In today's lecture we shift our focus towards optimization. Most of the material is an introduction to common terminology and problem formulations. The main focus is convex optimization.

1   Standard Notation

Optimization problem:

$$P: \quad \begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \quad \longleftarrow \text{ objective} \\ \text{s.t.} \quad & g_i(x) \le 0 \quad i = 1, \ldots, m \quad \longleftarrow \text{ inequality constraints} \\ & h_j(x) = 0 \quad j = 1, \ldots, p \quad \longleftarrow \text{ equality constraints} \end{aligned}$$

Feasible Set:

$$S = \left\{ x \in \mathbb{R}^n \mid g_i(x) \le 0 \text{ and } h_j(x) = 0 \text{ for all } i \text{ and } j \right\}$$

So we could rewrite the optimization problem P more compactly as $\min_{x \in S} f(x)$.

Some standard terminology:

$$\begin{aligned} \text{``}x \text{ is a feasible point''}: \quad & x \in S \\ \text{``}P \text{ is feasible''}: \quad & S \ne \emptyset \\ \text{``}P \text{ is infeasible''}: \quad & S = \emptyset \to \text{by convention, we say the optimal cost is } P_\star = +\infty \\ \text{``}P \text{ is unbounded''}: \quad & \text{the optimal cost is } P_\star = -\infty \end{aligned}$$

2   Affine/Linear Constraints

Constraints define the feasible set. Linear inequality constraints define *half-spaces*, which are all the points on one side of a $(n-1)$-dimensional affine space (which is also called a *hyperplane*). The intersection of multiple linear constraints forms a *polytope* (sometimes called a *polyhedron*).

$$\begin{aligned} a_1^{\mathsf{T}} x + b_1 &\le 0 \\ a_2^{\mathsf{T}} x + b_2 &\le 0 \\ &\vdots \\ a_m^{\mathsf{T}} x + b_m &\le 0 \end{aligned}$$

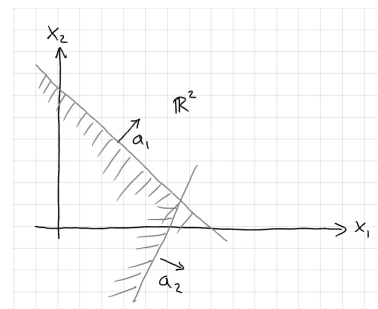write compactly as: $Ax + b \le 0$ (element-wise inequality)


Figure 1: polytope

3   Quadratic Constraints

These types of constraints lead to ellipsoidal, hyperbolic, and parabolic sets. They take the form:

$$x^\mathsf{T} Q_i x + 2p_i^\mathsf{T} x + r_i \leq 0$$

If $Q \succeq 0$, than the set is convex (more about this later!). Unlike linear constraints, these cannot be combined into a compact form and must be written separately.

When solving optimization problems, it is very helpful for $Q \succ 0$. This is true because most optimization algorithms take small steps towards the minimum. If there are different regions in the function like Figure 3 then this method because impossible.
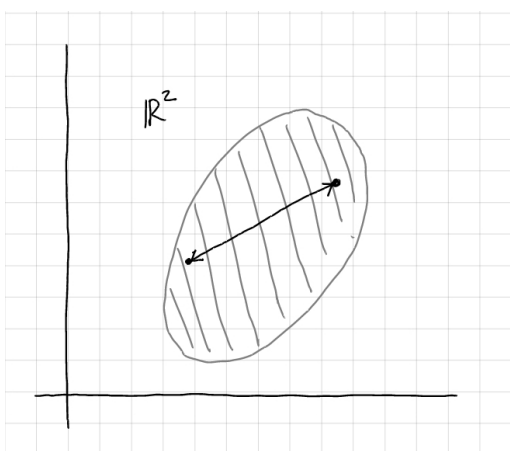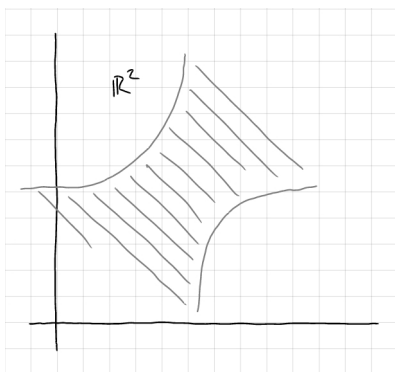

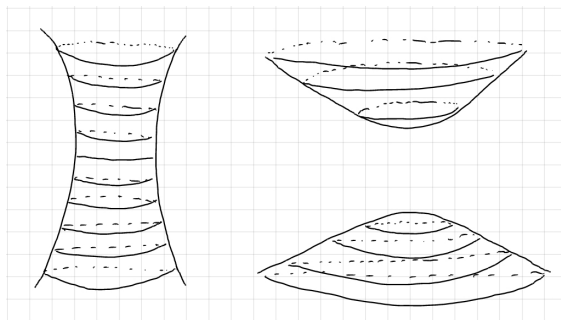(a) $Q \succ 0$ in $\mathbb{R}^2$; an ellipse.
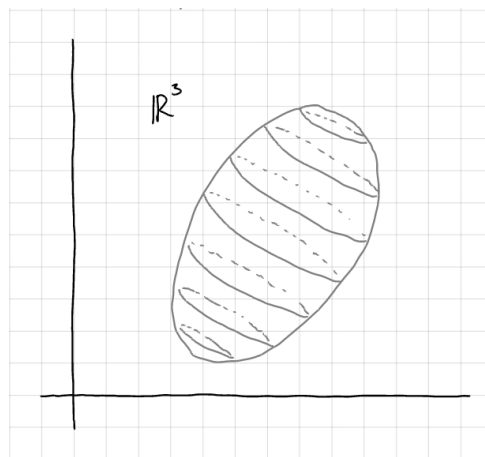
(b) $Q \succ 0$ in $\mathbb{R}^3$; an ellipsoid.

Figure 2: Positive-definite quadratic constraints (convex)

(a) Q indefinite in $\mathbb{R}^2$.

(b) Q indefinite in $\mathbb{R}^3$.

Figure 3: indefinite quadratic constraints (non-convex)

4 Defining Convex

We will define convexity for a set S and a function f. The definitions are related but their differences will become important in future lectures.

A set S is convex if the line segment joining any two points in the set also belongs to the set.

$$\underbrace{\alpha x + (1-\alpha)y}_{\text{convex combination}} \in S \quad \text{for all } x, y \in S \text{ and } \alpha \in [0,1]$$
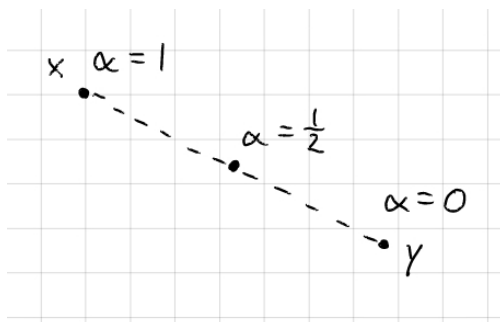

Figure 4: convex combination

If S_1 and S_2 are convex, then $S_1 \cap S_2$ is convex. Note that $S_1 \cup S_2$ is not necessarily convex. This can be seen in Fig. 5. Any line two points in the intersection can be connected with a line.
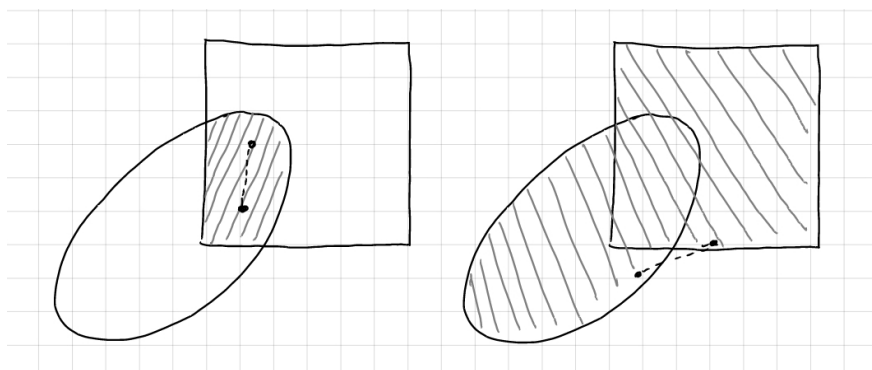

Figure 5: convex intersection and union

A function f is convex if for all pairs of points on the graph of the function, the line segment connecting them lies above the graph.

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) - (1-\alpha)f(y) \quad \text{for all } x, y \in \mathbb{R}^n \text{ and } \alpha \in [0,1]$$
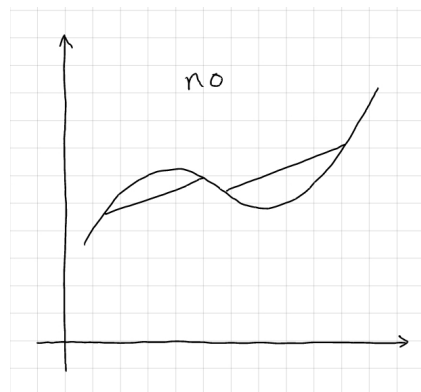
A function f is **concave** if $-f$ is convex. Some examples:

- convex: x^2, e^x, e^{-x}, $\frac{1}{x}$ for $x > 0$.
- concave: $-x^4$, $\log(x)$ for $x > 0$, $\frac{1}{x}$ for $x < 0$.
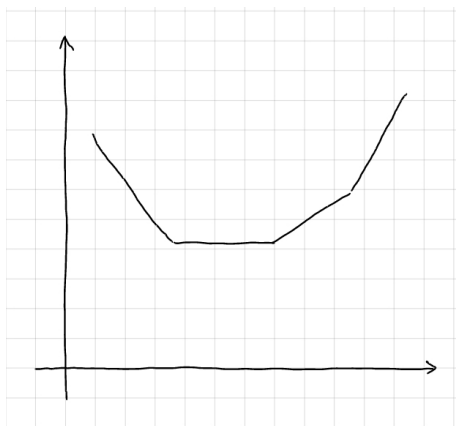
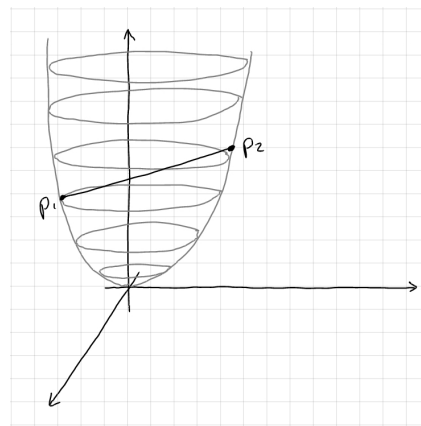The only functions that are both concave and convex are linear functions.
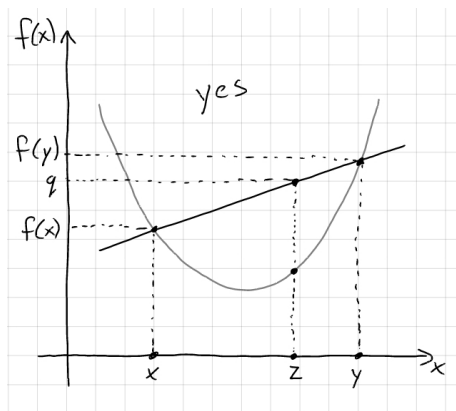
(a) Convex function

(b) Non-convex function

(c) Convex function in $\mathbb{R}^2$ that is not smooth.

(d) Convex function in $\mathbb{R}^3$.

Figure 6: convex function examples

5   Convexity relationships

For the definitions below, assume $f : \mathbb{R}^n \to \mathbb{R}$.

Graph of f.   The *graph* of a function is the set of pairs of points consisting of a point from the domain and the corresponding value of the function at that point.

$$\text{graph of } f : \left\{ (x,t) \in \mathbb{R}^{n+1} \mid f(x) = t \right\}$$

Sublevel sets of f.   The *sublevel sets* of a function are similar to the contour lines in an elevation map. A contour line represents points where the function value is the same.

$$\text{sublevel set of } f \text{ at level } t : \{ x \in \mathbb{R}^n \mid f(x) \le t \}$$

Epigraph of f: The epigraph of f is the set of points that lie above the graph of f.

$$\text{epi}(f) : \left\{ (x,t) \in \mathbb{R}^{n+1} \mid f(x) \le t \right\}$$

4

(a) sublevel sets in $\mathbb{R}^2$

(b) sublevel sets in $\mathbb{R}^3$

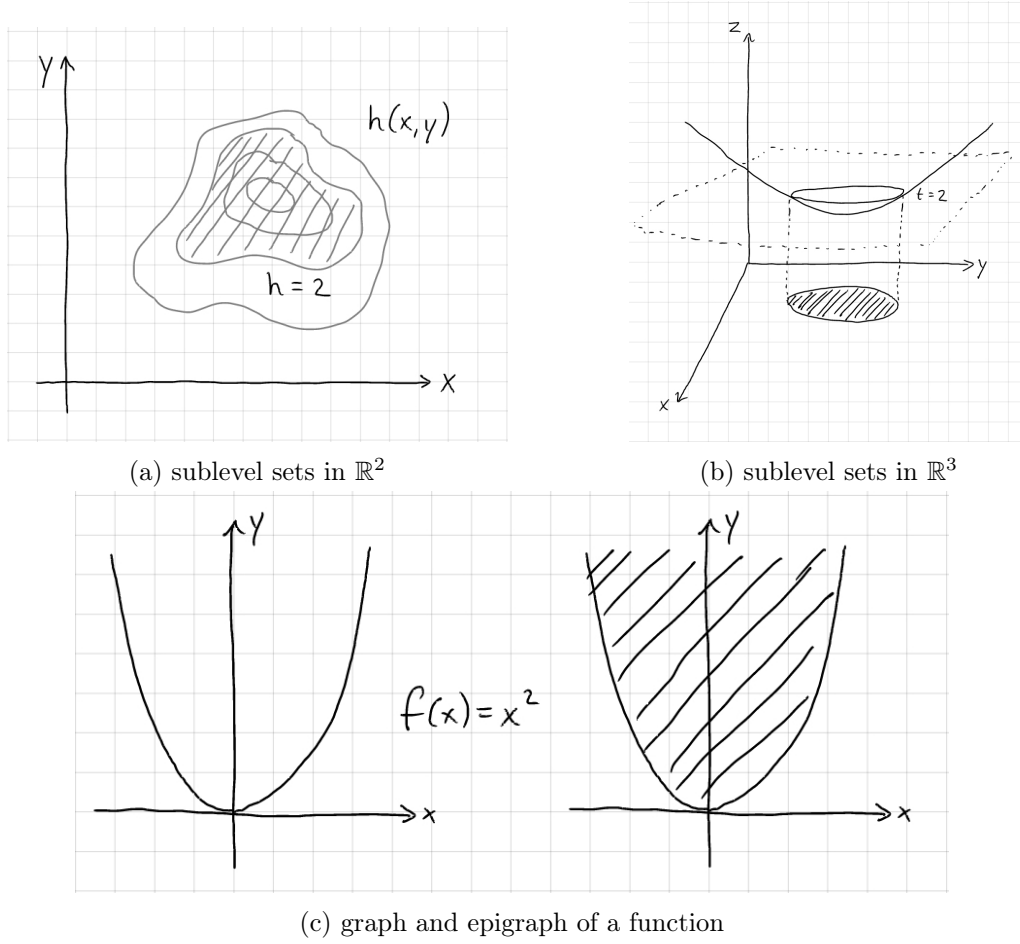(c) graph and epigraph of a function

Figure 7: Illustrations of the graph, epigraph, and sublevel sets of a function.

Key result: f is convex $\iff$ epi(f) is convex.

To prove this, first suppose f is convex and $(x, t), (y, \tau) \in$ epi(f). Then,

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \leq \alpha t + (1-\alpha)\tau$$

Where we used the fact that $f(x) \leq t$ and $f(y) \leq \tau$ and $\alpha \in [0, 1]$. We can now conclude that $\big(\alpha x + (1-\alpha)y, \alpha t + (1-\alpha)\tau\big) \in$ epi(f), so epi(f) is a convex set.

Conversely, suppose f is not convex, then there exists $x, y \in \mathbb{R}^n$ and $\alpha \in [0, 1]$ such that

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

Consider the points $(x, f(x)), (y, f(y)) \in$ epi(f). We have:

$$\alpha(x, f(x)) + (1-\alpha)(y, f(y)) = \underbrace{\big(\alpha x + (1-\alpha)y, \alpha f(x) + (1-\alpha)f(y)\big)}_{\notin \text{epi}(f) \text{ because of Eq. (1).}}$$

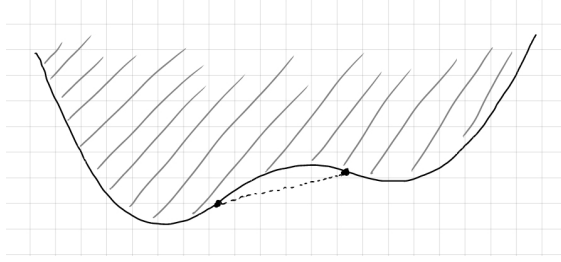This idea is show graphically in Fig. 8.

5

Figure 8: non-convex epigraph

Second key result: f is convex $\implies$ sublevel sets of (f) are convex. The proof is similar to that of the previous result.

The converse direction is *not true* in general. In other words, if a function has convex sublevel sets, the function is not necessarily convex. For counterexamples, see Fig. 9.

A function whose sublevel sets are convex is called *quasi-convex*. So convex functions are quasi-convex, but quasi-convex functions are not necessarily convex.
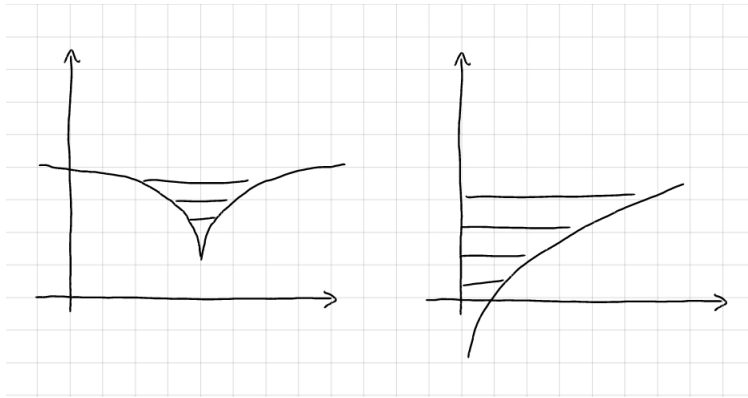

Figure 9: Not all convex sublevel sets are generated by convex functions.

6 Polytope representations

The convex hull of a set of points is the smallest convex set that contains all the points.

$$\text{conv}\{x_1, \ldots, x_m\} = \left\{ \sum_{i=1}^{m} \lambda_i x_i \,\middle|\, \lambda_1 + \cdots + \lambda_m = 1 \text{ and } \lambda_i \geq 0 \right\} \tag{2}$$

This is a general version of a convex combination. Another way to think about the convex hull is that it is the set of all convex combinations of the given set of points. The convex hull of a finite set of points is a polytope. For an illustration, see Fig. 10.
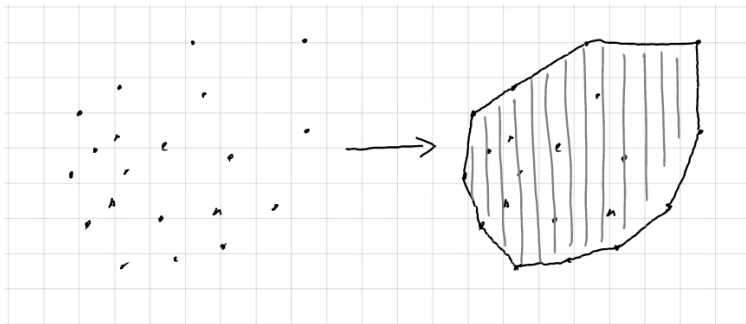

Figure 10: from a set to a convex hull

Faces vs. vertices. We have seen two different representations of polytopes now:

- As an intersection of half-spaces (face representation): $\{x \in \mathbb{R}^n \mid Ax \leq b\}$

- As the convex hull of a set of points (vertex representation): $\text{conv}\{x_1, \ldots, x_m\}$.

Depending on the context, sometimes the face representation is better, and other times the vertex representation is better. For example,

- A hypercube consists of the inequalities $-1 \leq x_i \leq 1$. Such a shape has $2n$ faces, but 2^n vertices. So representing a hypercube as a convex hull would require much more memory.

- A hyper-octahedron is the convex hull of the vertices $\pm e_i$. such a shape has $2n$ vertices, but 2^n faces. For example, there is a large difference between a cube and an octahedron. So representing a hyper-octahedron as a set of inequalities would require much more memory.