PRINCETON UNIVERSITY
ORFE
Instructor:
Amir Ali Ahmadi

This lecture:

- Convex optimization
  - Convex sets
  - Convex functions
  - Convex optimization problems
  - Why convex optimization? Why so early in the course?

Recall the general form of our optimization problems:

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

- In the last lecture, we focused on unconstrained optimization: $\Omega = \mathbb{R}^n$.
- We saw the definitions of local and global optimality, as well as first and second order optimality conditions.

- In this lecture, we consider a very important special case of constrained optimization problems known as "*convex optimization problems*".
- For these problems,
  - f will be a "*convex function*".
  - Ω will be a "*convex set*".
  - These notions are defined formally in this lecture.

- Roughly speaking, the high-level message is this:
  - Convex optimization problems are pretty much the broadest class of optimization problems that we know how to solve efficiently.
  - They have nice geometric properties;
    - e.g., a local minimum is automatically a global minimum.
  - Numerous important optimization problems in engineering, operations research, machine learning, etc. are convex.
  - There is available software that can take (a large subset of) convex problems written in very high-level language and solve it.
    - You should take advantage of this!
  - Convex optimization is one of the biggest success stories of modern theory of optimization.

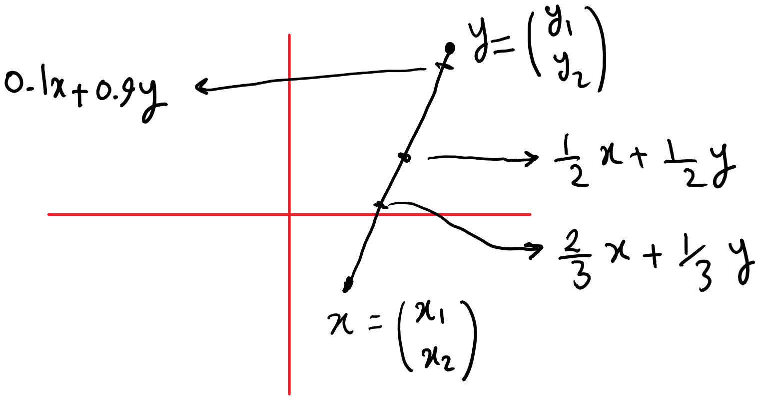
Convex sets

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

$$x, y \in \Omega, \lambda \in [0,1] \Rightarrow \lambda x + (1-\lambda)y \in \Omega$$

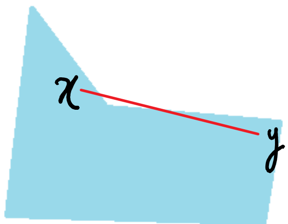
- A point of the form $\lambda x + (1-\lambda)y, \lambda \in [0,1]$ is called a *convex combination* of x and y.
- Note that when $\lambda = 0$, we are at y; when $\lambda = 1$, we are at x; for intermediate values of λ, we are on the line segment connecting x and y.

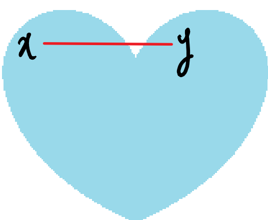
Illustration of the concept of a convex combination:



Convex:

Not convex:





Convex sets & Midpoint Convexity

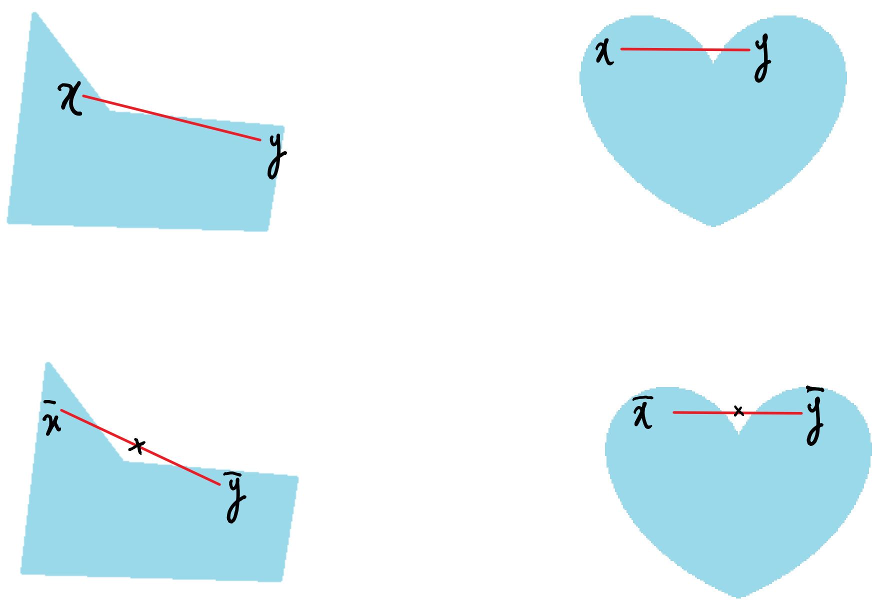
Midpoint convexity is a notion that is equivalent to convexity in most practical settings, but it is a little bit cleaner to work with.

Definition. 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,

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

- Obviously, convex sets are midpoint convex.
- Under mild conditions, midpoint convex sets are convex
  - e.g., a closed midpoint convex sets is convex.
  - What is an example of a midpoint convex set that is not convex? (The set of all rational points in [0,1].)

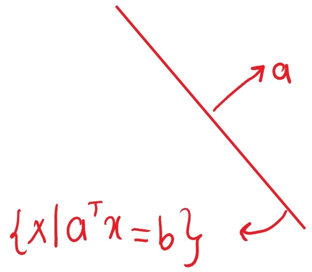
The nonconvex sets that we had are also not midpoint convex (why)?:



Common convex sets in optimization

(Prove convexity in each case.)

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

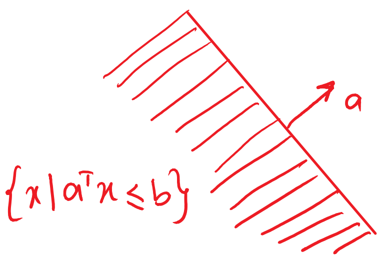


- **Halfspaces:** $\{x \mid a^T x \leq b\}$ $(a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0)$

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

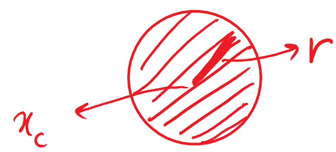
$a^T(\lambda x + (1-\lambda) y) = \lambda a^T x + (1-\lambda) a^T y \le \lambda b + (1-\lambda) b = b$

$\Rightarrow \lambda x + (1-\lambda) y \in H$. □



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

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



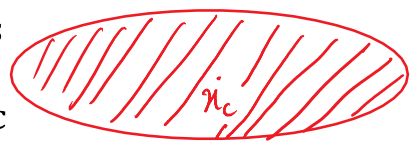
$\|\lambda x + (1-\lambda) y - x_c\| = \|\lambda(x - x_c) + (1-\lambda)(y - x_c)\|$

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

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

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

Proof hint: Wait until you see convex functions and quasiconvex functions. Observe that ellipsoids are sublevel sets of convex quadratic functions.



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

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

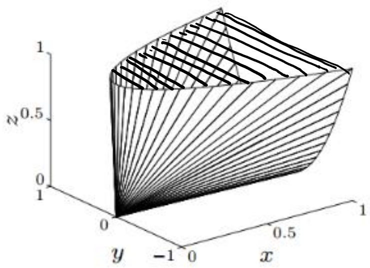
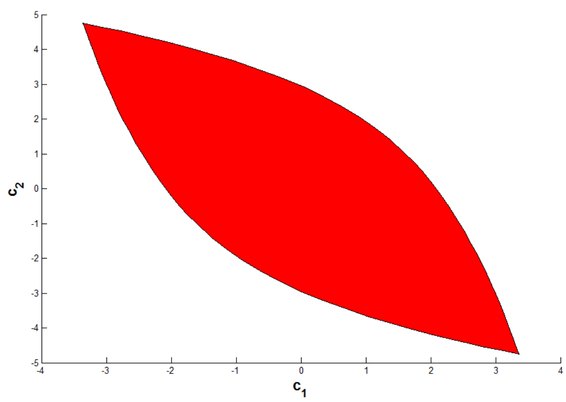


Image credit: [BV04]

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



Intersections of convex sets

- 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:

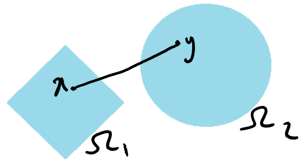
Pick $x \in \Omega_1 \cap \Omega_2$, $y \in \Omega_1 \cap \Omega_2$

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

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

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

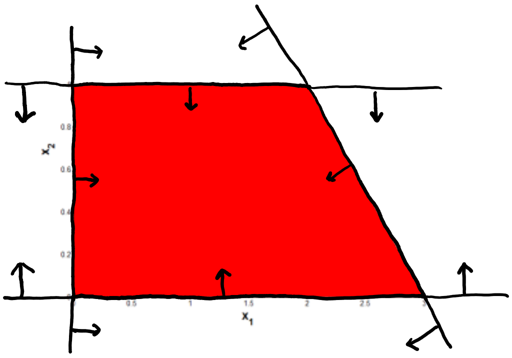
- Obviously, the union may not be convex:



Polyhedra

- A polyhedron is the solution set of finitely many linear inequalities.
  - Ubiquitous in optimization theory.
  - Feasible sets of "linear programs" (an upcoming subject).
- Such sets are 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: intersection of halfspaces $a_i^T x \leq b_i$, where a_i^T is the i-th row of A.

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

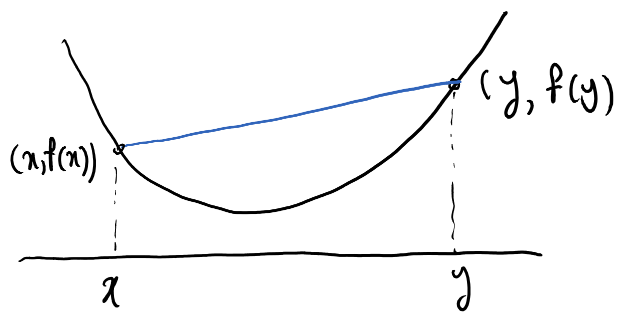


Convex functions

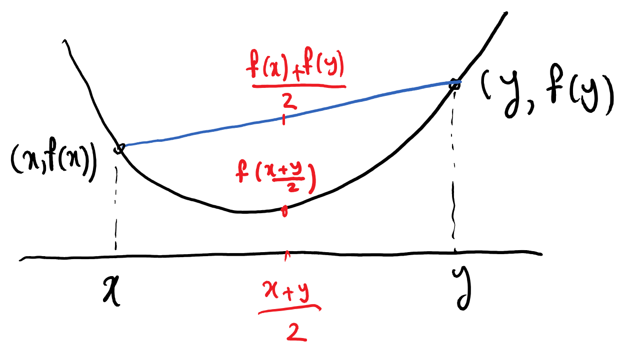
Definition. A function $f: \mathbb{R}^n \to \mathbb{R}$ is *convex* if its domain is a convex set and for all x,y in its domain, and all $\lambda \in [0,1]$, we have

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

- In words: take any two points x, y; f evaluated at any convex combination should be no larger than the same convex combination of $f(x)$ and $f(y)$.
- If $\lambda = \frac{1}{2}$, interpretation is even easier: take any two points x, y; f evaluated at the midpoint should be no larger than the average of $f(x)$ and $f(y)$.
- Geometrically, the line segment connecting $(x, f(x))$ to $(y, f(y))$ sits above the graph of f.



$(f: \mathbb{R} \to \mathbb{R})$



Definition. A function $f: \mathbb{R}^n \to \mathbb{R}$ is

- *Concave*, if $\forall x, y, \forall \lambda \in [0,1]$
  $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y)$.
- *Strictly convex*, if $\forall x, y, x \neq y, \forall \lambda \in (0,1)$
  $$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y).$$
- *Strictly concave*, if $\forall x, y, x \neq y, \forall \lambda \in (0,1)$
  $$f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda) f(y).$$

Note: f is concave if and only if $-f$ is convex. Similarly, f is strictly concave if and only if $-f$ is strictly convex.

The only functions that are both convex and concave are affine functions; i.e., functions of the form:
$f(x) = a^T x + b, \qquad (a \in \mathbb{R}^n, b \in \mathbb{R})$.

convex
(and strictly convex)

concave
(and strictly concave)

neither convex nor concave

both convex and concave (but not strictly)

Let's see some examples of convex functions (selection from [BV04]; see this reference for many more examples).

Examples of univariate convex functions ($f: \mathbb{R} \to \mathbb{R}$):

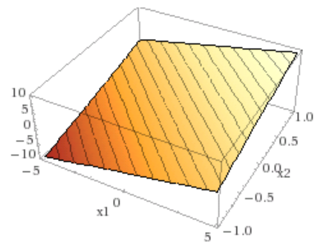
- e^{ax}
- $-\log x$
- x^a (defined on $\mathbb{R}_{++}$) $a \geq 1$ or $a \leq 0$
- $-x^a$ (defined on $\mathbb{R}_{++}$) $0 \leq a \leq 1$
- $|x|^a, a \geq 1$
- $x \log x$ (defined on $\mathbb{R}_{++}$)

- Try to plot the functions above and convince yourself of convexity visually.
- Can you formally verify that these functions are convex?
- We will soon see some characterizations of convex functions that make the task of verifying convexity a bit easier.

Examples of convex functions ($f: \mathbb{R}^n \to \mathbb{R}$)

- **Affine functions:** $f(x) = a^T x + b$ (for any $a \in \mathbb{R}^n, b \in \mathbb{R}$)

  (convex, but not strictly convex; also concave)

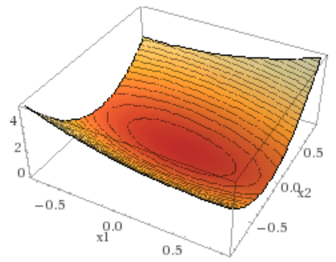


Proof: $\forall \lambda \in [0,1]$, $f(\lambda x + (1-\lambda) y) = a^T(\lambda x + (1-\lambda)y) + b$

$= \lambda a^T x + (1-\lambda) a^T y + \lambda b + (1-\lambda) b = \lambda f(x) + (1-\lambda) f(y)$. $\square$

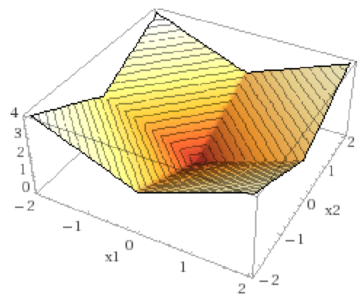
- **Some quadratic functions:**

$$f(x) = x^T Q x + c^T x + d$$



  - Convex if and only if $Q \succcurlyeq 0$.
  - Strictly convex if and only if $Q \succ 0$.
  - Concave iff $Q \preccurlyeq 0$; Strictly concave iff $Q \prec 0$.
  - Proofs are easy from the second order characterization of convexity (coming up).

- **Any norm:** meaning, any function f satisfying:
  a. $f(\alpha x) = |\alpha| f(x), \forall \alpha \in \mathbb{R}$
  b. $f(x+y) \leq f(x) + f(y)$
  c. $(f(x) \geq 0, \forall x, f(x) = 0 \Rightarrow x = 0)$



$||x||_1$

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

Proof: $f(\lambda x + (1-\lambda) y) \overset{b}{\leq} f(\lambda x) + f((1-\lambda) y)$

$\overset{a}{=} \lambda f(x) + (1-\lambda) f(y)$. $\square$

Examples:

- $||x||_\infty = \max_i |x_i|$
- $||x||_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, \quad p \geq 1$
- $||x||_Q = \sqrt{x^T Q x}, Q \succ 0$

Midpoint convex functions

Same idea as what we saw for midpoint convex sets.

Definition. A function $f: \mathbb{R}^n \to \mathbb{R}$ is *midpoint convex* if its domain is a convex set and for all x,y in its domain, we have

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2} f(x) + \frac{1}{2} f(y).$$

- Obviously, convex functions are midpoint convex.
- Continuous, midpoint convex functions are convex.

Convexity = Convexity along all lines

Theorem. A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function $g: \mathbb{R} \to \mathbb{R}$, given by $g(t) = f(x + ty)$ is convex (as a univariate function), for all x in domain of f and all $y \in \mathbb{R}^n$. (The domain of g here is all t for which $x + ty$ is in the domain of f.)

- This should be intuitive geometrically:
  - The notion of convexity is defined based on line segments.
- The theorem simplifies many basic proofs in convex analysis.
- But it does not usually make verification of convexity that much easier; the condition needs to hold for *all* lines (and we have infinitely many).
- Many of the algorithms we will see in future lectures work by iteratively minimizing a function over lines. It's useful to remember that the restriction of a convex function to a line remains convex. Here is a proof:

Suppose for some x, y, $g(\alpha) = f(x + \alpha y)$ was not convex.

$\Rightarrow \exists \lambda \in [0,1], \alpha_1, \alpha_2$ s.t. $g(\lambda \alpha_1 + (1-\lambda)\alpha_2) > \lambda g(\alpha_1) + (1-\lambda) g(\alpha_2)$.

$\Rightarrow f(x + (\lambda \alpha_1 + (1-\lambda)\alpha_2) y) = f(\lambda (x + \alpha_1 y) + (1-\lambda)(x + \alpha_2 y)) > \lambda f(x + \alpha_1 y) + (1-\lambda) f(x + \alpha_2 y)$.

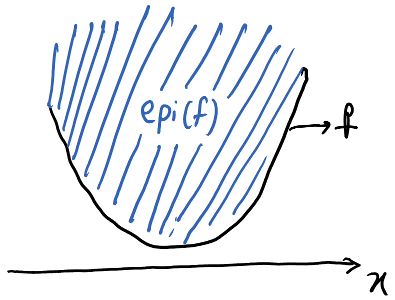
$\square$

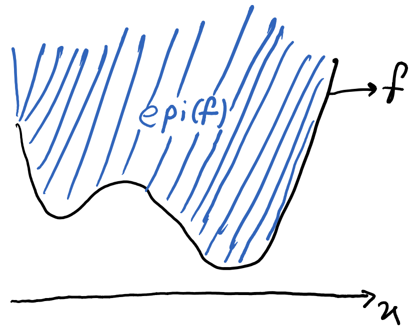
Epigraph

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

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





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

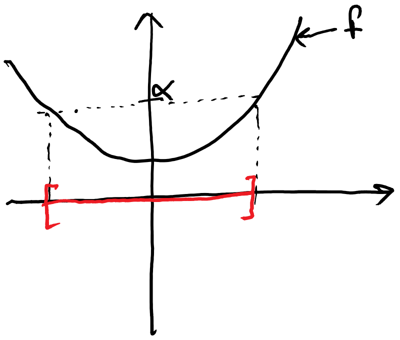
$\geq \lambda f(x) + (1-\lambda) f(y)$

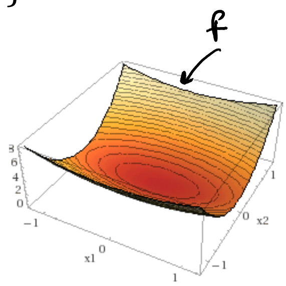
$\Rightarrow f$ not convex. $\square$

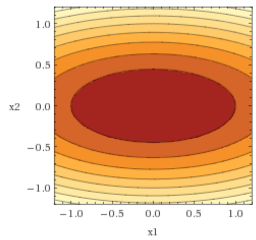
Convexity of sublevel sets

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

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



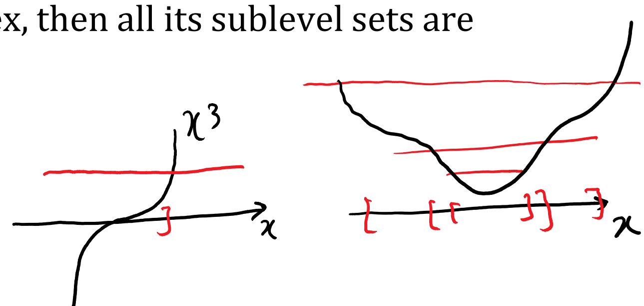




Several sublevel sets (for different values of α)

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

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

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

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

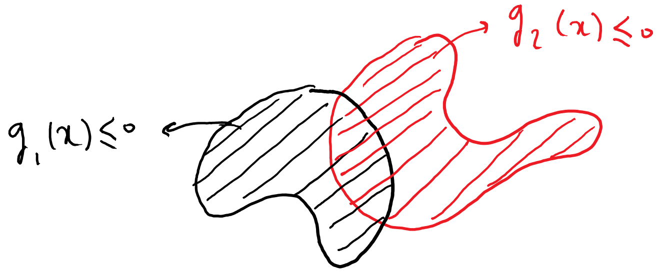
Convex optimization problems

A convex optimization problem is an optimization problem of the form

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

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: $\Omega = \{x \in \mathbb{R}^n | \ g_i(x) \leq 0, h_j(x) = 0\}$.
    - Observe that for a convex optimization problem Ω is a convex set (why?)
    - But the converse is not true:
        - Consider for example, $\Omega = \{x \in \mathbb{R} | \ x^3 \leq 0\}$. 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 $\Omega = \{x \in \mathbb{R} | \ x \leq 0\}$, 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:



$\Omega = \{x \mid g_1(x) \leq 0, \ g_2(x) \leq 0\}$ 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 $\leq$ Concave**
- **Affine $==$ Affine**

This is really the same as:

- **Convex ≤ 0**
- **Affine $== 0$**

Why?
(Hint: Sum of two convex functions is convex, and sums and differences of affine functions are affine.)

Notes:

- Further reading for this lecture can include the first few pages of Chapters 2,3,4 of [BV04]. Your [CZ13] book defines convex sets in Section 4.3. Convex optimization appears in Chapter 22. The relevant sections are 22.1-22.3.

References:

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

- [CZ13] E.K.P. Chong and S.H. Zak. An Introduction to Optimization. Fourth edition. Wiley, 2013.