ORF 523	Lecture 4	Princeton University
Instructor: A.A. Ahmadi		Scribe: G. Hall

Any typos should be emailed to a_a_a@princeton.edu.

Consider the general form of an optimization problem:

$$\begin{aligned} &\text{min. } f(x) \\ &\text{s.t. } x \in \Omega. \end{aligned}$$

The few optimality conditions we've seen so far characterize *locally* optimal solutions. (In fact, they do not even do that since we did not have a "necessary and sufficient" condition). But ideally, we would like to make statements about *global* solutions. This comes at the expense of imposing some additional structure on f and Ω. By and large, the most successful structural property that we know of that achieves this goal is *convexity*. This motivates an in-depth study of convex sets and convex functions. In short, the reasons for focusing on convex optimization problems are as follows:

- They are close to being the broadest class of problems we know how to solve efficiently.
- They enjoy nice geometric properties (e.g., local minima are global).
- There's excellent software that readily solves a large class of convex problems.
- Numerous important problems in a variety of application domains are convex!

1 From local to global minima

1.1 Definition

Definition 1. *A set $\Omega \subseteq \mathbb{R}^n$ is* convex*, if for all $x, y \in \Omega$ and $\forall \lambda \in [0,1]$*

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

A point of the form $\lambda x + (1-\lambda)y,\ \lambda \in [0,1]$ is called a *convex combination* of x and y. As λ varies between $[0,1]$, a "line segment" is being formed between x and y as shown in Figure 1.

Definition 2. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain dom(f) is a convex set and for all $x, y \in$ dom(f) and $\forall \lambda \in [0,1]$, we have*

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

Figure 1: Convex sets and convex combinations.

Geometrically, the line segment connecting $(x, f(x))$ to $(y, f(y))$ should sit above the graph of the function.

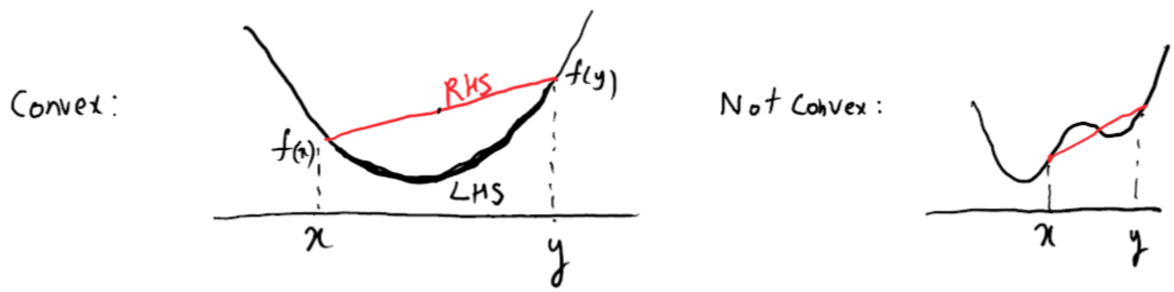


Figure 2: An illustration of the definition of a convex function.

Theorem 1. *Consider an optimization problem*

$$\begin{aligned} &\textit{min. } f(x) \\ &\textit{s.t. } x \in \Omega, \end{aligned}$$

where f is a convex function and Ω is a convex set. Then, any local minimum is also a global minimum.

Proof: Let $\bar{x}$ be a local minimum.

$$\Rightarrow \bar{x} \in \Omega \text{ and } \exists \epsilon > 0 \text{ s.t. } f(\bar{x}) \leq f(x)\ \forall x \in B(\bar{x}, \epsilon).$$

Suppose for the sake of contradiction that $\exists z \in \Omega$ with

$$f(z) < f(\bar{x}).$$

Because of convexity of Ω, we have

$$\lambda \bar{x} + (1 - \lambda) z \in \Omega,\ \forall \lambda \in [0, 1].$$

By convexity of f, we have

$$\begin{aligned}
f(\lambda\bar{x} + (1-\lambda)z) &\le \lambda f(\bar{x}) + (1-\lambda)f(z) \\
&< \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x}).
\end{aligned}$$

But, as $\lambda \to 1$, $(\lambda\bar{x} + (1-\lambda)z) \to \bar{x}$ and the previous inequality contradicts local optimality of $\bar{x}$. $\square$

This theorem, as simple as it is, is one of the most important theorems in convex analysis. Let's now delve deeper in the theory of convex sets and convex functions.

2 Convex sets

If you refer back to the definition of a convex set, you see that the condition is required $\forall x, y \in \Omega$ and $\forall \lambda \in [0,1]$. Under mild conditions, it is possible to fix λ to a constant.

2.1 Midpoint convexity

Definition 3. *A set $\Omega \subseteq \mathbb{R}^n$ is* midpoint convex *if for all $x, y \in \Omega$, the midpoint between x and y is also in Ω. In other words,*

$$x, y \in \Omega \Rightarrow \frac{1}{2}x + \frac{1}{2}y \in \Omega.$$

It's clear that convex sets are midpoint convex. But the converse is also true except in pathological cases.

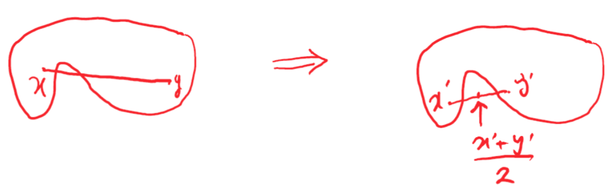


Figure 3: Intuition

Theorem 2. *A closed midpoint convex set Ω is convex.*

<u>Proof:</u> Pick $x, y \in \Omega$, $\lambda \in [0,1]$. For any integer k, define λ_k to be the k-bit rational number closest to λ:

$$\lambda_k = c_1 2^{-1} + c_2 2^{-2} + \ldots + c_k 2^{-k},$$

where $c_i \in \{0,1\}$. Then for all k, $\lambda_k x + (1-\lambda_k)y \in \Omega$ as we can apply midpoint convexity k times recursively.
As $k \to \infty$, $\lambda_k \to \lambda$. By closedness of Ω, we conclude that $\lambda x + (1-\lambda)y \in \Omega$. $\Box$

Remark: An example of a midpoint convex set that's not convex is the set of all rationals in $[0,1]$.

2.2 Common examples of convex sets

The following sets are convex (prove convexity in each case):

- **Hyperplanes:**$\{x \mid a^T x = b\}$ $(a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0)$
- **Halfspaces:** $\{x \mid a^T x \leq b\}$ $(a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0)$
- **Euclidian balls:** $\{x \mid ||x - x_c|| \leq r\}$ $(x_c \in \mathbb{R}^n, r \in \mathbb{R}, ||.||$ is the 2-norm)
- **Ellipsoids:** $\{x \mid \sqrt{(x-x_c)^T P (x - x_c)} \leq r\}$ $(x_c \in \mathbb{R}^n, r \in \mathbb{R}, P \succ 0)$
  Proof: Define $||x||_P = \sqrt{x^T P x}$. This is a norm (as you proved on the homework!). As this set is closed, it suffices to show midpoint convexity. Pick $x, y \in S$. We have $||x - x_c||_P \leq r$ and $||y - x_c||_P \leq r$. Then

$$\begin{aligned} \left|\left|\frac{x+y}{2} - x_c\right|\right|_P &= \left|\left|\frac{1}{2}x - \frac{1}{2}x_c + \frac{1}{2}y - \frac{1}{2}x_c\right|\right|_P \\ &\leq \frac{1}{2}||x - x_c||_P + \frac{1}{2}||y - x_c||_P \\ &\leq \frac{1}{2}r + \frac{1}{2}r = r. \end{aligned}$$

- **The set of symmetric positive semidefinite matrices**:

$$S_+^{n\times n} = \{P \in S^{n \times n} \mid P \succeq 0\}.$$

  Proof: Let $A \succeq 0$, $B \succeq 0$ and $\lambda \in [0,1]$. (Again, it would be enough to look at $\lambda = \frac{1}{2}$ as the set is closed.) Pick any $y \in \mathbb{R}^n$. Then,

$$y^T\left(\lambda A + (1-\lambda)B\right)y = \lambda y^T A y + (1-\lambda) y^T B y \geq 0.$$

- **The set of nonnegative polynomials in** n **variables and of degree** d. (A polynomial p is nonnegative, if $p(x) \geq 0, \forall x \in \mathbb{R}^n$.)

- **The set of optimal solutions of the problem** $\min_{x\in\Omega} f(x)$ where f is convex and Ω is a convex set.

- Proving convexity of a set is not always easy like our previous examples. For instance, for $n > 2$, fix any $Q_1, Q_2 \in S^{n\times n}$: try to show that the following set in $\mathbb{R}^2$ is convex

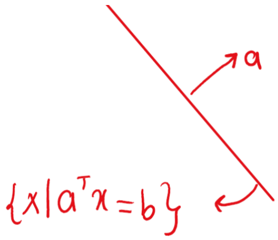
$$S = \{(x^T Q_1 x, x^T Q_2 x) \mid ||x|| = 1\}.$$

  In Figure 4g, you can see an example of what the set S can look like in the case of two quadratics in four variables.

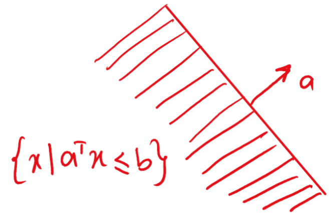
- Interestingly, the analogue of the statement above would fail to be true if we had three quadratic maps or for two polynomial maps of higher degree.

The following theorem (which we will prove later) shows that indeed testing convexity of sets can be a very computationally demanding task.

Theorem 3 ([1])**.** *Given a multivariate polynomial p of degree 4, it is NP-hard to test whether the set $\{x|\ p(x) \leq 1\}$ is convex.*



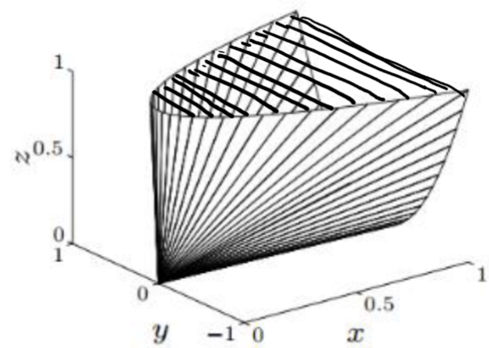
(a) A hyperplane



(b) A halfspace

(c) A Euclidian ball

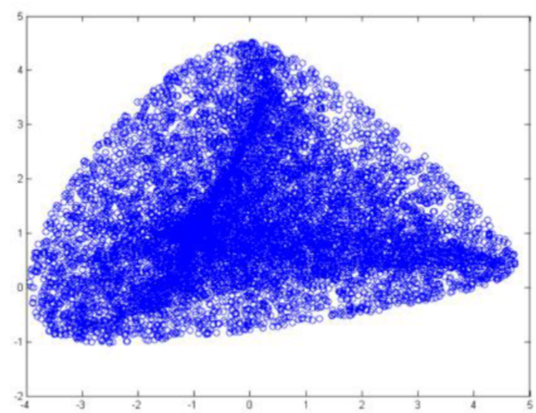
(d) An ellipsoid



(e) $\left\{ (x, y, z) \mid \begin{pmatrix} x & y \\ y & z \end{pmatrix} \succeq 0 \right\}$
(Image credit: [3])

```
n=4;
a=randn(n);a=a+a';
b=rand(n);b=b+b';
for i=1:10000
    x=randn(n,1);x=x/norm(x);
    plot(x.'*a*x,x.'*b*x,'o')
    hold on
end
```

(f) Matlab code to generate S



(g) The set S generated by this Matlab code

Figure 4: Examples of convex sets

2.3 Operations on convex sets

It is very easy to see that the intersection of two convex sets is convex:

$$\Omega_1 \text{ convex}, \Omega_2 \text{ convex} \Rightarrow \Omega_1 \cap \Omega_2 \text{ convex}.$$

Proof: Pick $x \in \Omega_1 \cap \Omega_2$, $y \in \Omega_1 \cap \Omega_2$. We have, for all $\lambda \in [0,1]$:

$$\lambda x + (1-\lambda)y \in \Omega_1 \text{ because } \Omega_1 \text{ is convex}, \quad \lambda x + (1-\lambda)y \in \Omega_2 \text{ because } \Omega_2 \text{ is convex}$$

$\Rightarrow \lambda x + (1-\lambda)y \in \Omega_1 \cap \Omega_2$. $\square$

This statement is also true for infinite intersections. Obviously the union of two convex sets may not be convex.

Similarly, it is easy to show that the Minkowski sum of two convex sets is convex.
Example: Polyhedra

- A polyhedron is the solution set of finitely many linear inequalities. These sets are very important sets in optimization theory. as they form the feasible sets of linear programs.
- A polyhedron can be written in the form

$$\{x|\ Ax \leq b\}$$

  where A is an $m \times n$ matrix and b is an $m \times 1$ vector.
- These sets are convex as they are the intersection of halfspaces $a_i^T x \leq b_i$, where a_i^T is the i^{th} row of A.

$$\text{e.g., } A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix}$$

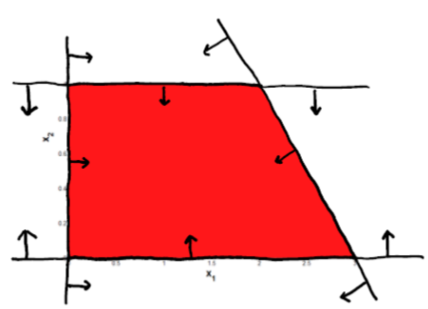


Figure 5: An example of a polyhedron

2.4 Epigraphs

The notion of epigraphs nicely connects the concepts of convex functions and convex sets.

Definition 4. *The epigraph epi(f) of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a subset of $\mathbb{R}^{n+1}$ defined as*

$$epi(f) = \{(x,t) \mid x \in \ domain(f), f(x) \le t\}.$$

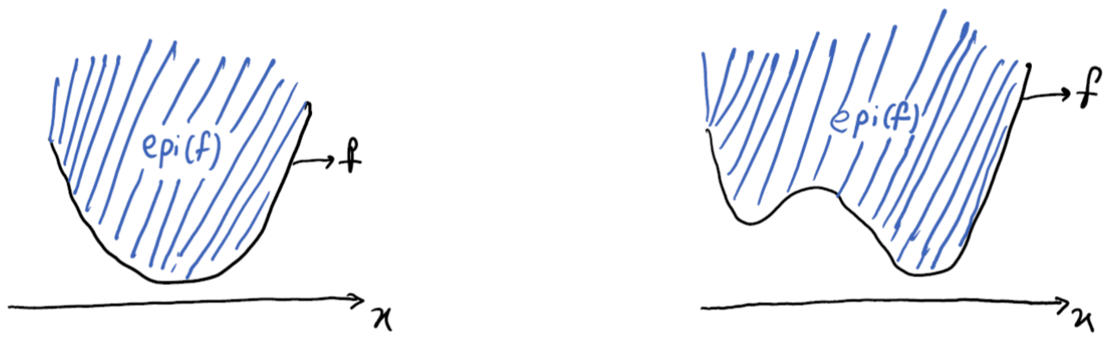


Figure 6: Examples of epigraphs

Theorem 4. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph is convex (as a set).*

Proof:
($\Rightarrow$) Suppose epi(f) not convex $\Rightarrow \exists (x, t_x), (y, t_y), \lambda \in [0,1]$ s.t. $f(x) \le t_x, f(y) \le t_y$ and

$$(\lambda x + (1-\lambda)y, \lambda t_x + (1 - \lambda t_y) \notin epi(f)$$
$$\Rightarrow f(\lambda x + (1-\lambda)y) > \lambda t_x + (1-\lambda)t_y \ge \lambda f(x) + (1-\lambda) f(y).$$

This implies that f is not convex.

($\Leftarrow$) Suppose f is not convex $\Rightarrow \exists x, y \in \text{dom}(f), \lambda \in \ [0,1]$ s.t.

$$f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda) f(y). \tag{1}$$

Pick $(x, f(x)), (y, f(y)) \in \text{epi}(f)$. Then

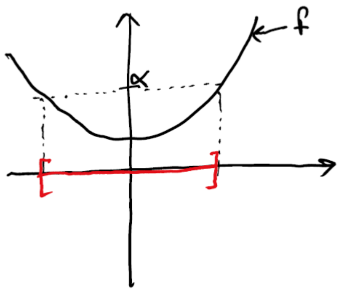
$$(1) \Rightarrow (\lambda x + (1-\lambda)y, \lambda f(x) + (1-\lambda) f(y)) \notin \ \text{epi}(f).$$

$\square$

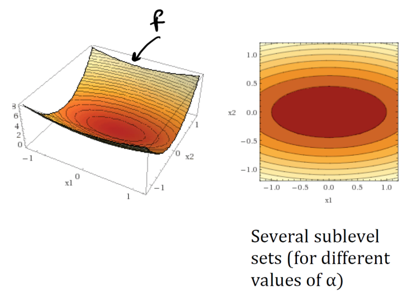
2.5 Convexity of sublevel sets

Definition 5. *The α-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set*

$$S_\alpha = \{x \in domain(f) \mid f(x) \leq \alpha\}.$$



(a) Sublevel set of a function in one dimension



(b) Sublevel sets of a bivariate function

Figure 7: Examples of sublevel sets

Theorem 5. *If a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex then all its sublevel sets are convex sets.*

<u>Proof:</u> Pick $x, y \in S_\alpha$, $\lambda \in [0, 1]$.

$$x \in S_\alpha \Rightarrow f(x) \leq \alpha; \quad y \in S_\alpha \Rightarrow f(y) \leq \alpha$$

$$\begin{aligned} f \text{ convex} \Rightarrow & f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \\ & \leq \lambda\alpha + (1-\lambda)\alpha \\ & = \alpha \end{aligned}$$

$\Rightarrow \lambda x + (1-\lambda)y \in S_\alpha$.

A function whose sublevel sets are all convex is called *quasiconvex*. Although convexity implies quasiconvexity, the converse is not true. See Figure 8.

2.6 Convex hulls

Given $x_1, \ldots, x_m \in \mathbb{R}^n$, a point of the form $\lambda_1 x_1 + \ldots + \lambda_m x_m$ with $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$ is called a *convex combination* of the points $x_1, \ldots, x_m$.

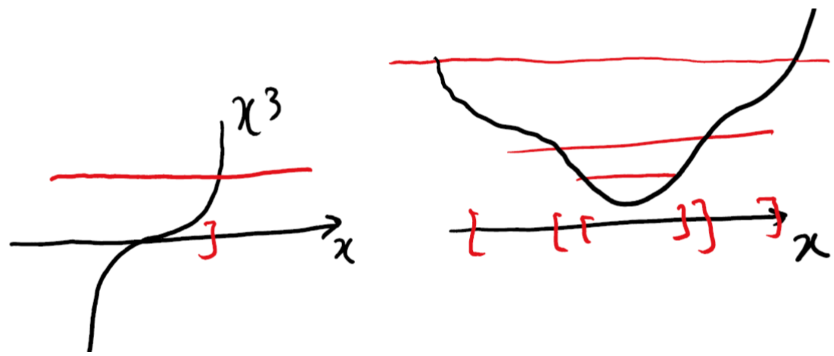


Figure 8: Examples of quasiconvex functions that are not convex

Lemma 1. *A set $S \subseteq \mathbb{R}^n$ is convex iff it contains every convex combination of its points.*

Definition 6. *The* convex hull *of a set $S \subseteq \mathbb{R}^n$, denoted by $conv(S)$, is the set of all convex combinations of the points in S:*

$$conv(S) = \{\sum_{i=1}^{m} \lambda_i x_i \mid x_i \in S, \lambda_i \geq 0, \sum \lambda_i = 1\}.$$

Theorem 6 (Carathéodory, 1907)**.** *Consider a set $S \subseteq \mathbb{R}^d$. Then every point in $conv(S)$ can be written as a convex combinaiton of $d+1$ points in S.*

Proof: We give the standard proof as it appears, e.g., in [2]. Let $x \in conv(S)$, then

$$x = \alpha_1 y_1 + \ldots + \alpha_m y_m,$$

for some $\alpha_i \geq 0, \sum_i \alpha_i = 1$, and $y_i \in S$. If $m \leq d+1$, we are done (why?). Suppose $m > d+1$. We'll give another representation of x using $m-1$ points. An iteration of this idea would finish the proof.
Consider the system of $d+1$ equations:

$$\begin{cases} \gamma_1 y_1 + \ldots + \gamma_m y_m = 0 \\ \gamma_1 + \ldots + \gamma_m = 0 \end{cases}$$

in m variables $\gamma_i \in \mathbb{R}$. As $m > d+1$, this system has infinitely many solutions. Let $\gamma_1, \ldots, \gamma_m$ be any nonzero solution. We must have $\gamma_i > 0$ for some i (why?). Let

$$\tau = \min_i \left\{ \frac{\alpha_i}{\gamma_i} : \gamma_i > 0 \right\} := \frac{\alpha_{i_0}}{\gamma_{i_0}}.$$

Let $\lambda_i = \alpha_i - \tau\gamma_i$ for $i = 1, \ldots, m$.

<u>Claims:</u>

$$(i)\ \lambda_i \geq 0, \quad (ii)\ \sum_{i=1}^{m} \lambda_i = 1, \quad (iii)\ \sum \lambda_i y_i = x_i, \quad (iv)\ \lambda_{i_0} = 0.$$

(i) $\lambda_i = \alpha_i - \frac{\alpha_{i_0}}{\gamma_{i_0}}\gamma_i \geq \alpha_i - \frac{\alpha_i}{\gamma_i}\gamma_i = 0.$

(ii) $\sum \lambda_i = \sum \alpha_i - \tau \sum \gamma_i = \sum \alpha_i = 1.$

(iii) $\sum \lambda_i y_i = \sum(\alpha_i - \tau\gamma_i)y_i = \sum \alpha_i y_i - \tau \sum \gamma_i y_i = x.$

(iv) $\lambda_{i_0} = \alpha_{i_0} - \frac{\alpha_{i_0}}{\gamma_{i_0}}\gamma_{i_0} = 0.$

□

Theorem 7. *The convex hull of S is the smallest convex set that contains S; i.e., if B is convex and $S \subseteq B$, then $conv(S) \subseteq B$.*

The proof is an exercise on the homework. Let us just show that $conv(S)$ is convex.
Pick $x, y \in conv(S)$. This implies:

$$x = \mu_1 u_1 + \ldots + \mu_k u_k,\ u_i \in S, \mu_i \geq 0, \sum \mu_i = 1$$
$$y = \eta_1 v_1 + \ldots + \eta_m v_m,\ v_i \in S, \eta_i \geq 0, \sum \eta_i = 1$$

Let $\lambda \in [0, 1]$.

$$\lambda x + (1-\lambda)y = \lambda\mu_1 u_1 + \ldots + \lambda\mu_k u_k + (1-\lambda)\eta_1 v_1 + \ldots + \eta_m v_m$$

where $u_i, v_i \in S$, $\lambda_i \mu_i \geq 0$, $(1-\lambda_i)\eta_i \geq 0$ and

$$\lambda \sum \mu_i + (1-\lambda)\sum \eta_i = \lambda + (1-\lambda) = 1.$$

□

Theorem 8. *Let $l : \mathbb{R}^n \to \mathbb{R}$ be a linear function and $\Omega \subset \mathbb{R}^n$ a compact set. Then,*

$$\min_{x\in\Omega} l(x) = \min_{x \in conv(\Omega)} l(x).$$

(You can remove the compactness assumption and replace "min" with "inf".)

<u>Proof:</u> It is clear that RHS $\leq$ LHS as we are optimizing over a larger set ($S \subseteq conv(S)$). To show that LHS $\leq$ RHS, let

$$\bar{x} \in \text{argmin}_{x \in conv(\Omega)} l(x).$$

Then,

$$\bar{x} = \sum_{i=1}^{k} \lambda_i y_i, \text{ with } y_i \in \Omega, \sum \lambda_i = 1, \lambda_i \geq 0.$$

$$\begin{aligned} RHS = l(\bar{x}) = l(\sum \lambda_i y_i) &= \sum \lambda_i l(y_i) \\ &\geq \sum \lambda_i \min_i l(y_i) \\ &= \min_i l(y_i) \\ &\geq LHS. \end{aligned}$$

where we used for the last inequality that $y_i \in S$. $\square$.

Consider a general optimization problem

$$\begin{aligned} &\min_x f(x) \\ &\text{s.t. } x \in \Omega. \end{aligned}$$

We can first rewrite it in the so-called "epigraph form" :

$$\begin{aligned} &\min_{x,\alpha} \alpha \\ &\text{s.t. } x \in \Omega, f(x) \leq \alpha. \end{aligned}$$

These problems are equivalent in the sense that they achieve the same optimal value and one can map any optimal solution of one to the other. Notice that in the new problem, the objective is linear. We can rewrite it again (via Theorem 8):

$$\begin{aligned} &\min_{x,\alpha} \alpha \\ &\text{s.t. } (x, \alpha) \in conv\{x \in \Omega, f(x) \leq \alpha\}. \end{aligned}$$

Note that the objective is linear and the feasible set is convex! So we rewrote an arbitrary optimization problem as an optimization problem with a convex objective and a convex feasible set. But there is a catch: this transformation is not algorithmic at all. We are hiding all the difficulty in the convex hull operation. In general, it is not easy to write down a description for the convex hull a set, even if the set has a simple description (let's say it's described by quadratic inequalities).
Note that the argument we gave also goes against the common belief about "convex problems being easy". Indeed, the structure and functional description of the feasible set, beyond convexity, cannot be ignored.

3 Convex optimization problems

Motivated in part by our discussion above, we will define a *convex optimization problem* to be any optimization problem of the form

$$\begin{aligned} \text{min.}\ & f(x) \\ \text{s.t.}\ & g_i(x) \leq 0, i = 1, \ldots, m, \\ & h_j(x) = 0, j = 1, \ldots, k, \end{aligned}$$

where $f, g_i : \mathbb{R}^n \to \mathbb{R}$ are convex functions and $h_i : \mathbb{R}^n \to \mathbb{R}$ are affine functions.

Let Ω denote the feasible set, i.e., $\Omega = \{x \in \mathbb{R}^n |\ g_i(x) \leq 0, h_j(x) = 0\}$.

- For a convex optimization problem, the set Ω is always a convex set (why?).
- The converse is not true:
  - Consider for example, $\Omega = \{x \in \mathbb{R} |\ x^3 \leq 0\}$. Then Ω is a convex set, but minimizing a convex function over Ω is not a convex optimization problem per our definition.
  - However, the same set can be represented as $\Omega = \{x \in \mathbb{R}^n |\ x \leq 0\}$, and then this would be a convex optimization problem with our definition.
- We require this stronger notion because otherwise many abstract and complicated optimization problems can be formulated as optimization problems over a convex set. (Think, e.g., of the set of nonnegative polynomials). The stronger definition is much closer to what we can solve algorithmically.

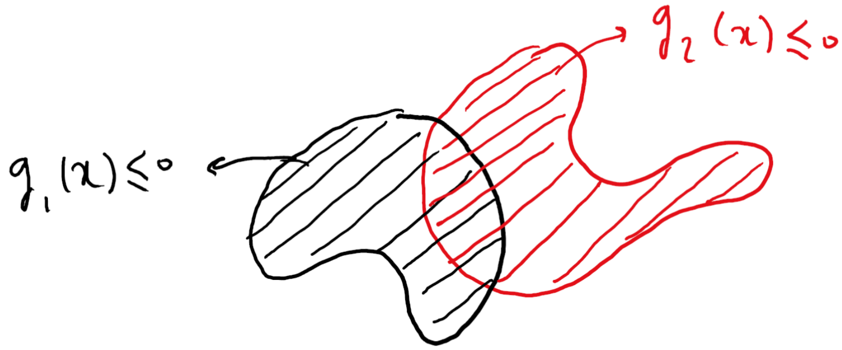


Figure 9: The feasible set $S = \{x \mid g_1(x) \leq 0, g_2(x) \leq 0\}$ can be convex even when the defininig inequalities are not even quasiconvex.

The software CVX that we will be using only accepts convex optimization problems defined as above; i.e, CVX accepts the following constraints:

- Convex ≤ 0.
- Affine $== 0$.

Notes

Further reading for this lecture can include Chapter 2 of [3].

References

[1] Amir Ali Ahmadi, Alex Olshevsky, Pablo A Parrilo, and John N Tsitsiklis. NP-hardness of deciding convexity of quartic polynomials and related problems. *Mathematical Programming*, 137(1-2):453–476, 2013.

[2] Alexander Barvinok. *A Course in Convexity*, volume 54. American Mathematical Soc., 2002.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, http://stanford.edu/ boyd/cvxbook/, 2004.