# Convex Optimization
# CMU-10725

## 2. Linear Programs

Barnabás Póczos & Ryan Tibshirani

MACHINE LEARNING DEPARTMENT
Carnegie Mellon.
School of Computer Science

# Administrivia

- **Please ask questions!**
- Lecture = 40 minutes part 1 - 5 minutes break – 35 minutes part 2
- Slides: http://www.stat.cmu.edu/~ryantibs/convexopt/
- Anonym feedback survey will be on black board next week. **Please use it!** Constructive feedback and suggestions are always welcome!
- Subscribe for scribing!
- My office hour is after the class.

# Basic Definitions

- More and more complicated optimization problems
- Definition of LP

# Simplest Optimization Problems

Goal:

- Constant function
- 1-dim linear function
- 1-dim linear function with bound constraints

# Linear Programs

- n-dim linear function with m linear constraints

**Inequality form**:

Cost function:

Constraints:

Bounds:

# Linear Programs

**Inequality form using matrix notation**:

Cost function:

Constraints:

Bounds:

**Example**:

# Goal of this (...and next) lecture(s)

- **To be able to solve Linear Programs**

  Simplex Algorithm (Phase I and Phase II)
  (Later we will see other algorithms too)

- **Understand why LP is useful**
  - Motivation
  - Applications in Machine Learning

- **Understand the difficulties**
  - Convergence? Polynomial or Exponential many operations?
  - Will algorithms find the exact solutions, or only approximate ones?

# Table of Contents

# Linear Programs

- Motivation
- History
- Sketching LP

# History

**Dantzig 1947 (Simplex method)**

**(**one of the top 10 algorithms of the twentieth century**)**

**Motivated by World War II:**

- Job scheduling (Assign 70 men to 70 jobs)
- Blending problem
  (produce a blend (30% Lead, 30% Zinc, 40% Tin) out of 9 different alloys (different mixture, different costs) such that the cost is minimal)
- Network flow optimization (Max flow min cut)

# The product mix problem

A furniture company manufactures four models of desks

Number of man hours and profit:

| | Desk 1 | Desk 2 | Desk 3 | Desk 4 | Available hrs |
|---|---|---|---|---|---|
| Carpentry shop hrs | 4 | 9 | 7 | 10 | 6000 |
| Finishing shop hrs | 1 | 1 | 3 | 40 | 4000 |
| Profit | $12 | $20 | $18 | $40 | |

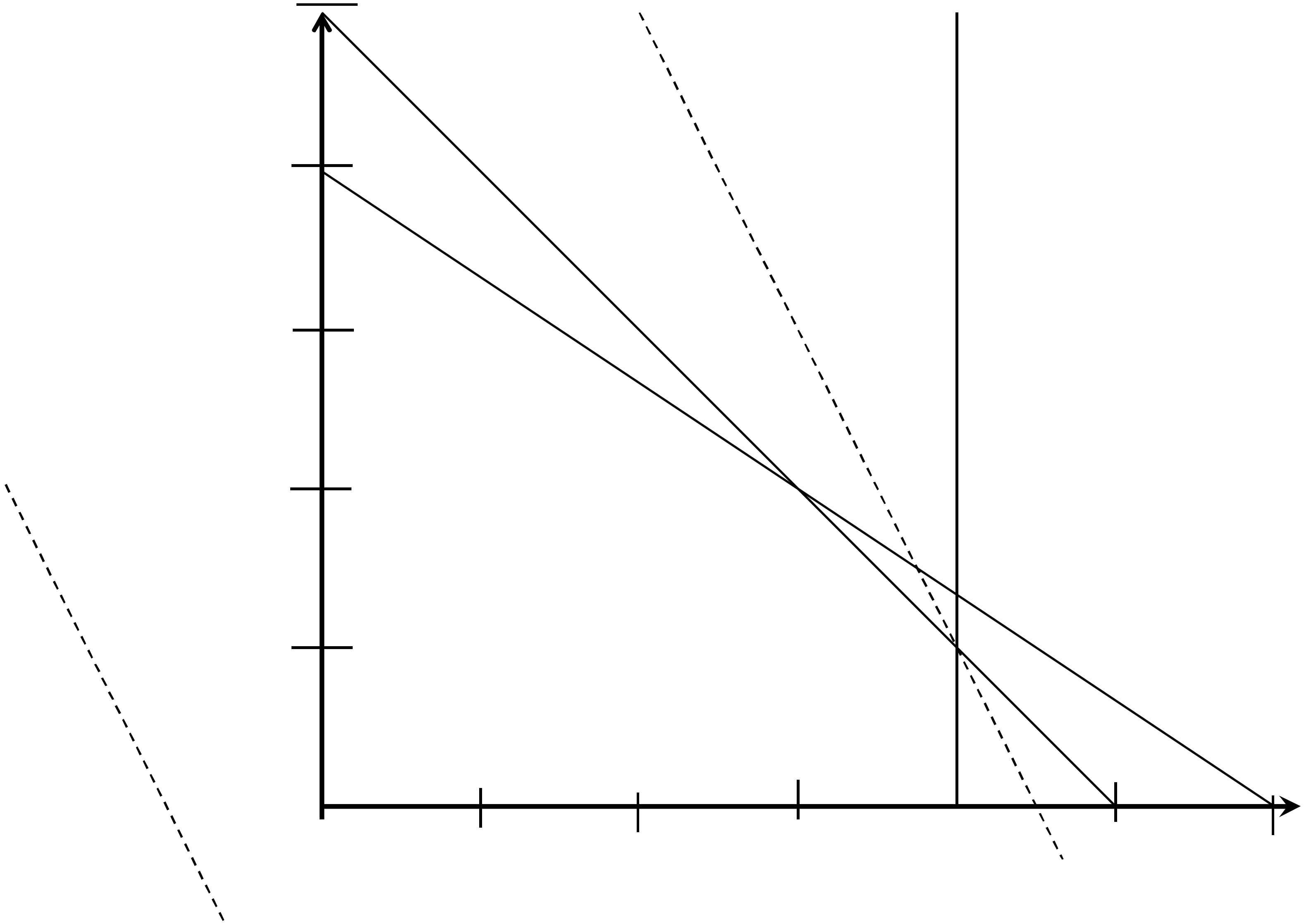
**Why is it called Linear Programing???**

# **Motivation**: Why Linear Programing?

- The simplest, nontrivial optimization problem
- Many complex system (objective and constraints) can be well approximated with linear equations
- Important applications
- There are efficient toolboxes that can solve LPs

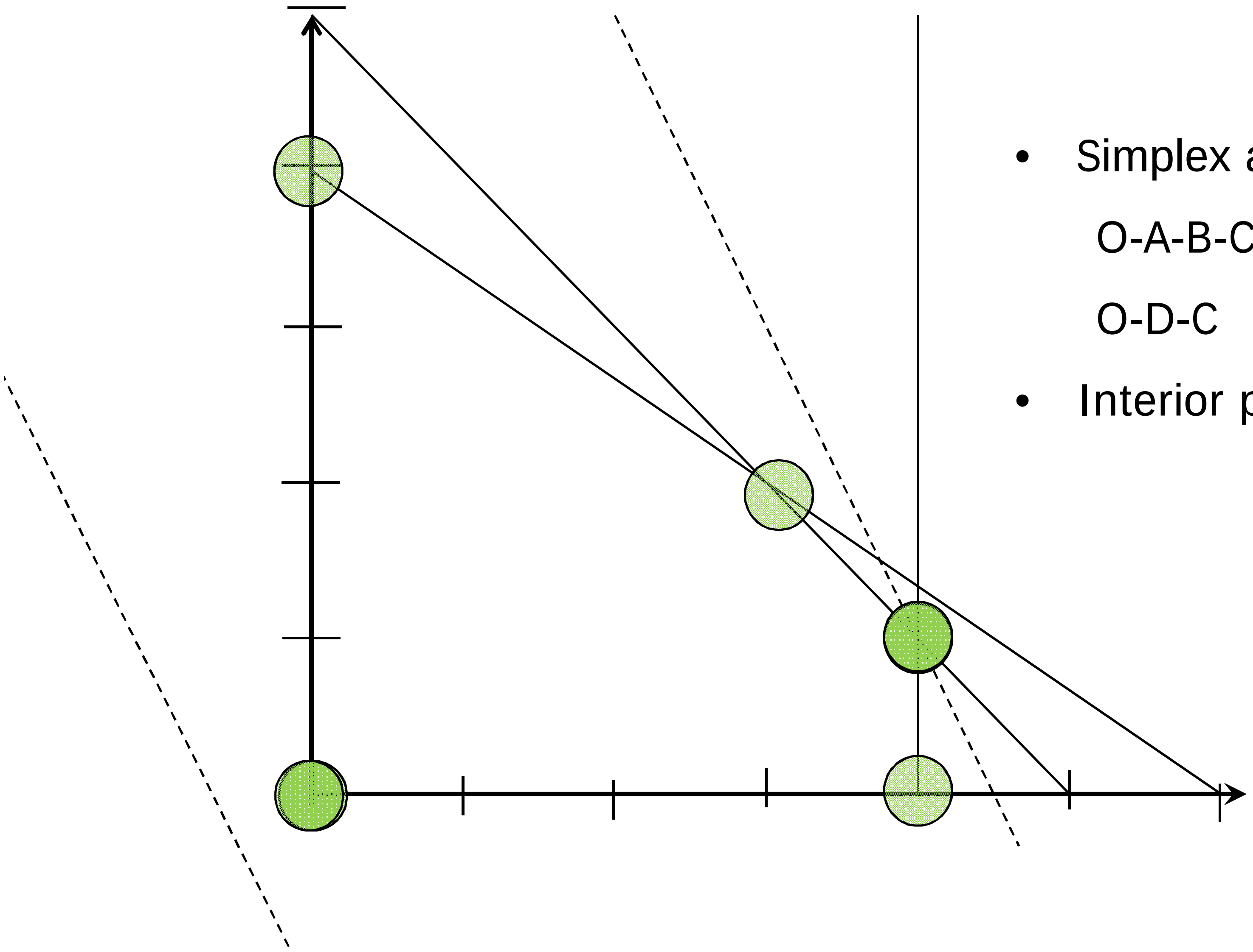
# Sketching Linear Programs

**Example**:



# Simplex Algorithm

**Example**:



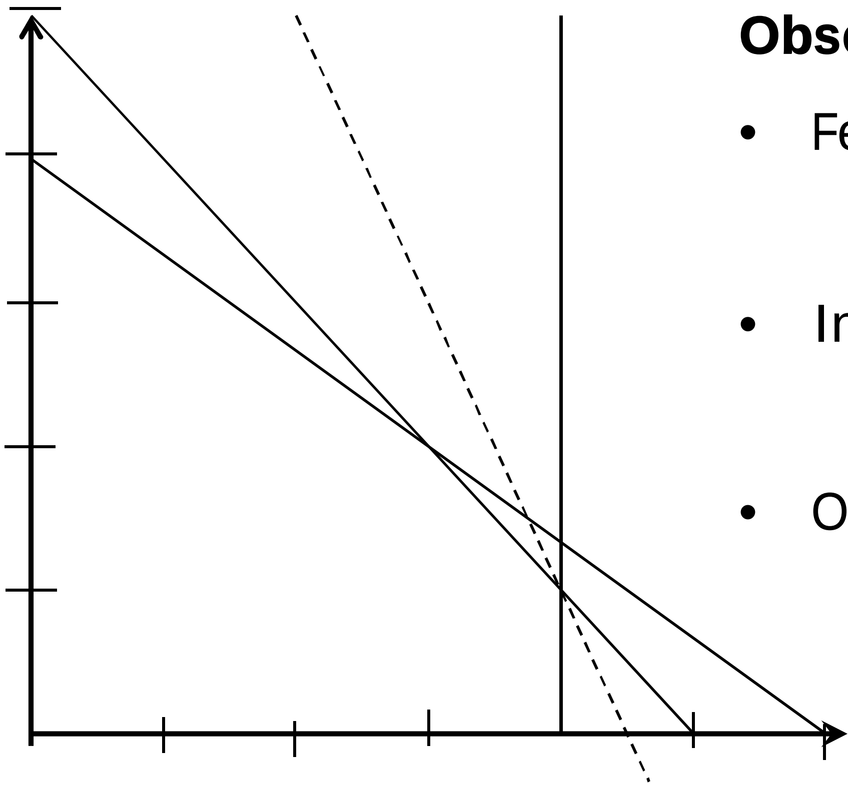
- Simplex algorithm:

  O-A-B-C

  O-D-C

- Interior point methods

# Linear Program



**Observations, Difficulties:**

- Feasible set might not exist, no solution

  (Inconsistency in the constraints)

- Infinite many global optimum

  (Optimum is on an edge)

- Optimum can be – 1 , 1

  (Unbounded optimum)

# Linear Program

**High dimensional case is similar:**

faces, facets instead of edges

cost function = hyperplane

# Applications

**Pattern Classification via Linear Programming**

# Application

**Pattern Classification via Linear Programming**

More info can be found on: cgm.cs.mcgill.ca/~beezer/cs644/main.html

**Goal**: show how LP can be used for linear classification.

**Why LP?**

There are many efficient LP solver software packages

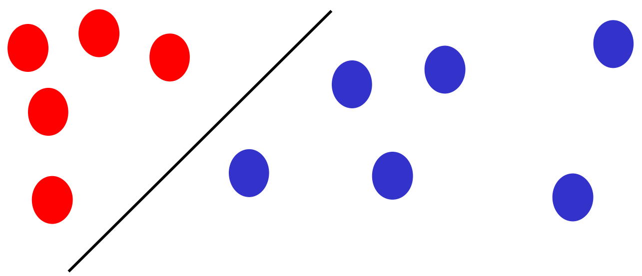
# Pattern Classification via LP

**Formal goal:**

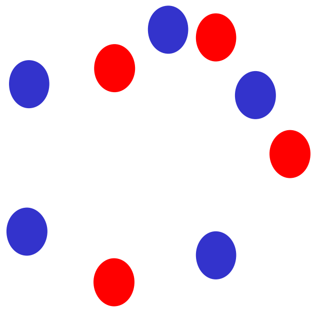
**Problem 1**: Determine whether H and M are linearly separable

**Problem 2**: If H and M are linearly separable,
then find a separating hyper plane

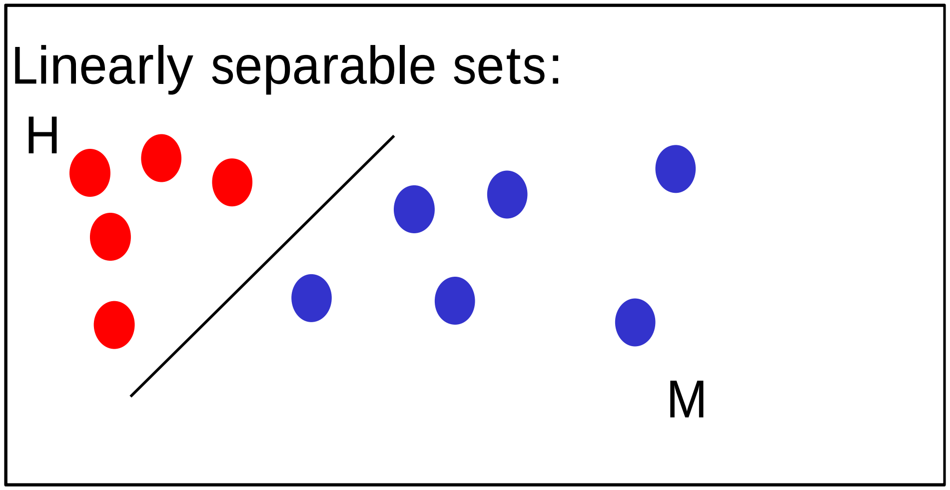
Linearly separable sets:

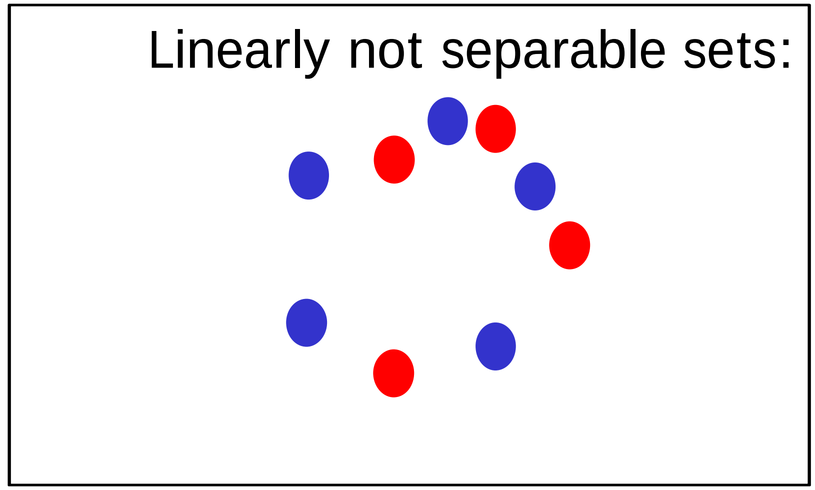


Linearly not separable sets:



# Pattern Classification via LP




**Observation**:

H and M are linearly separable

# Pattern Classification via LP

**Lemma 1:**

H and M are linearly separable

**Proof**

# Pattern Classification via LP

**Lemma 1:**

H and M are linearly separable

**Proof**

# Pattern Classification via LP

**Proof continued**

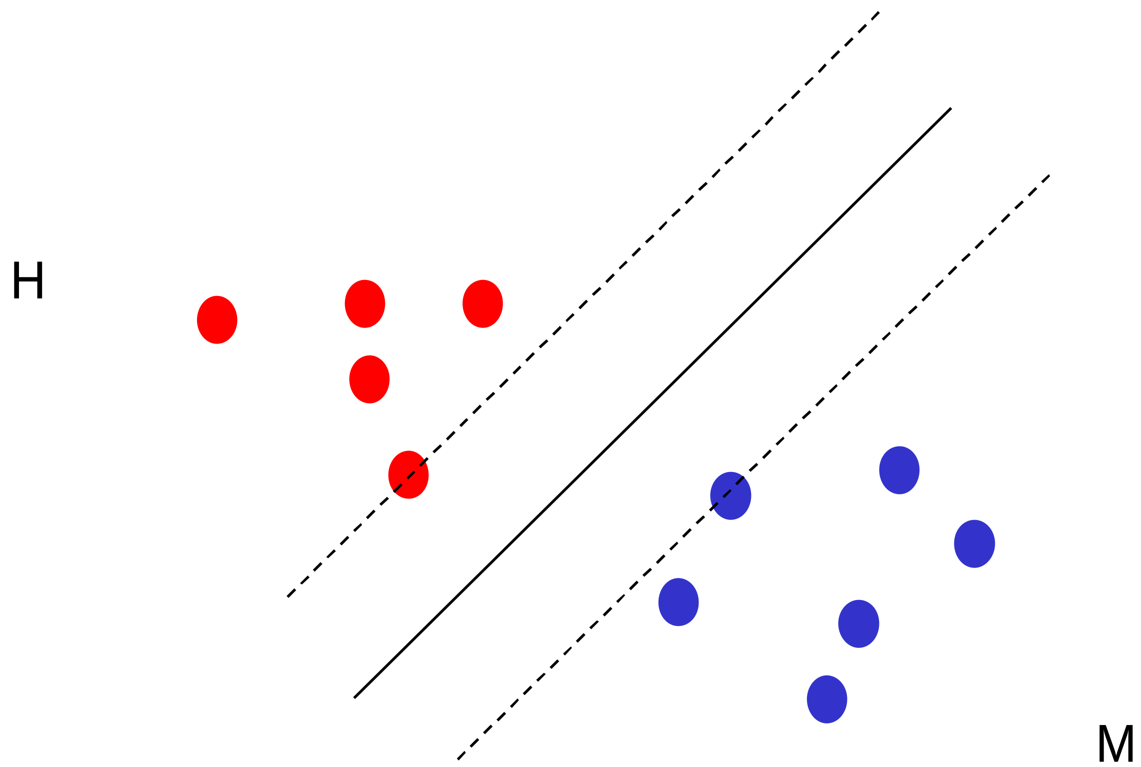
# Pattern Classification via LP

**Proof continued**

**Similarly,**

# Pattern Classification via LP

**Proof continued**



# Pattern Classification via LP

We will see that the following linear problem solves Problem 1 & 2:

[Mansgarian 1995]

# Pattern Classification via LP

**Theorem 1**

H and M are linearly separable iff the optimal value of LP is 0.

**Theorem 2**

H and M are linearly separable

$y^*, z^*, a^*, b^*$ is an optimal solution of (LP)

⇒ $f(x)=a^{*T}x+b^*$ is a separating hyperplane

# Pattern Classification via LP

**Proof of Theorems 1 and 2**

The optimal value of (LP) is 0

# Pattern Classification via LP

**Application:** Breast Cancer Diagnosis

Used at the University of Wisconsin Hospital

[Mangasarian et al 1995]

1. Fluid sample from breast.
2. Placed on a glass and stained the highlight the nuclei of cells
3. Image is taken
4. 30D features: Area, Radius, perimeter, etc

Goal: Classification between benign lumps and malignant lumps

Results: 97.5% accuracy

# Pattern Classification via LP

**Example 1: Linearly Separable Case**

# Pattern Classification via LP

**Example 2: Linearly nonseparable case**

# Linear Programs

- Standard from, Canonical form, Inequality form
- Transforming LPS
  - Pivot transformation

# Linear Programs

**Inequality form** of LPs using matrix notation:

**Standard form** of LPs:

**Theorem:** Any LP can be rewritten to an equivalent standard LP

# Transforming LPs

**Theorem:** Any LP can be rewritten to an equivalent standard LP

- Getting rid of inequalities (except variable bounds)

- Getting rid of equalities

# Transforming LPs

- Getting rid of negative variables

- Getting rid of bounded variables

- Max to Min

- Negative $b_i$