for all $a > 0$ and $x \in S$.
(sometimes $a \geqslant 0$ is required)

Cones need not be convex.





Cones and conic hulls

A **conic combination** of points $x_1, \ldots, x_m \in \mathbb{R}^n$ is the sum,

$$\sum_{j=1}^{m} \mu_j x_j,$$

Recall: "convex" $\Longrightarrow \sum_{j=1}^{m} \mu_j = 1$

for some $\mu \in \mathbb{R}^m_+$.

Definition

The **conic hull**, $\mathrm{cone}(S)$, is either of the following equivalent sets:

- The set of all conic combinations of points in S
- The intersection of all convex cones containing S, unioned with 0

Cones and conic hulls

A **conic combination** of points $x_1, \ldots, x_m \in \mathbb{R}^n$ is the sum,

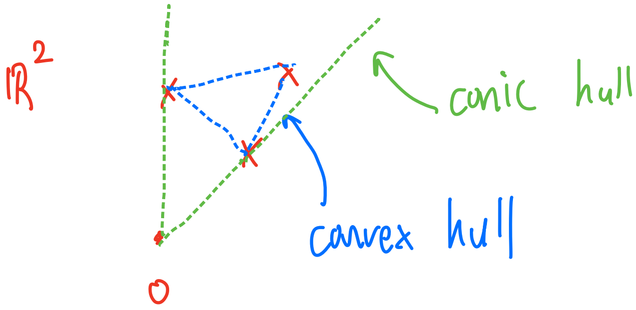
$$\sum_{j=1}^{m} \mu_j x_j,$$

for some $\mu \in \mathbb{R}_+^m$.

Definition

The **conic hull**, $\mathrm{cone}(S)$, is either of the following equivalent sets:

- The set of all conic combinations of points in S
- The intersection of all convex cones containing S, unioned with 0



Convex functions

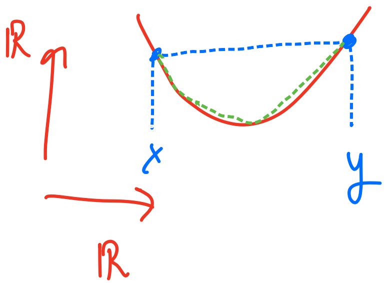
Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex over a convex set S if for all $x, y \in S$ and $\lambda \in [0,1]$, then

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

f is **concave** if $-f$ is convex.

"The graph of f over any line segment in S lies below a secant line connecting the graph values at the endpoints."



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

Convex functions

Various flavors of convexity:

- *Strictly* convex functions have a strict inequality above for all $x \neq y$ and $\lambda \in (0,1)$.
- *Strongly* convex functions are "at least as convex" as a definite quadratic function:

$$f(\lambda x + (1-\lambda) y) \leqslant \lambda f(x) + (1-\lambda) f(y) - m\lambda(1-\lambda)||x-y||_2^2, \quad \overset{\exists}{m} > 0$$

$\forall \lambda \in [0,1]$

There's also quasiconvexity, pseudoconvexity, ...

Strictly convex

$f(x) = x^4$: not strongly convex

convex, but not strictly

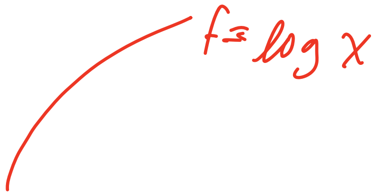
Examples: Convex functions

The following are examples of convex functions:

- $x \mapsto \|x\|$ for any valid norm $\|\cdot\|$
- $f(x) = x^2$, $f(x) = |x|$

- $f(x) = a^T x + b$ (affine functions)
- $f(x) = \log\left(\sum_{j=1}^{n} \exp(x_j)\right)$ ("LogSumExp", whose gradient is the softmax function)

Examples: Convex functions



The following are examples of non-convex functions:

- $f(x) = \log x$ (concave)
- $f(x) = x_1^2 - x_2^2$ (saddle pt @ $x = (0,0)$)
- $f(x) = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}$ for $0 < p < 1$ (ℓ^p "quasinorm")
- $f(A) = \det A$ for $n \times n$ matrices A

Convexity-preserving operations

Complicated convex functions can be build from simple ones using some properties.

- Conic combinations of convex functions are convex
- The (multivariate) maximum of convex function is convex: $\sup_{a \in \mathcal{A}} f_a(x)$ is convex if f_a is convex for each a.
- Compositions with affine maps: If f is convex and $\phi(x) = Ax + b$, then $f(\phi(x))$ is convex. (Is $\phi(f(x))$ convex for ϕ affine and f convex?)
- Compositions of certain convex functions: If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are both convex, and g is non-decreasing, then $g(f(x))$ is convex.

Convexity-preserving operations

Complicated convex functions can be build from simple ones using some properties.

- Conic combinations of convex functions are convex
- The (multivariate) maximum of convex function is convex: $\sup_{a \in \mathcal{A}} f_a(x)$ is convex if f_a is convex for each a. ($\mathcal{A}$ can be discrete or a continuum)
- Compositions with affine maps: If f is convex and $\phi(x) = Ax + b$, then $f(\phi(x))$ is convex. (Is $\phi(f(x))$ convex for ϕ affine and f convex?)
- Compositions of certain convex functions: If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are both convex, and g is non-decreasing, then $g(f(x))$ is convex.

Convexity-preserving operations

Complicated convex functions can be build from simple ones using some properties.

- Conic combinations of convex functions are convex
- The (multivariate) maximum of convex function is convex: $\sup_{a \in \mathcal{A}} f_a(x)$ is convex if f_a is convex for each a.
- Compositions with affine maps: If f is convex and $\phi(x) = Ax + b$, then $f(\phi(x))$ is convex. (Is $\phi(f(x))$ convex for ϕ affine and f convex?)
- Compositions of certain convex functions: If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are both convex, and g is non-decreasing, then $g(f(x))$ is convex.

$\phi(x) = -x$

Convexity-preserving operations

Complicated convex functions can be build from simple ones using some properties.

- Conic combinations of convex functions are convex
- The (multivariate) maximum of convex function is convex: $\sup_{a \in \mathcal{A}} f_a(x)$ is convex if f_a is convex for each a.
- Compositions with affine maps: If f is convex and $\phi(x) = Ax + b$, then $f(\phi(x))$ is convex. (Is $\phi(f(x))$ convex for ϕ affine and f convex?)
- Compositions of certain convex functions: If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are both convex, and g is non-decreasing, then $g(f(x))$ is convex.

Equivalent characterizations of convexity

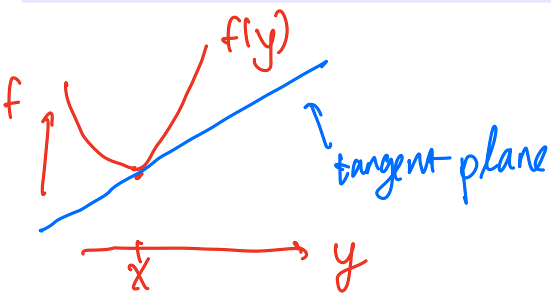
Sometimes alternative formulations of convexity are easier/simpler to show.

Theorem

Assume $f \in C^2(S)$ with S an open set. Then the following are equivalent:

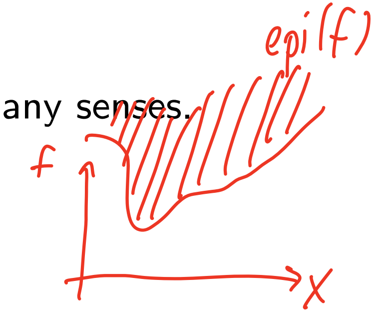
- *f is convex over S.*
- *(Gradient inequality) $f(y) \geqslant \underline{f(x) + \nabla f(x)^T (y - x)}$ for all $x, y \in S$*
- *(Definite Hessian) $\nabla^2 f(x) \geq 0$ for all $x \in S$*

tangent plane to graph at x



Cool things about convex functions

Like convex sets, convex functions are "nice" sets in many senses.



Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex:

- It's epigraph, i.e., the set of points,

$$\mathrm{epi}(f) := \left\{ z = (x, y) \in \mathbb{R}^{n+1} \;\middle|\; y \geqslant f(x) \right\},$$

  is a convex set. (Convexity of $\mathrm{epi}(f)$ also implies convexity of f.)

- Sublevel sets of f, i.e., sets of the form,

$$f^{-1}\left((-\infty, a]\right) = \left\{ x \in \mathbb{R}^n \;\middle|\; f(x) \leqslant a \right\}, \quad f^{-1}\left((-\infty, a)\right) = \left\{ x \in \mathbb{R}^n \;\middle|\; f(x) < a \right\},$$

  for arbitrary $a \in \mathbb{R}$ are convex.
  (Functions whose sublevel sets are convex are *quasiconvex*, but need not be convex.)

Cool things about convex functions

Like convex sets, convex functions are "nice" sets in many senses.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex:

- It's epigraph, i.e., the set of points,

$$\operatorname{epi}(f) := \left\{ z = (x, y) \in \mathbb{R}^{n+1} \mid y \geqslant f(x) \right\},$$

  is a convex set. (Convexity of $\operatorname{epi}(f)$ also implies convexity of f.)

- Sublevel sets of f, i.e., sets of the form,

$$f^{-1}\left((-\infty, a]\right) = \left\{ x \in \mathbb{R}^n \mid f(x) \leqslant a \right\}, \quad f^{-1}\left((-\infty, a)\right) = \left\{ x \in \mathbb{R}^n \mid f(x) < a \right\},$$

  for arbitrary $a \in \mathbb{R}$ are convex.
  (Functions whose sublevel sets are convex are *quasiconvex*, but need not be convex.)

$f(x) = \sqrt{x}$ is quasi convex on $[0, \infty)$, not convex.

Cool things about convex functions

Like convex sets, convex functions are "nice" sets in many senses.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex:
Jensen's inequality is a convex function property for general convex combinations.

Theorem (Jensen's inequality)

If f is a convex function, then for any $x_1, \ldots, x_m$ and $\lambda \in \Delta^m$, we have,

$$f\left(\sum_{j=1}^{m} \lambda_j x_j\right) \leqslant \sum_{j=1}^{m} \lambda_j f(x_j).$$

Jensen's inequality also applies to "infinite sums", i.e., integrals,

$$f\left(\int g(x) \mathrm{d}x\right) \leqslant \int f(g(x)) \mathrm{d}x,$$

where f is convex and g is integrable.

Cool things about convex functions

Like convex sets, convex functions are "nice" sets in many senses.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex:
Jensen's inequality is a convex function property for general convex combinations.

Theorem (Jensen's inequality)

If f is a convex function, then for any $x_1, \ldots, x_m$ and $\lambda \in \Delta^m$, we have,

$$f\left(\sum_{j=1}^{m} \lambda_j x_j\right) \leqslant \sum_{j=1}^{m} \lambda_j f(x_j).$$

Jensen's inequality also applies to "infinite sums", i.e., integrals,

$$f\left(\int g(x)\mathrm{d}x\right) \leqslant \int f(g(x))\mathrm{d}x,$$

where f is convex and g is integrable.

Cool things about convex functions

Like convex sets, convex functions are "nice" sets in many senses.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex: A particularly salient result about convexity:

Theorem

Assume f is convex over (a convex set) S. If x^* is a local minimum of f over S, then it is a global minimum of f over S.

I.e., local and global minimizers are **equivalent** notions for convex functions.

Back to optimization!

General optimization problem:

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \\ & S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\} \end{aligned}$$

Definition

The above is a **convex optimization** problem if f is convex over the convex set S.

Convexity results in strong statements about optimality.

(Also want g_i to be convex for all i)

Convex optimization

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \\ & S := \left\{ x \mid g_i(x) \leqslant 0,\ i \in [m] \right\} \end{aligned}$$

Theorem (Convexity of optimal sets)

Assume the above is a convex optimization problem. Then the set of minimizers,

$$\arg\min_S f = \left\{ x \in S \;\middle|\; f(x) \leqslant f(y)\ \forall\ y \in S \right\}$$

is a convex set.

Convex optimization

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x \in S \end{aligned}$$

$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

Theorem (Unconstrained optimization)

Assume the above is a convex optimization problem with $S = \mathbb{R}^n$ and $f \in C^1(\mathbb{R}^n)$. Then x^* is a stationary point of f if and only if x^* is a global minimizer of f.

$$\nabla f(x) = 0 \iff x \in \operatorname*{argmin}_{x \in S} f(x)$$

Proof: $\nabla f(x^*) = 0$

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

Convex optimization

$$
\begin{aligned}
\text{minimize } & f(x) \\
\text{subject to } & x \in S
\end{aligned}
$$

$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

Theorem (Constrained optimization, sufficiency)

Assume the above is a convex optimization problem and $f, g_i \in C^1(\mathbb{R}^n)$. Then if x^* is a KKT point, it is a global minimizer.

Note: the above assumes no constraint qualifications.

Proof: (x^*, λ^*) : KKT

$$L(x,\lambda) = f(x) + \sum_{i\in[m]} \lambda_i g_i(x)$$

("Lagrangian")

$\lambda_i^* g_i(x^*) = 0$
$\forall\, i \in [m]$

$$f(x^*) = L(x^*, \lambda^*)$$

$$= f(x^*) + \sum_{i\in[m]} \lambda_i^* g_i(x^*)$$

$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*)$
$= 0$

$$= L(x^*, \lambda^*)$$

$$\leq L(x, \lambda^*) \quad \forall\, x \in S$$

$$= f(x) + \sum_{i\in[m]} \lambda_i^* g_i(x) \leq f(x)$$

$g_i(x) \leq 0$, $\lambda_i^* \geq 0$

Convex optimization

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \{x \mid g_i(x) \leqslant 0, \; i \in [m]\}$$

We need some additional assumption for necessity. We say that the problem above satisfies **Slater's condition** if $\exists \; x \in S$ such that $g_i(x) < 0$ for all $i \in [m]$.

Theorem (Constrained optimization, necessity)

Assume the above is a convex optimization problem satisfying Slater's condition, and that $f, g_i \in C^1(S)$*. Then* x^* *is a KKT point if and only if it is a global optimizer.*

Note: Slater's condition doesn't require *a priori* knowledge of KKT points.

Summary

For optimization problems, convexity results in,

- Convexity of optimal sets
- Sufficiency (and typically necessity) of local optimality conditions

References I

Amir Beck, *Introduction to Nonlinear Optimization*, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, October 2014.

Stephen Boyd and Lieven Vandenberghe, *Convex Optimization, With Corrections 2008*, 1 edition ed., Cambridge University Press, Cambridge, UK ; New York, March 2004 (English).

Ralph Tyrell Rockafellar, *Convex Analysis:*, Princeton University Press, Princeton, NJ, December 1996 (English).