PRINCETON UNIVERSITY
ORFE

Instructor:
Amir Ali Ahmadi
Spring 2017

TA: Georgina Hall

Today:

Start convex analysis:

- Convex sets and convex functions, local and global minima of convex problems.
- Midpoint convex sets, epigraphs, quasiconvex functions.
- Convex hulls, Carathéodory's theorem, rewriting any problem as a convex problem.

Consider the general form of an optimization problem:

$$\min_{x \in \Omega} \; f(x)$$

The few optimality conditions we've seen so far characterize locally optimal solutions. (In fact, they don't 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 and acheives 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:

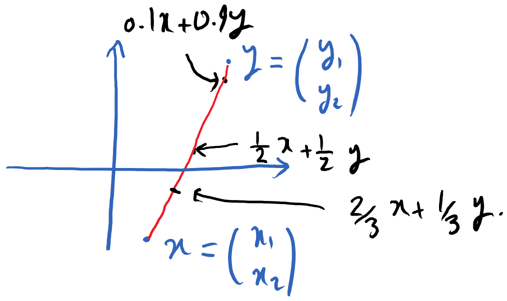
- They are pretty much the broadest class of problems we can solve efficiently.
- They enjoy nice geometric properties (e.g., local minima are global).
- There's excellent software that readily solves (a large subset of) convex problems.
- Numerous important problems in various application domains are convex!

From local to global minima

Defn. A set $\Omega \subseteq \mathbb{R}^n$ is <u>convex</u>, if $\forall x, y \in \Omega$ and $\forall \lambda \in [0,1]$

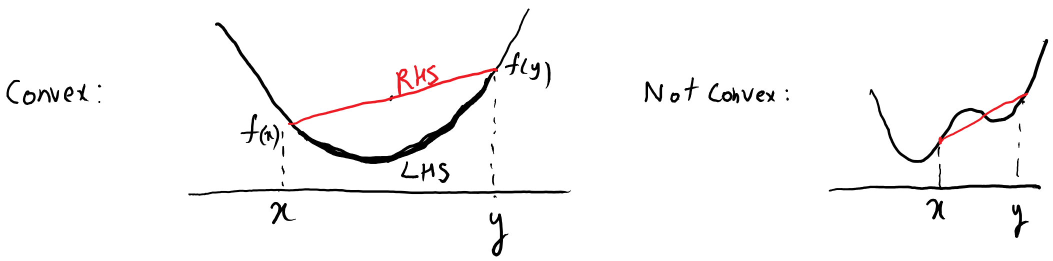
$$\lambda x + (1-\lambda) y \in \Omega .$$

- A point of the form $\lambda x + (1-\lambda) y$ is called a <u>convex combination</u> of x and y. As λ varies between $[0,1]$, a "line segment" is being formed between x and y.



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

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



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

Theorem. Consider an optimization problem

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

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.

$\Rightarrow \bar{x} \in \Omega$ and $\exists\ \epsilon > 0$ 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

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

But because of convexity of Ω we have

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

By convexity of f we have

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

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 optimization. Let's now delve deeper in convex sets and convex functions to see how far we can get.

Convex sets

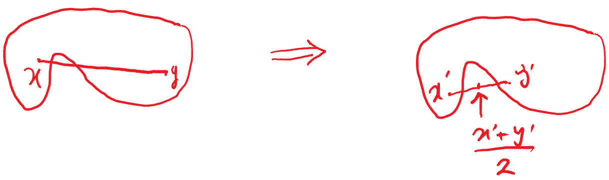
The definition of convexity requires a condition $\forall x, y \in \Omega$ and $\forall \lambda \in [0,1]$. Under mild conditions, it is possible to fix λ.

<u>Defn.</u> A set $\Omega \subseteq \mathbb{R}^n$ is <u>midpoint convex</u> if $\forall x, y \in \Omega$,

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

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

Intuition:

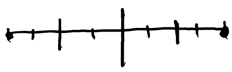


<u>Theorem.</u> A closed midpoint convex set Ω is convex.

<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:

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

where $c_i \in \{0, 1\}$. Then $\forall k$, $\lambda_k x + (1-\lambda_k) y \in \Omega$ as



we can apply midpoint convexity k times recursively.

As $k \to \infty$, $\lambda_k \to \lambda$. By closedness of Ω, we conclude that $\lambda x + (1-\lambda) y \in \Omega$. $\square$

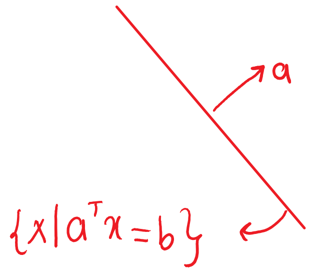
- An example of a midpoint convex set that's not convex:

$$\Omega = \{\text{all rationals in } [0,1]\}.$$

Common convex sets in optimization

(Prove convexity in each case.)

- **Hyperplanes:** $\{x \mid a^T x = b\}$ $(a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0)$



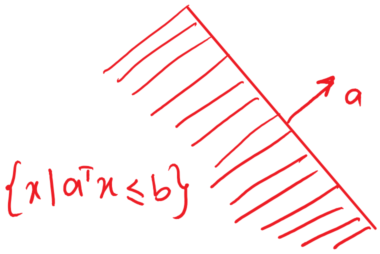
- **Halfspaces:** $\{x \mid a^T x \leq b\}$ $(a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0)$

Proof: Let $\mathcal{H} := \{x \mid a^T x \leq b\}$. Take $x, y \in \mathcal{H}$

$$a^T(\lambda x + (1-\lambda) y) = \lambda a^T x + (1-\lambda) a^T y \leq \lambda b + (1-\lambda) b = b$$

$$\Rightarrow \lambda x + (1-\lambda) y \in \mathcal{H}. \quad \square$$

(Could also just work with $\lambda = \frac{1}{2}$)



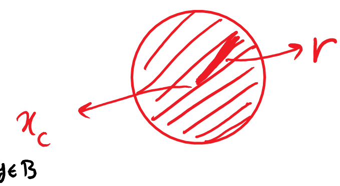
- **Euclidean balls:** $\{x \mid \|x - x_c\| \leq r\}$ $(x_c \in \mathbb{R}^n, r \in \mathbb{R}, \|.\|$ 2-norm$)$

Proof: Let $B := \{x \mid \|x - x_c\| \leq r\}$. Take $x, y \in B$.

$$\|\lambda x + (1-\lambda) y - x_c\| = \|\lambda (x - x_c) + (1-\lambda)(y - x_c)\|$$

$$\underset{\text{Triangle ineq.}}{\leq} \|\lambda (x - x_c)\| + \|(1-\lambda)(y - x_c)\| \overset{\text{homog.}}{=} \lambda \|x - x_c\| + (1-\lambda)\|y - x_c\| \overset{x, y \in B}{\leq} \lambda r + (1-\lambda) r = r. \Rightarrow \lambda x + (1-\lambda) y \in B.$$

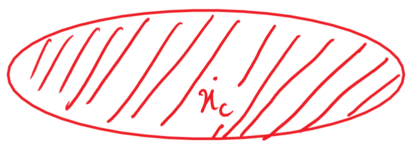
$\square$



- **Ellipsoids:** $\{x \mid \sqrt{(x - x_c)^T P (x - x_c)} \leq r\}$ $(x_c \in \mathbb{R}^n, r \in \mathbb{R}, P \succ 0)$

(P here is an $n \times n$ symmetric matrix)

Proof: Prove that $\|x\|_P = \sqrt{x^T P x}$ is a norm and repeat as above.



Fancier convex sets

Many fundamental objects in mathematics have surprising convexity properties.

For example, prove that the following two sets are convex.

- The set of (symmetric) positive semidefinite matrices:
  $S_+^{n\times n} = \{P \in S^{n\times n} |\ P \succcurlyeq 0\}$

Proof. Let $A \succcurlyeq 0$, $B \succcurlyeq 0$. Let $\lambda \in [0,1]$. $x^T(\lambda A + (1-\lambda)B)x = \lambda \underbrace{x^TAx}_{\geq 0} + (1-\lambda)\underbrace{x^TBx}_{\geq 0} \geq 0.$ $\square$

e.g., $\{x, y, z |\begin{bmatrix} x & y \\ y & z \end{bmatrix} \succcurlyeq 0\}$:

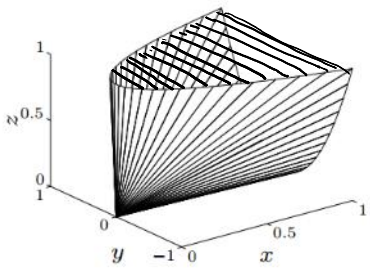
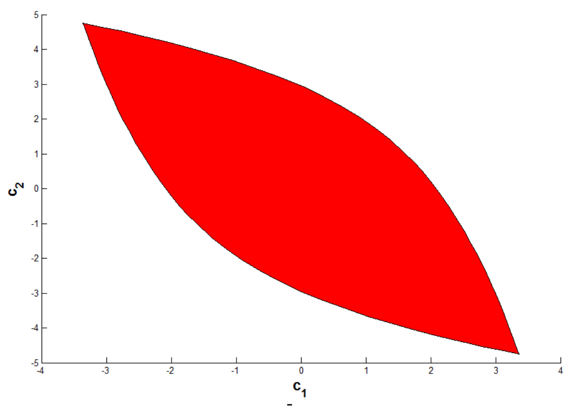


Image credit: [BV04]

- The set of nonnegative polynomials in n variables and of degree d.
  (A polynomial $p(x_1, \ldots, x_n)$ is nonnegative, if $p(x) \geq 0, \forall x \in \mathbb{R}^n$.)

e.g., $\{(c_1, c_2) |\ 2x_1^4 + x_2^4 + c_1 x_1 x_2^3 + c_2 x_1^3 x_2 \geq 0, \forall (x_1, x_2) \in \mathbb{R}^2\}$:

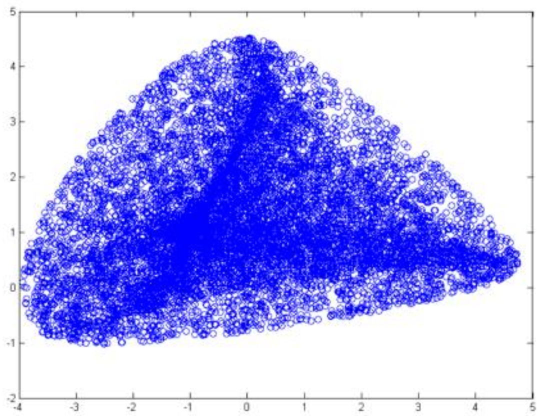


Proving convexity of a set is not always easy like our previous examples.

For example, try to show that the following set in $\mathbb{R}^2$ is convex for any pair of matrices $Q_1, Q_2 \in S^{n\times n}$, $n>2$:

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

```
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
```



- Would the same statement hold if we had three quadratic forms?

$$\tilde{S} = \{ (x^T Q_1 x, x^T Q_2 x, x^T Q_3 x) \mid \|x\|=1 \}$$

- Would the same statement hold if we had two cubic or quartic forms (homogeneous polynomials) p_1, p_2?

$$\hat{S} = \{ (p_1(x), p_2(x)) \mid \|x\|=1 \}$$

Here's a result on complexity of testing convexity that we'll prove later:

Thm [AOPT13]. Given a polynomial p of degree 4, it is NP-hard to test whether the following set is convex: $\{x \mid p(x) \le 1\}$.

- Easy to see that intersection of two convex sets is convex:

  Ω_1 convex, Ω_2 convex $\Rightarrow \Omega_1 \cap \Omega_2$ convex.

Proof:

Pick $x \in \Omega_1 \cap \Omega_2$, $y \in \Omega_1 \cap \Omega_2$

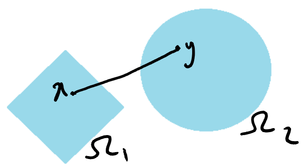
$\forall \lambda \in [0,1]$, $\lambda x + (1-\lambda) y \in \Omega_1$ (b/c Ω_1 is convex)

$\lambda x + (1-\lambda) y \in \Omega_2$ (b/c Ω_2 is convex)

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

True also for infinite intersections. True also for Minkowski sums.

- Obviously, the union may not be convex:



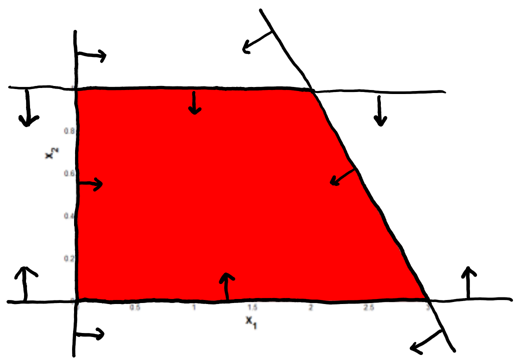
Polyhedra

- A polyhedron is the solution set of finitely many linear inequalities.
  - Ubiquitous in optimization theory.
  - Feasible sets of "linear programs".
- Such sets are written in the form:

$$\{x \mid Ax \leq b\},$$

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

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



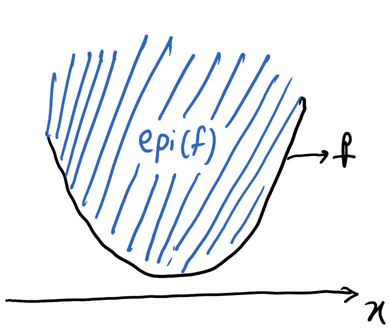
Epigraph

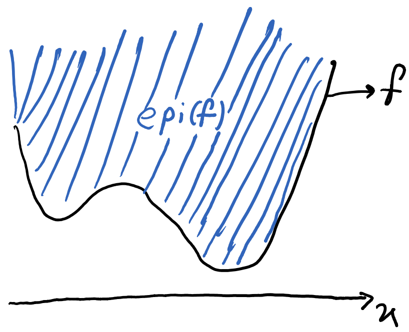
Is there a connection between convex sets and convex functions?

- We will see a couple; via epigraphs, and sublevel sets.

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

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





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

Proof: Suppose f not convex $\Rightarrow \exists\, x, y \in dom(f), \lambda \in [0,1]$

s.t. $f(\lambda x + (1-\lambda) y) > \lambda f(x) + (1-\lambda) f(y)$. ①

Pick $(x, f(x)), (y, f(y)) \in epi(f)$.

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

Suppose $epi(f)$ not convex $\Rightarrow \exists\, (x, t_x), (y, t_y), \lambda \in [0,1]$

s.t. $f(x) \le t_x$, $f(y) \le t_y$, $f(\lambda x + (1-\lambda) y) > \lambda t_x + (1-\lambda) t_y$

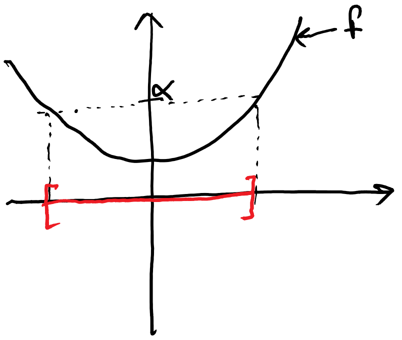
$\ge \lambda f(x) + (1-\lambda) f(y)$

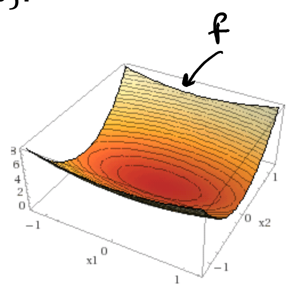
$\Rightarrow f$ not convex. $\square$

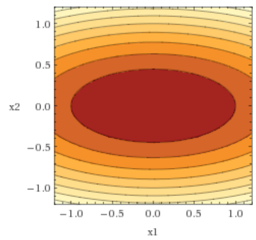
Convexity of sublevel sets

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

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



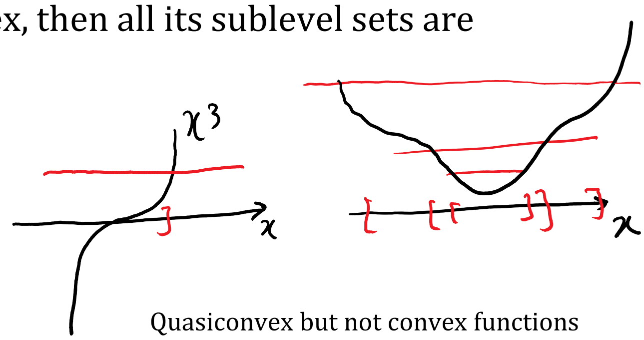




Several sublevel sets (for different values of α)

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

- Converse *not* true.
- A function whose sublevel sets are all convex is called *quasiconvex*.



Quasiconvex but not convex functions

Proof of theorem:

Pick $x, y \in S_\alpha$, $\lambda \in [0,1]$

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

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

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

Convex hulls

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

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

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

$$\mathrm{conv}(S) = \left\{ \sum_{i=1}^{m} \lambda_i x_i \;\middle|\; x_i \in S,\ \lambda_i \geq 0,\ \sum \lambda_i = 1 \right\}.$$

Thm (Carathéodory ~1907). Consider a set $S \subseteq \mathbb{R}^d$. Then every point in $\mathrm{conv}(S)$ can be written as a convex combination of $d+1$ points in S.

Proof: We give the standard proof as, e.g., in [Bar00]. Let $x \in \mathrm{conv}(S)$.

$$\Rightarrow \quad x = \alpha_1 y_1 + \cdots + \alpha_m y_m, \quad \text{where } \sum \alpha_i = 1, \alpha_i \geq 0, \text{ and } y_i \in S.$$

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

in m variables $\gamma_i \in \mathbb{R}$. As $m > d+1$, this system has infinitely many solutions.

Let $\gamma_1, \dots, \gamma_m$ be any nonzero solution. (We must have $\gamma_i > 0$ for some i.)

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 $i = 1, \dots, m$.

__Claims:__ (i) $\lambda_i \geq 0$, (ii) $\sum_{i=1}^m \lambda_i = 1$, (iii) $\sum \lambda_i y_i = x$, (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) $\sum \lambda_i = \sum \alpha_i - \tau \cancel{\sum \gamma_i}^{0} = \sum \alpha_i = 1$.

(iii) $\sum \lambda_i y_i = \sum (\alpha_i - \tau \gamma_i) y_i = \sum \alpha_i y_i - \tau \cancel{\sum \gamma_i y_i}^{0} = x$.

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

__Thm__. Convex hull of S is the smallest convex set that contain S; i.e, if B is convex and $S \subseteq B$, then $\mathrm{Conv}(S) \subseteq B$.

Proof is an exercise on the homework. Let us just show that $\mathrm{conv}(S)$ is convex.

Pick $x, y \in \mathrm{conv}(S) \Rightarrow$ $x = \mu_1 u_1 + \dots + \mu_k u_k$, $u_i \in S$, $\mu_i \geq 0$, $\sum \mu_i = 1$

$y = \eta_1 v_1 + \dots + \eta_m v_m$ $v_i \in S$, $\eta_i \geq 0$, $\sum \eta_i = 1$

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

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

$u_i, v_i \in S$, $\lambda_i \mu_i \geq 0$, $(1-\lambda_i)\eta_i \geq 0$, and $\lambda \sum \mu_i + (1-\lambda) \sum \eta_i = \lambda + (1-\lambda) = 1$. $\square$

Theorem. Let $\ell : \mathbb{R}^n \to \mathbb{R}$ be a linear function, and $\Omega \subseteq \mathbb{R}^n$ be compact. Then,
$$\min_{x \in \Omega} \ell(x) = \min_{x \in \text{conv}\,\Omega} \ell(x).$$

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

Proof. It is clear that RHS $\leq$ LHS.

To show that LHS $\leq$ RHS, let $\bar{x} = \arg\min_{x \in \text{conv}(\Omega)} \ell(x)$.

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

$$\begin{aligned} \text{RHS} = \ell(\bar{x}) = \ell\left(\sum \lambda_i y_i\right) &= \sum \lambda_i \, \ell(y_i) \\ &\geq \sum \lambda_i \min_i \ell(y_i) \\ &= \min_i \ell(y_i) \\ &\geq \text{LHS}. \quad \square \end{aligned}$$

since $y_i \in S$

Consider a general optimization problem:

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

We can rewrite it as a "convex optimization" problem (in abstract form):

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

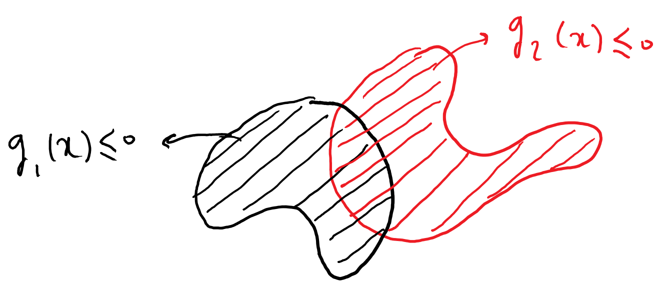
Convex optimization problems

A convex optimization problem is an optimization problem of the form

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

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

- Let Ω denote the feasible set: $\Omega = \{x \in \mathbb{R}^n |\ g_i(x) \le 0, h_j(x) = 0\}$.
  - Observe that for a convex optimization problem Ω is a convex set (why?)
  - But the converse is not true:
    - Consider for example, $\Omega = \{x \in \mathbb{R}|\ x^3 \le 0\}$. 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 $\Omega = \{x \in \mathbb{R}|\ x \le 0\}$, and then this would be a convex optimization problem with our definition.
- Here is another example of a convex feasible set that fails our definition of a convex optimization problem:



$\Omega = \{x \mid g_1(x) \le 0,\ g_2(x) \le 0\}$

is a convex set. But neither

g_1 nor g_2 are convex functions (why?).

Convex optimization problems (cont'd)

- We require this stronger definition because otherwise many abstract and complex 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 actually solve efficiently.
- The software CVX that we'll be using ONLY accepts convex optimization problems defined as above.
- Beware that [CZ13] uses the weaker and more abstract definition for a convex optimization problem (i.e., the definition that simply asks Ω to be a convex set.)

Acceptable constraints in CVX:

- **Convex $\leq$ 0**
- **Affine == 0**

Notes:

- Further reading for this lecture can include Chapter 2 of [BV04].

References:

- [BV04] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.
  http://stanford.edu/~boyd/cvxbook/

- [AOPT13] A.A. Ahmadi, A. Olshevsky, P.A. Parrilo, and J.N. Tsitsiklis. NP-hardness of checking convexity of quartic forms and related problems. *Mathematical Programming,* 2013.
  http://web.mit.edu/~a_a_a/Public/Publications/convexity_nphard.pdf