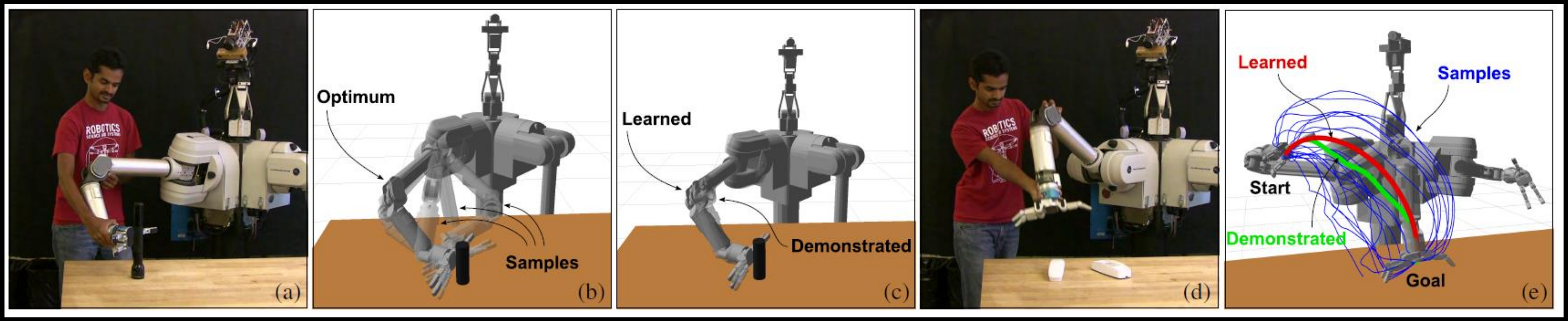
Convex Optimization

Using material from Stephen Boyd

This lecture is being recorded

Why do we need optimization in robotics?

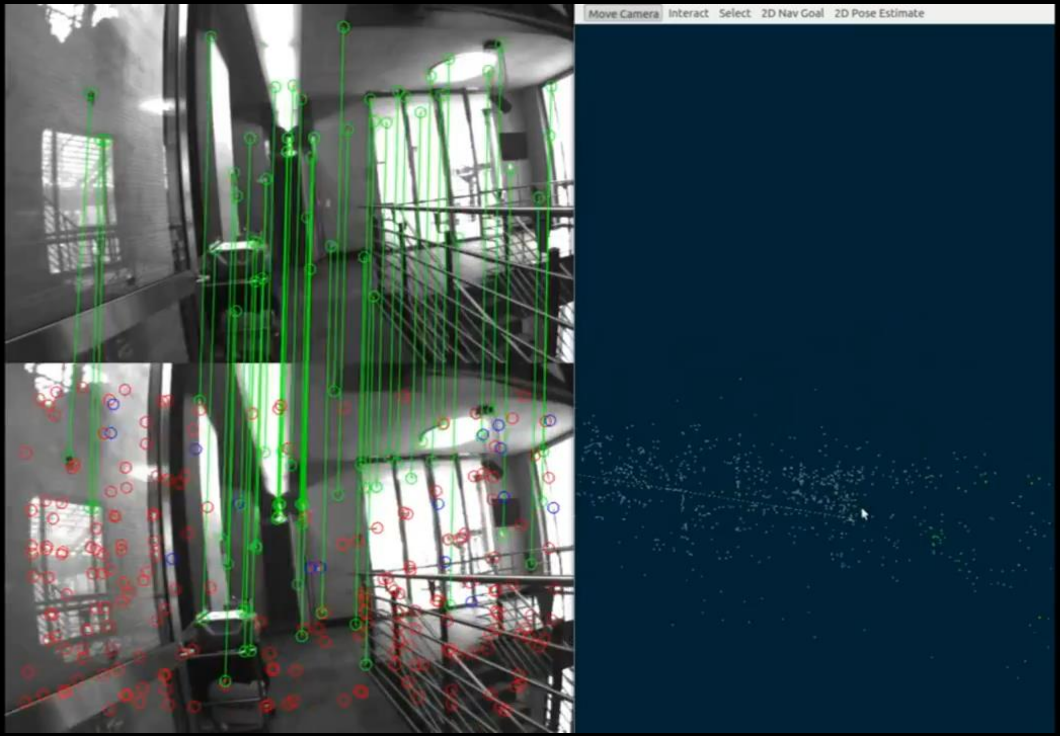
- Gives us a way to frame robotics problems mathematically
- VERY widely used
- Example: Inverse Optimal Control:



Learning Objective Functions for Manipulation
[Kalakrishnan et al., ICRA 2013]

Why do we need optimization in robotics?

- Example: Simultaneous Localization and Mapping (SLAM)



Keyframe-Based Visual-Inertial SLAM Using Nonlinear Optimization
[Leutenegger et al., RSS 2013]

Convex Optimization

- Convex optimization is a mature field with deep mathematical foundations

- It is so powerful that it's often worth it to
  - Work hard to reformulate your problem as convex
  - Approximate non-convex objective functions as convex
  - Use solution to approximation to start search for solution to the real problem

- It scales well with dimensionality
  - Convex optimization routinely solves problems with 1000s of variables

- Convex optimizers are fast (usually)

Outline

- Calculus Review
- Convex Sets
- Convex Functions

Set Notation

$$X = \{x | a^T x \leq b, x \in C, \quad a \in \boldsymbol{R}^n\}$$

X is the set of *xs* such that $a^T x \leq b$ is true for x in the set C where *a* is a vector in a Euclidian space of dimension *n*

Review: Functions

- Functions are defined as:

$$f : A \rightarrow B$$

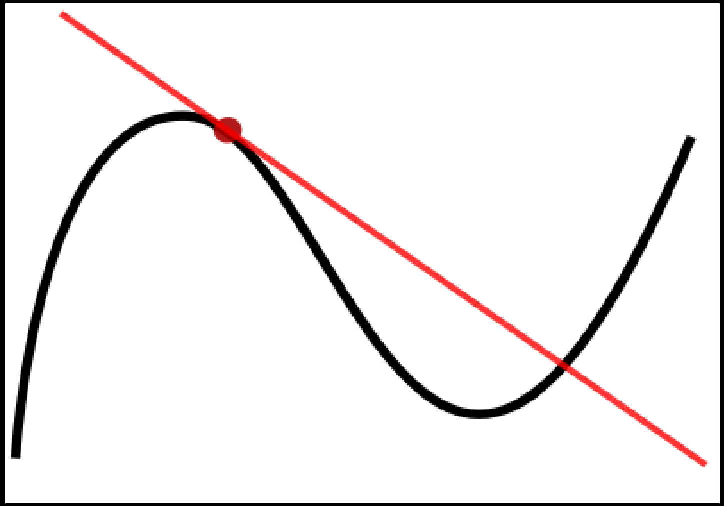
- "f maps elements in the set A to elements in the set B"
- The set A is the domain of f
- The set B is the range of f
- Example:

$$f : \mathbf{R}^n \rightarrow \mathbf{R}^m$$

"Function f maps n-dimensional vectors to some m-dimensional vectors"

Review: Derivatives

- Derivatives can get complicated!
- Keep this in mind: A derivative is a linear approximation of how a function changes a certain point



- The derivative of f(x) is the ratio between an infinitesimal change in an input variable x and the resulting change in the output f(x)

Review: Derivatives

- Recall the definition for a derivative $f: \mathbf{R} \to \mathbf{R}$

$$Df(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- We can write a similar definition for $f: \mathbf{R}^n \to \mathbf{R}^{\boldsymbol{m}}$

Review: Derivatives

- Suppose $f : \mathbf{R}^n \to \mathbf{R}^m$
- The function *f* is differentiable at *x* if there exists a matrix $\boxed{Df(x)} \in \mathbf{R}^{m \times n}$ that satisfies

This is a matrix

$$\lim_{z \in \mathbf{dom}\, f,\; z \neq x,\; z \to x} \frac{\|f(z) - f(x) - Df(x)(z - x)\|_2}{\|z - x\|_2} = 0$$

- *Df(x)* is called the derivative (or Jacobian) of the function
- *Df(x)* can be computed by computing partial derivatives

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \qquad i = 1, \ldots, m, \quad j = 1, \ldots, n$$

Review: Gradient

- When *f* is real-valued $(i.e.,\ f : \mathbf{R}^n \to \mathbf{R})$ the derivative *Df(x)* is a row vector (a 1 x n matrix)

  Range must be 1-dimensional!

- The transpose of the derivative is the gradient:

$$\nabla f(x) = Df(x)^T$$

- Again, you can compute the gradient by taking partial derivatives:

$$\nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, \ldots, n$$

Review: Second Derivative

- When f is real-valued $(i.e.,\ f : \mathbf{R}^n \rightarrow \mathbf{R})$ the **second** derivative is called the Hessian Matrix: $\nabla^2 f(x)$

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \qquad i = 1, \ldots n, \quad j = 1, \ldots, n,$$

- Recall that the second derivative is the derivative of the first derivative:

$$D \nabla f(x) = \nabla^2 f(x)$$

Questions

- Suppose we have a real-valued function $f : \mathbf{R}^n \to \mathbf{R}$

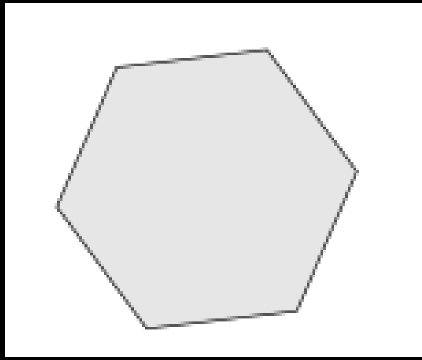
1. What are the dimensions of the gradient vector $\nabla f(x)$?

2. What are the dimensions of the Hessian matrix $\nabla^2 f(x)$?

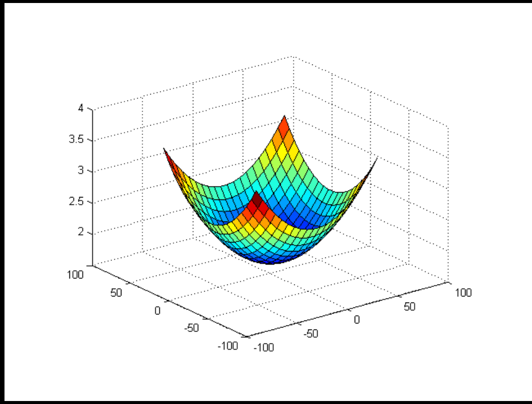
Convex Sets

Convex sets and functions

- Convexity is a restriction on shapes and functions
  - Convex optimization only works when everything is convex!
- We will cover definitions of convexity for shapes and functions and convexity-preserving operations
- You can use these to build convex functions for the problems you care about
- You can also use them to check if a function is convex
  - If f can be decomposed into convex functions and convexity-preserving operators, f is convex



A convex set



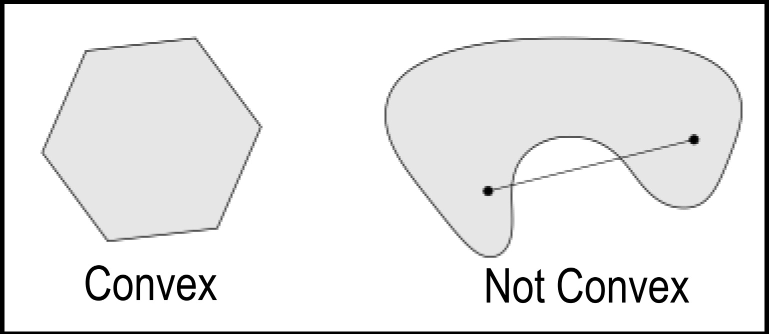
A convex function

Convex Sets

- Convex set: contains line segment between any two points in the set. *C* is a *convex set* if:

$$x_1, x_2 \in C, \quad 0 \le \theta \le 1 \quad \Longrightarrow \quad \theta x_1 + (1-\theta) x_2 \in C$$

- Examples:



Important Types of Convex Sets: Hyperplane

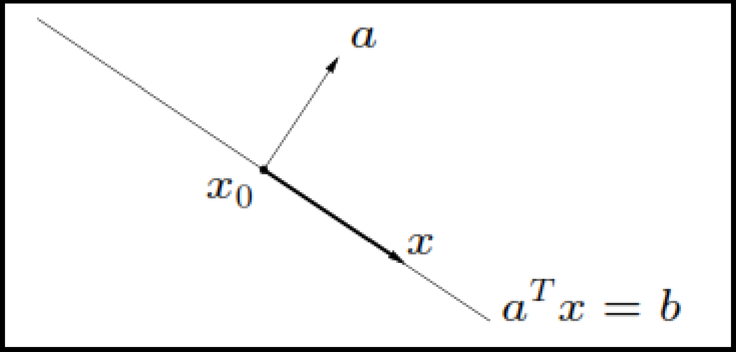
- Hyperplane: A set of points that have a constant inner product with vector a

$$\{x \mid a^T x = b\} \ (a \neq 0)$$

same as $a \cdot x$

- Another way to define it:

$$\{x \mid a^T (x - x_0) = 0\}$$



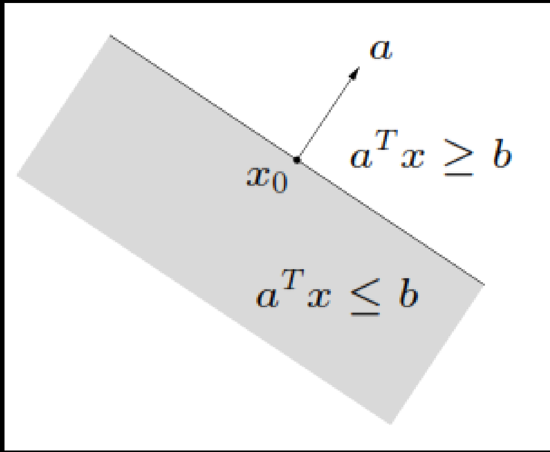
Important Types of Convex Sets: Halfspace

- Halfspace: A hyperplane with an inequality

$$\{x \mid a^T x \leq b\} \ (a \neq 0)$$

- Another way to define it :

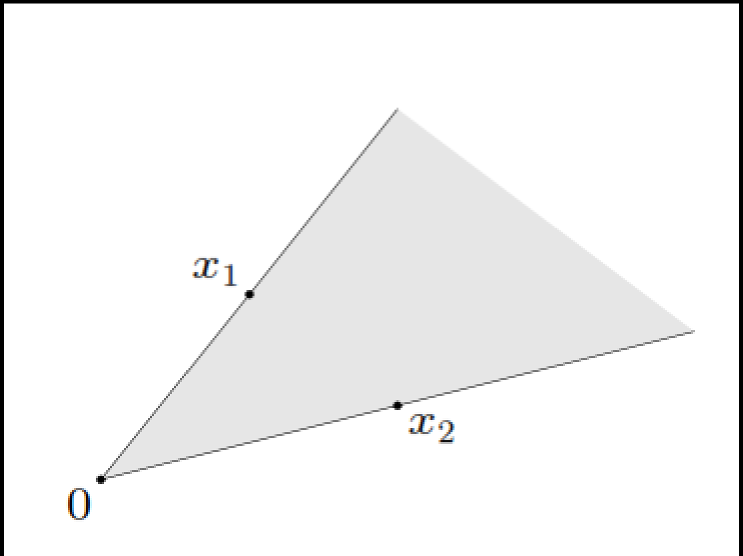
$$\{x \mid a^T (x - x_0) \leq 0\}$$



Important Types of Convex Sets: Convex Cone

- Convex Cone: A set C is a *convex cone* if

$$\theta_1 x_1 + \theta_2 x_2 \in C \qquad x_1,\ x_2 \in C \text{ and } \theta_1,\ \theta_2 \geq 0$$

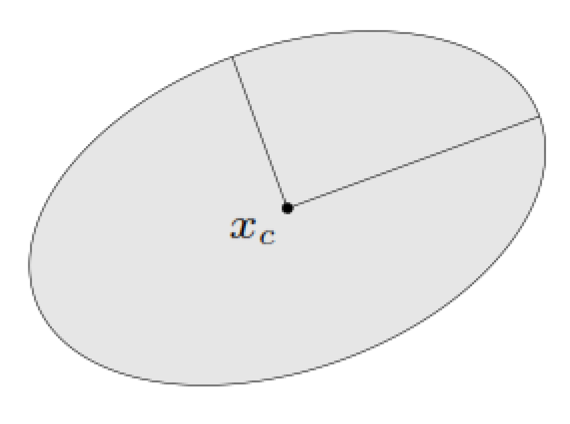


Important Types of Convex Sets: Ellipsoid

- Ellipsoid: Set of the form

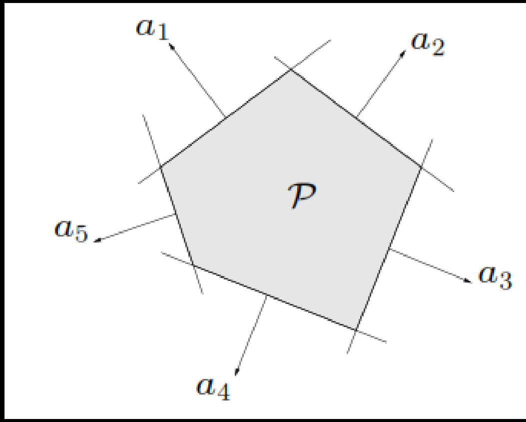
$$\{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$$

- P is a symmetric (i.e. $P = P^T$) positive definite matrix
  - Matrix P is positive definite if $z^T P z$ is positive for any non-zero z



Important Types of Convex Sets: Polyhedron

- Polyhedron: The intersection of a finite number of halfspaces and hyperplanes



- Another way to define it: The set of solutions to a set of linear inequalities and equalities:

$$Ax \leq b$$

$$Cx = d$$

Important Convexity-Preserving Operations on Sets

- Intersection preserves convexity

  If S_1 and S_2 are convex, then $S_1 \cap S_2$ is convex

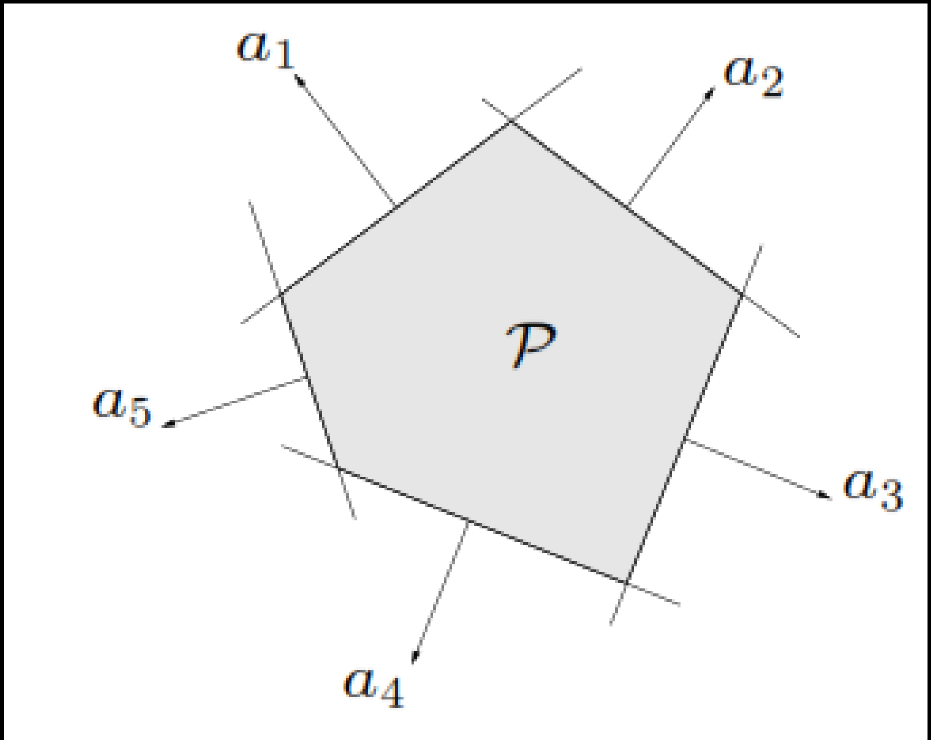
- It follows that the intersection of any number of convex sets is convex

- Affine functions preserve convexity

$$f : \mathbf{R}^n \to \mathbf{R}^m \qquad f(x) = Ax + b \text{ with } A \in \mathbf{R}^{m \times n},\ b \in \mathbf{R}^m$$

- Examples of affine functions
  - Scaling
  - Translation
  - Projection

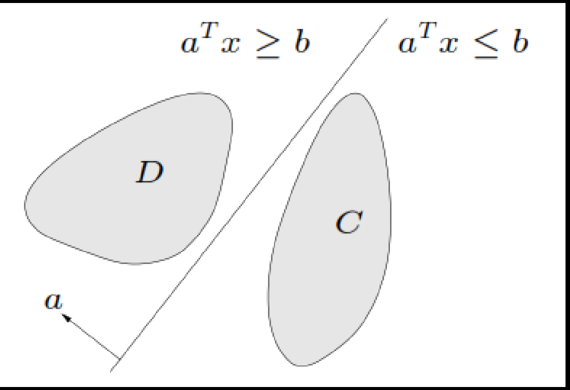
How do we know a polyhedron is always convex?



Separating Hyperplane Theorem

- If C and D are disjoint (i.e. $C \cap D = \emptyset$) convex sets, then there exists $a \neq 0$, b such that

$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$

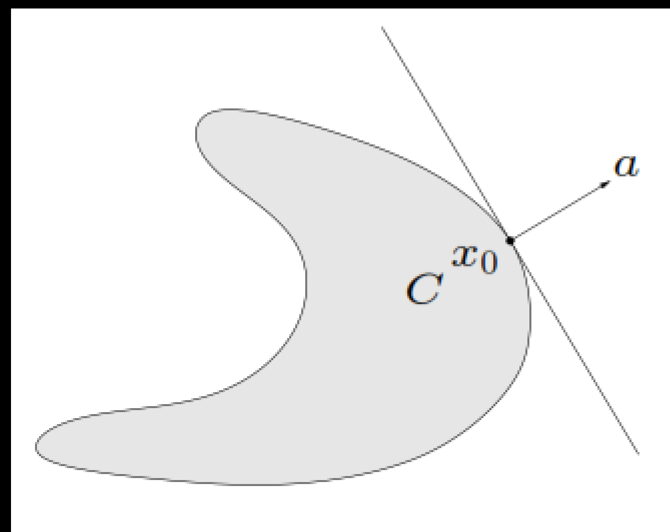


- C and D are separated by the hyperplane

$$\{x \mid a^T x = b\}$$

Supporting Hyperplane Theorem

If $a \neq 0$ satisfies $a^T x \leq \overset{b}{a^T x_0}$ for all $x \in C$, then the hyperplane $\{x \mid a^T x = a^T x_0\}$ is called a *supporting hyperplane* to C at the point x_0.



- Supporting Hyperplane Theorem: If C is convex, then there exists a supporting hyperplane at **every** boundary point of C.

Break

Convex Functions

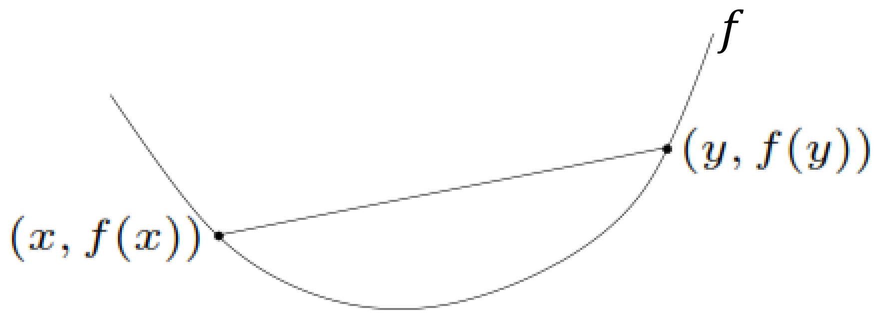
Convex Functions

The domain of the function

$f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and

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

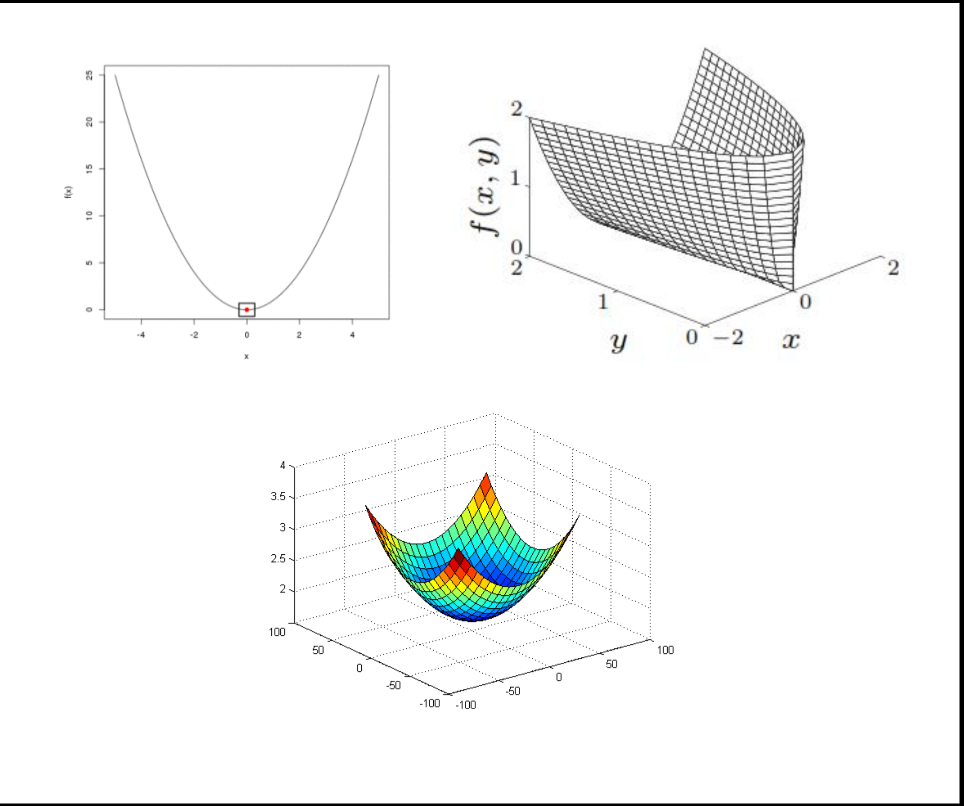
for all $x, y \in \mathbf{dom}\, f$, $0 \leq \theta \leq 1$



- I.e. the line segment between $(x, f(x))$ and $(y, f(y))$ lies above the graph of f

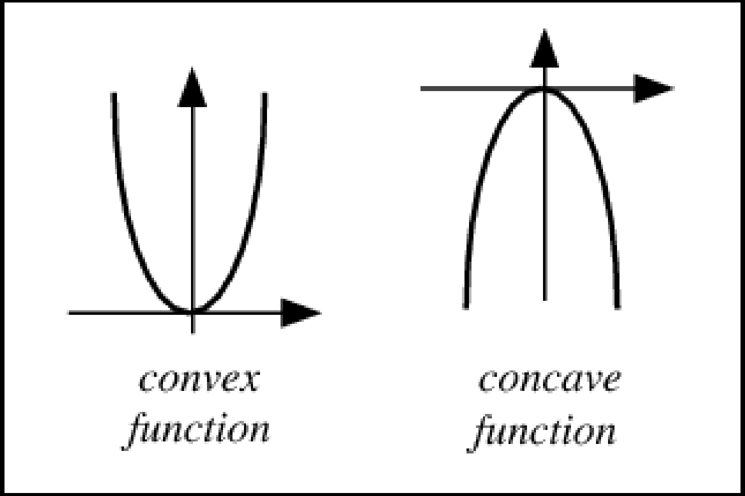
Advantage of convex functions

- Convex functions have only one local minimum!
  - That means local methods can find the global optimum!
    - (More on this in the next lecture)



Concave Functions

- Concave functions are convex functions that are “upside down”



- If $f(x)$ is convex, $-f(x)$ is concave.
- Some $f(x)$ are **both** concave and convex
  - Example?

Common Convex and Concave Functions

convex:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- exponential: e^{ax}, for any $a \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on $\mathbf{R}$, for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$

concave:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$

Convexity-Preserving Operations for Functions

- Non-negative multiplication

$$\alpha f \text{ is convex if } f \text{ is convex, } \alpha \geq 0$$

- Sum (extends to infinite sums and integrals)

$$f_1 + f_2 \text{ convex if } f_1, f_2 \text{ convex}$$

- Point-wise Maximum

$$\text{if } f_1, \ldots, f_m \text{ are convex, then } f(x) = \max\{f_1(x), \ldots, f_m(x)\} \text{ is convex}$$

Convexity-Preserving Operations

- Composition with affine functions

$$\boxed{f(Ax+b) \text{ is convex if } f \text{ is convex}}$$

- Composition in general

composition of $g : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$:

$$f(x) = h(g(x))$$

f is convex if $\begin{array}{l} g \text{ convex, } h \text{ convex, } \tilde{h} \text{ nondecreasing} \\ g \text{ concave, } h \text{ convex, } \tilde{h} \text{ nonincreasing} \end{array}$

"Extended-value extension of h"
We won't worry about it, just assume this is h

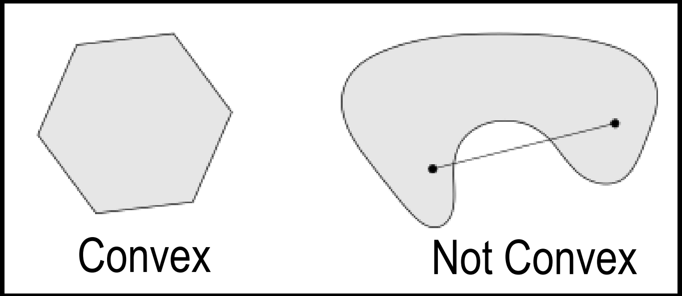
How do we use these functions/operators?

- Can use them to build convex functions for the problems you care about

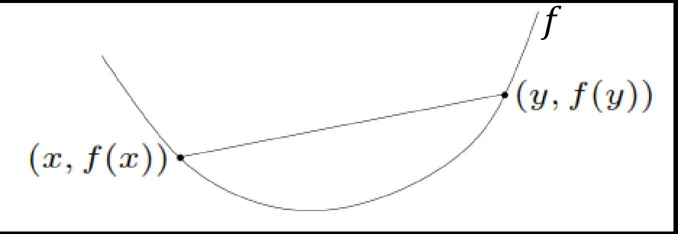
- Can use them to check if a function is convex
  - If f can be decomposed into convex functions and convexity-preserving operators, f is convex

Summary

- Convex sets are sets where a line segment between any two points is part of the set



- Convex functions are functions where the line segment between any two points is above the graph of the function



- Certain operators can be used to transform convex sets/functions while preserving convexity
  - Use them to assemble/decompose more complex functions

Homework

- Reading from optimization book
  - Descent Methods (Ch. 9.1-9.1.1, 9.2, 9.3-9.3.1, 9.5-9.5.2, 9.5.4)
- Subgradients

  https://see.stanford.edu/materials/lsocoee364b/01-subgradients_notes.pdf

  (everything except Section 4)
- Numerical differentiation

  https://en.wikipedia.org/wiki/Numerical_differentiation

  (up to “Complex-variable Methods”)