Optimization Problems

LING 572
Advanced Statistical Methods for NLP
January 28, 2020

Announcements

- HW2 grades posted: 85.7 avg
  - Historically the hardest / most time-consuming assignment of 572
  - [NB: we accept +/- 5% differences from our results on test data]
- Format moving forward (no points this time; will update the specs):
  - Each line needs to have the specified format
  - Blocks for the classes need to be in the same order as example file (e.g. talk.politics.guns, then talk.politics.mideast, talk.politics.misc)
  - Include the same commented lines as the example files

Performance

- A fair number of people struggled to get reasonable performance
  - Depth + recursion = explosion
- General lesson: think about what *repeated operations* you will be doing a lot, and choose data structures that do those efficiently
  - e.g. dicts/sets are hash tables, so very efficient lookup / insertion (O(1) avg)
  - Useful for the built-in datatypes: https://wiki.python.org/moin/TimeComplexity
- Common issue: pandas data-frames can be slow
- To the quick live demo!

Linguistics Twitter Field Day

Tweet

San Miguel Sheriff
@SheriffAlert

Large boulder the size of a small boulder is completely blocking east-bound lane Highway 145 mm78 at Silverpick Rd. Please use caution and watch for emergency vehicles in the area.

link

Optimization

What is an optimization problem?

- The problem of finding the best solution from all feasible solutions.
- Given a function $f : X \to \mathbb{R}$, find $x_0 \in X$ that optimizes f.
- f is called
  - an objective function,
  - a loss function or cost function (minimization), or
  - a utility function or fitness function (maximization), etc.

- X is an *n*-dimensional vector space:
  - discrete (possible values are countable): combinatorial optimization problem
  - continuous: e.g., constrained problems

Components of each optimization problem

- Decision variables X: describe our choices that are under our control.
  - We normally use n to represent the number of decision variables, and x_i to represent the i-th decision variable.

- Objective function f: the function we wish to optimize

- Constraints: describe the limitations that restrict our choice for decision variables.

Standard form of a continuous optimization problem

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m \\
& h_j(x) = 0, \quad j = 1, \dots, p
\end{aligned}
$$

where

- $f: \mathbb{R}^n \to \mathbb{R}$ is the **objective function** to be minimized over the n-variable vector x,
- $g_i(x) \leq 0$ are called **inequality constraints**
- $h_j(x) = 0$ are called **equality constraints**, and
- $m \geq 0$ and $p \geq 0$.

Common types of optimization problem

- Linear programming (LP) problems:
  - Definition: Both objective function and constraints are linear
  - The problems can be solved in polynomial time.
  - https://en.wikipedia.org/wiki/Linear_programming

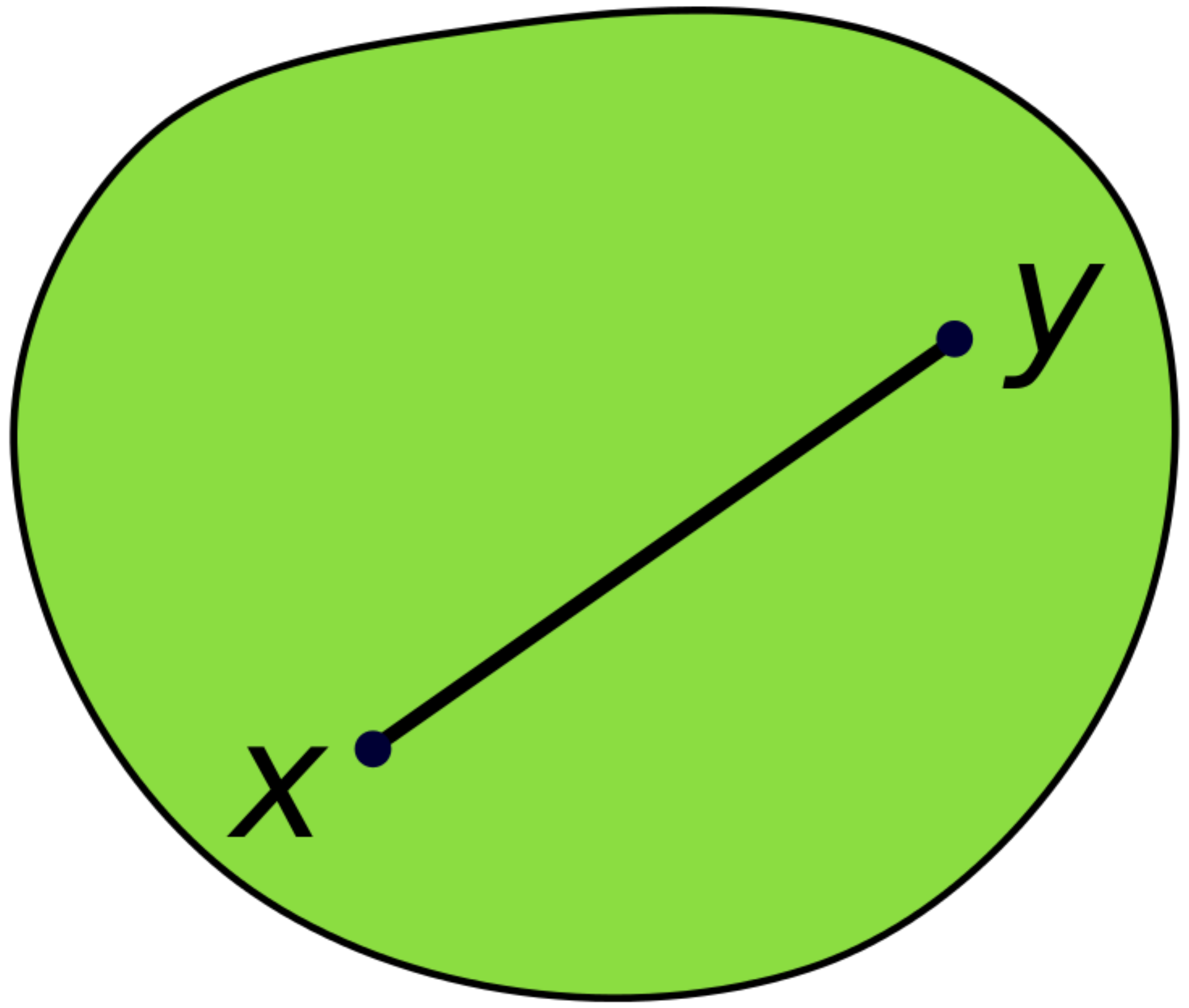
- Integer linear programming (ILP) problems:
  - Definition: LP problem in which some or all of the variables are restricted to be integers
  - Often, solving ILP problem is NP-hard.
  - https://en.wikipedia.org/wiki/Integer_programming

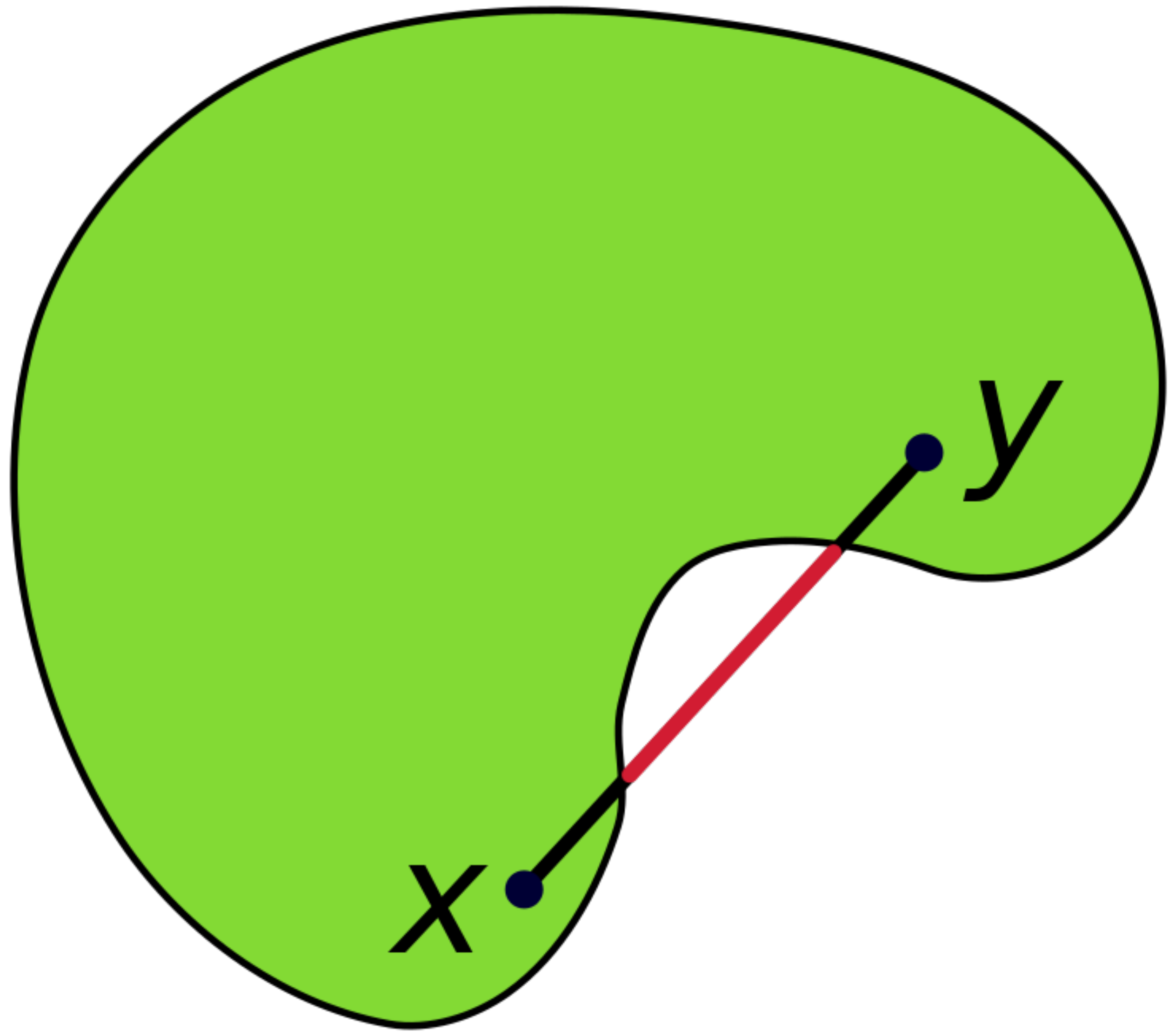
Common types of optimization problem (cont'd)

- Quadratic programming (QP):
  - Definition: The objective function is quadratic, and the constraints are linear
  - Solving QP problems is simple under certain conditions
  - https://en.wikipedia.org/wiki/Quadratic_programming

- Convex optimization:
  - Definition: f(x) is a convex function, and X is a convex set.
  - Property: if a local minimum exists, then it is a global minimum.
  - https://en.wikipedia.org/wiki/Convex_optimization

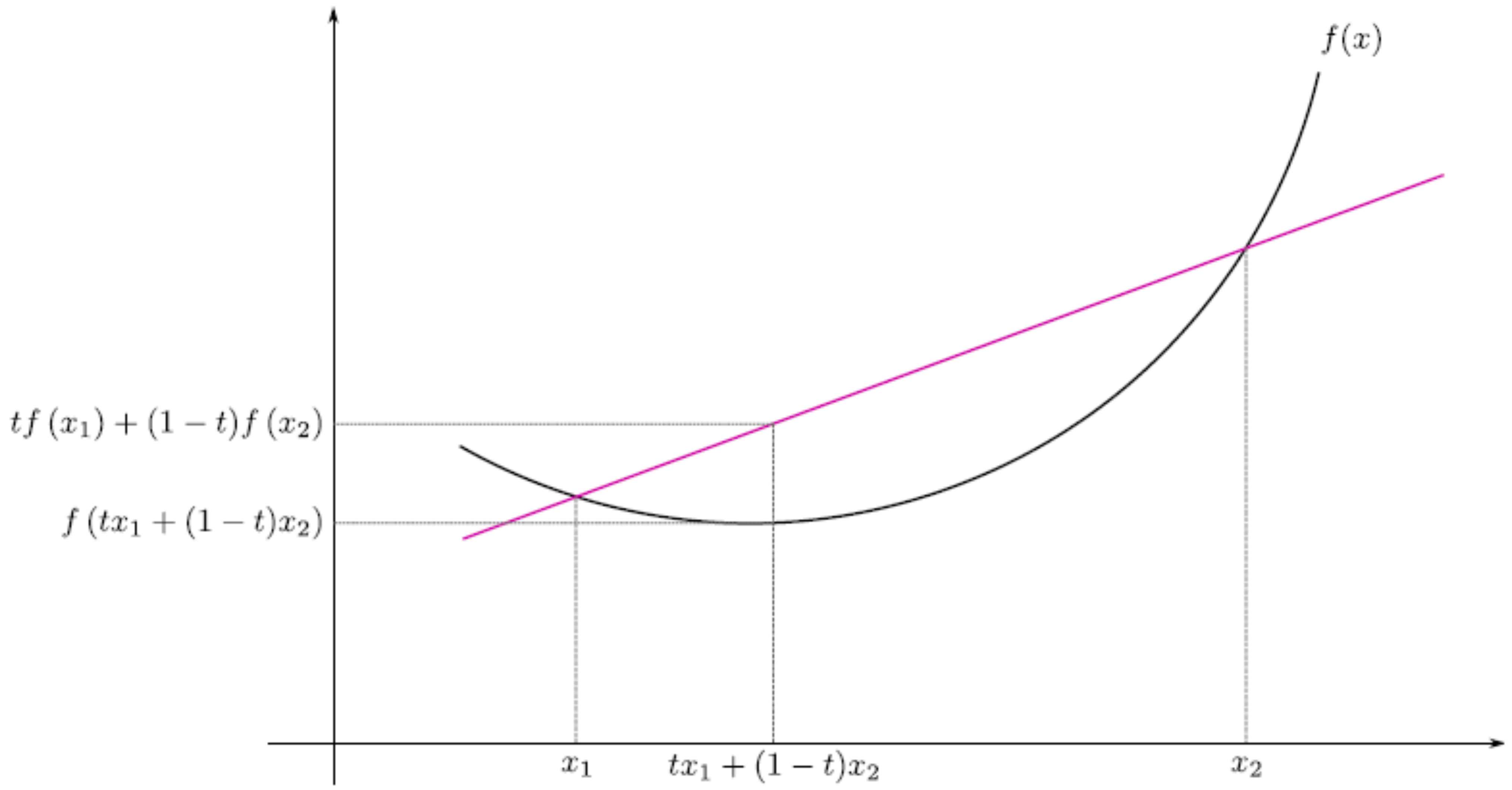
Convex set

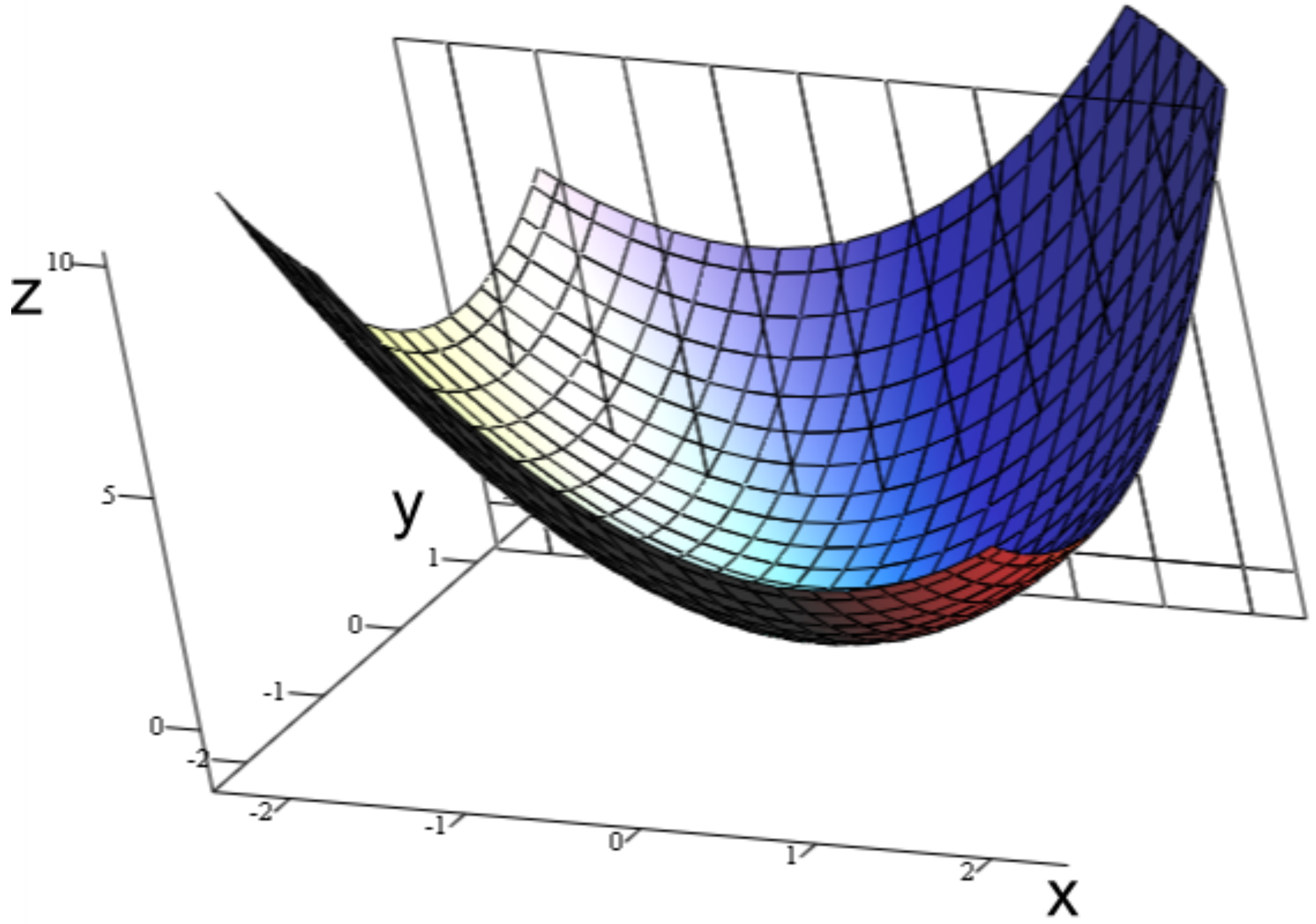




A set C is said to be **convex** if, for all x and y in C and all t in the interval (0, 1), the point $(1 - t)x + ty$ also belongs to C

Convex function





- Let X be a convex set in a real vector space and $f : X \to \mathbb{R}$ a function.
- f is convex just in case:
  - $\forall x_1, x_2 \in X, \forall t \in [0,1],\ f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$
- (strictly convex: strict inequality, with t ranging in (0, 1), excluding endpoints.)

Terms

- A solution is the assignment of values to all the decision variables
- A solution is called feasible if it satisfies all the constraints.
- The set of all the feasible solutions forms a feasible region.
- A feasible solution is called optimal if $f(x)$ attains the optimal value at the solution.

Terms

- If a problem has no feasible solution, the problem itself is called infeasible.

- If the value of the objective function can be infinitely large, the problem is called unbounded.

Linear programming

Linear Programming

- The linear programming method was first developed by Leonid Kantorovich in late 1930s.
- Main applications: diet problem, supply problem
- A primary method for solving LP is the simplex method.
- LP problems can be solved in polynomial time.

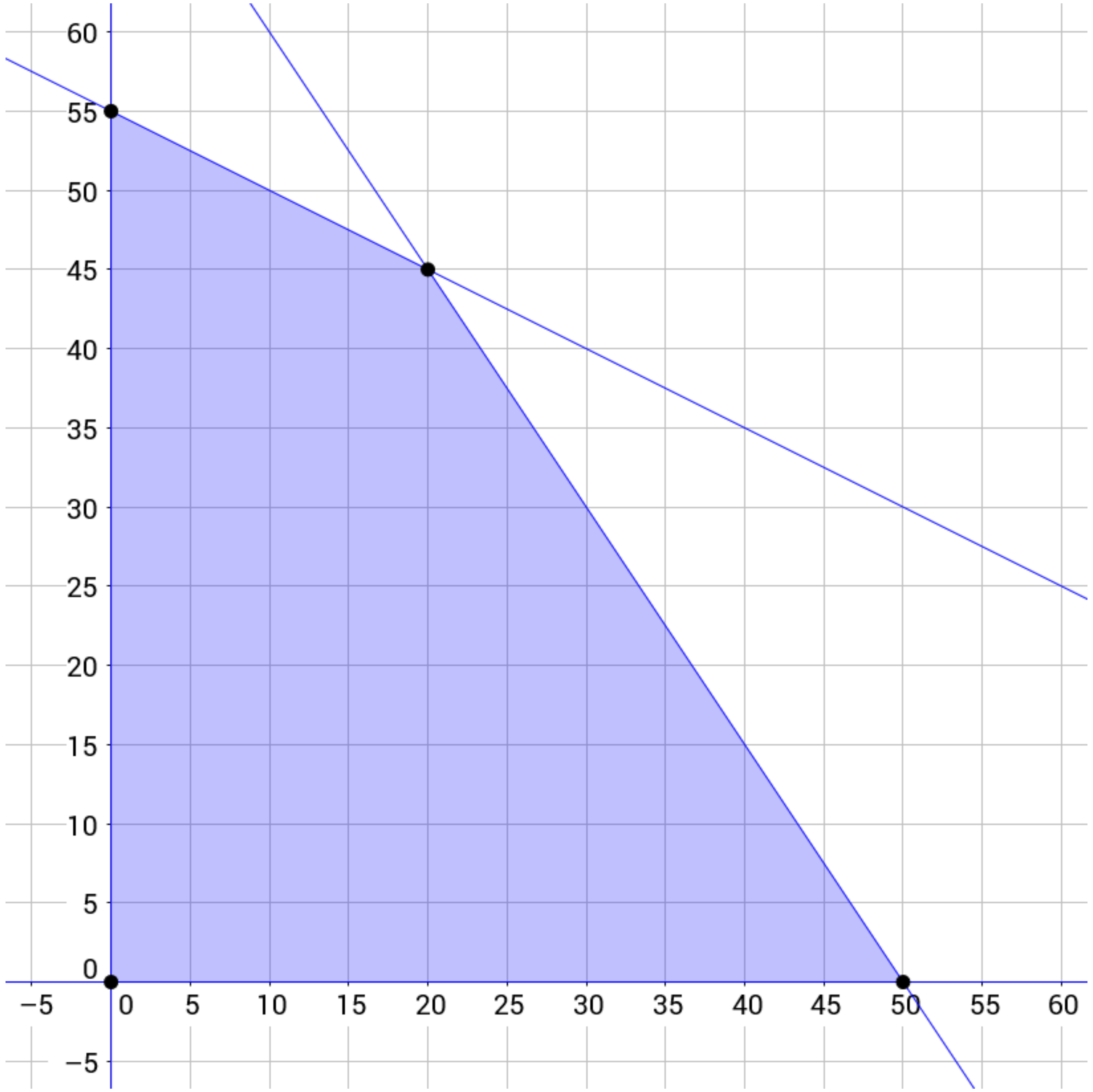
An example

Suppose that a farmer has a piece of farm land, say L km^2, to be planted with either wheat or barley or some combination of the two. The farmer has a limited amount of fertilizer, F kilograms, and pesticide, P kilograms. Every square kilometer of wheat requires F_1 kilograms of fertilizer and P_1 kilograms of pesticide, while every square kilometer of barley requires F_2 kilograms of fertilizer and P_2 kilograms of pesticide. Let S_1 be the selling price of wheat per square kilometer, and S_2 be the selling price of barley. If we denote the area of land planted with wheat and barley by x_1 and x_2 respectively, then profit can be maximized by choosing optimal values for x_1 and x_2. This problem can be expressed with the following linear programming problem in the standard form:

Maximize: $S_1 \cdot x_1 + S_2 \cdot x_2$	(maximize the revenue—revenue is the "objective function")
Subject to: $x_1 + x_2 \leq L$	(limit on total area)
$F_1 \cdot x_1 + F_2 \cdot x_2 \leq F$	(limit on fertilizer)
$P_1 \cdot x_1 + P_2 \cdot x_2 \leq P$	(limit on pesticide)
$x_1 \geq 0, x_2 \geq 0$	(cannot plant a negative area).

Feasible region

$$2x + 4y \leq 220$$
$$3x + 2y \leq 150$$
$$x \geq 0$$
$$y \geq 0$$



source

Property of LP

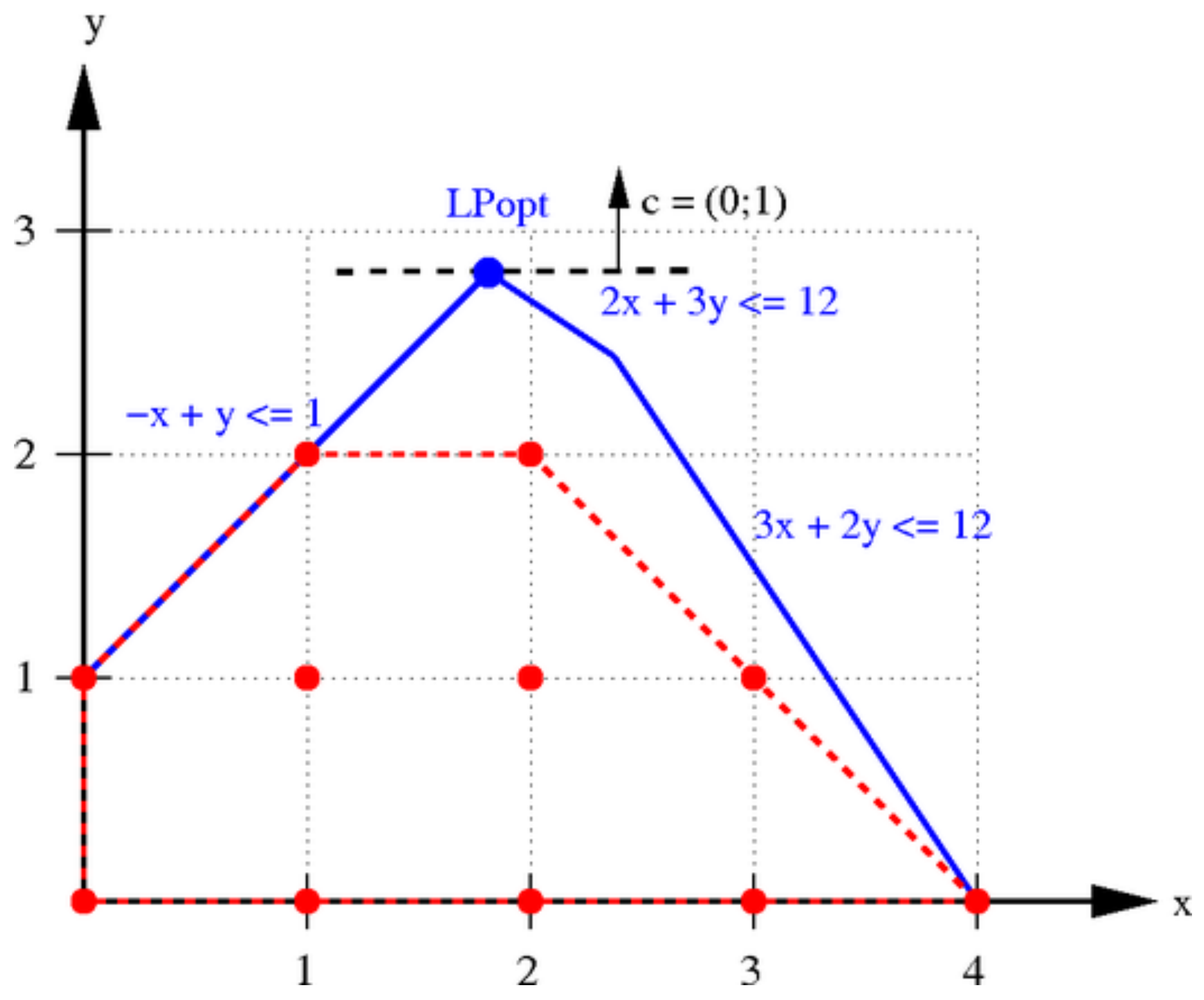
- The feasible region is convex

- If the feasible region is non-empty and bounded, then
  - optimal solutions exist, and
  - there is an optimal solution that is a corner point

  ➔ We only need to check the corner points

- The most well-known method is called the simplex method.

Simplex method



Simplex method:

- Start with a feasible solution, move to another one to increase f(x)

Integer linear programming

Integer programming

- IP is an active research area and there are still many unsolved problems.
- Many applications: scheduling, “diet” problems, NLP, …
- IP is more difficult to solve than LP.

- Methods:
  - Branch and Bound
  - Use LP relaxation

Example: Investment Decisions

Four investment options. Over 3 months, we want to invest up to 14, 12, and 15k.

requires an investment of \$5,000, \$8,000, and \$2,000 in month 1, 2, and 3, respectivey, and has a present value (a time-discounted value) of \$8,000; investment 2 requires \$7,000 in month 1 and \$10,000 in period 3, and has a value of \$11,000; investment 3 requires \$4,000 in period 2 and \$6,000 in period 3, and has a value of \$6,000; and investment 4 requires \$3,000, \$ 4,000, and \$5,000 and has a value of \$4,000. The corresponding integer program is

$$
\begin{array}{ll}
\text{Maximize} & 8x_1 + 11x_2 + 6x_3 + 4x_4 \\
\text{subject to} & 5x_1 + 7x_2 \qquad\qquad + 3x_4 \leq 14 \\
& 8x_1 \qquad\qquad + 4x_3 + 4x_4 \leq 12 \\
& 2x_1 + 10x_2 + 6x_3 + 4x_4 \leq 15 \\
& x_j \in \{0, 1\}.
\end{array}
$$

link

Example: Maximum Spanning Tree

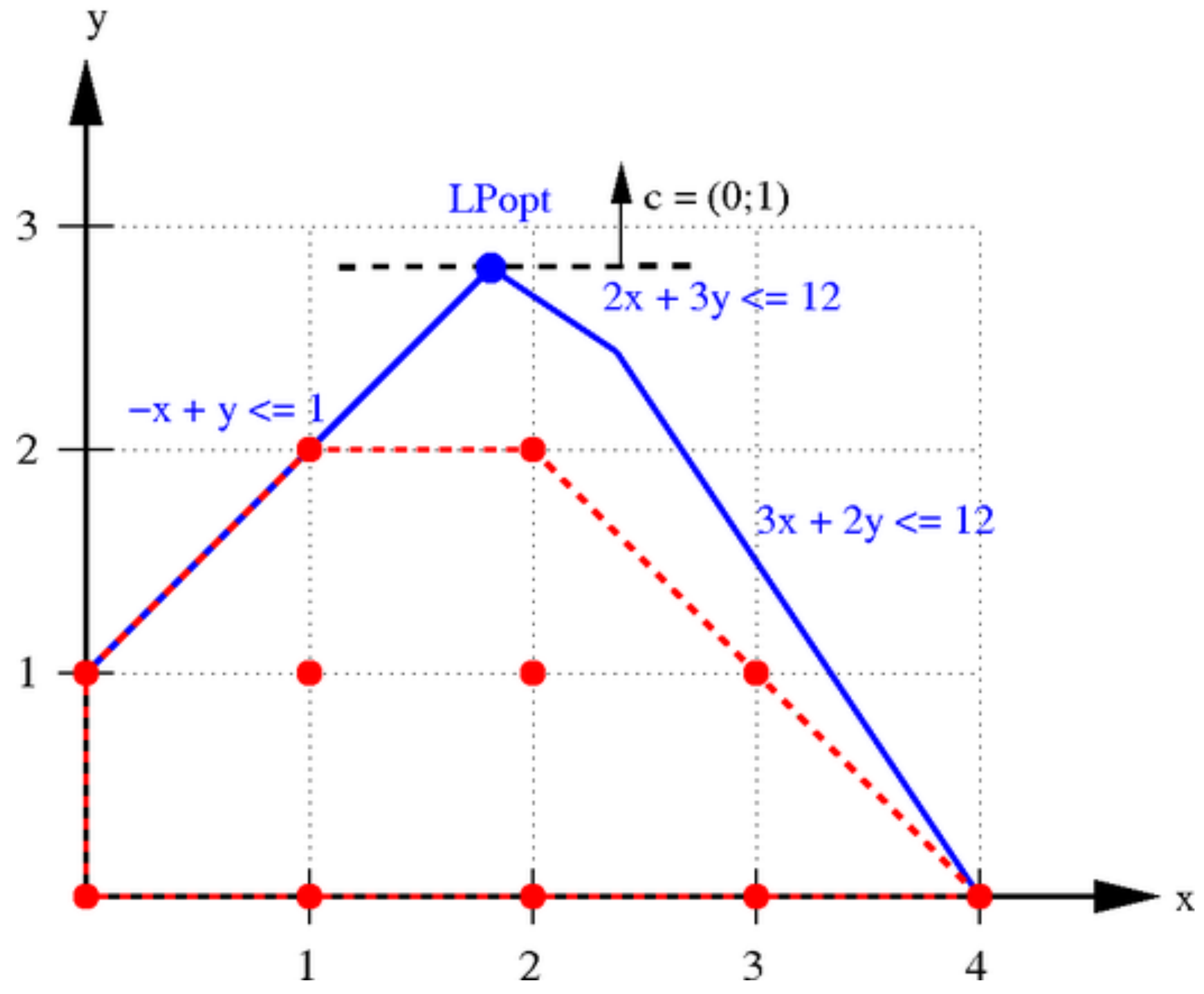
An approach to dependency parsing (see, e.g. 571 slides)

$$\max s(G) = \sum_{(w_1, w_2, l) \in G} s(w_1, w_2, l)$$

Constraint: G is a tree (no cycles)

More constraints possible: heads cannot have more than one outgoing label of each type

LP vs. ILP



Summary

- Optimization problems have many real-life applications.
- Common types: LP, IP, ILP, QP, Convex optimization problem
- LP is easy to solve; the most well-known method is the simplex method.
- IP is hard to resolve.
- QP and Convex optimization are used the most in our field.