IN4344 Advanced Algorithms

Lecture 2 – Some modeling techniques

Yuki Murakami

Delft University of Technology

11 September 2023

Types of optimization models

Recap:

- **Linear Programming (LP)** models have a **linear** objective function and **linear** constraints and **real** variables.

- **Integer Linear Programming (ILP)** models have a **linear** objective function and **linear** constraints and **integer** variables.

- **Mixed Integer Linear Programming (MILP)** models have a **linear** objective function and **linear** constraints and **integer and real** variables.

Why do we emphasize **linear** functions?

Modelling LPs and ILPs

Distinguish between the **parameters/input** (what you know) and the **decision variables** (what you want to know)!

Once you have specified the input, take the following steps:

- **Step 1: define the decision variables**

  What do we want to know?
  How can we describe a solution with variables?

- **Step 2: formulate the objective function**

  What do we want to maximize or minimize?
  Make sure that the objective function is linear! And that there is only one!

- **Step 3: formulate the constraints**

  What do solutions need to satisfy?
  Make sure the constraints are linear!

Combinatorial optimization problems

An important type of (M)ILPs is a **Combinatorial Optimization Problem** (COP).

Example of a COP: The subset sum problem

$N = \{1, \dots, n\}$, set of "items".

Each item $j \in N$ has value c_j, and weight $w_j \geq 0$.

Given is a number b. $\mathcal{S}$ consists of all subsets $S \subseteq N$ such that

$$\sum_{j \in S} w_j = b.$$

Problem

$$\max\{c(S) \mid S \in \mathcal{S}\}$$

$$N = \{1, \dots, 6\},\ c = (5, 3, 4, 2, 8, 10)$$

$$w = (4, 1, 8, 5, 6, 12),\ b = 12$$

What is the collection $\mathcal{S}$ of feasible subsets?

Example of a COP: The subset sum problem

$$N = \{1, \dots, 6\},\ c = (5, 3, 4, 2, 8, 10)$$

$$w = (4, 1, 8, 5, 6, 12),\ b = 12$$

What is the collection $\mathcal{S}$ of feasible subsets?

$$\mathcal{S} = \{\{1, 3\},\ \{2, 4, 5\},\ \{6\}\}$$

Here we have:

$$\begin{aligned} c(S_1) &= c(\{1, 3\}) = 9, \\ c(S_2) &= 13, \\ c(S_3) &= 10 \end{aligned}$$

So, subset $S_2 = \{2, 4, 5\}$ is the optimal one.
Notice, $\mathcal{S}$ can contain **exponentially many subsets**!

Example of a COP: The subset sum problem

$N = \{1, \ldots, n\}$, set of "items".

Each item $j \in N$ has value c_j, and weight $w_j \geq 0$.

Given is a number b. $\mathcal{S}$ consists of all subsets $S \subseteq N$ such that

$$\sum_{j \in S} w_j = b.$$

Problem

$$\max\{c(S) \mid S \in \mathcal{S}\}$$

Let's try and formulate this as an ILP.

Formulating subset sum as an ILP

We can easily formulate the subset sum problem as an ILP;
we then **define the feasible subsets implicitly** using linear constraints.

- **Step 1: define the decision variables**
  Let

$$x_j = \begin{cases} 1, & \text{if item } j \text{ is included in the subset} \\ 0, & \text{otherwise.} \end{cases}$$

- **Step 2: formulate the objective function**

$$\max z = \sum_{j \in N} c_j x_j$$

- **Step 3: formulate the constraints**

$$\sum_{j \in N} w_j x_j = b,$$

$$x_j \in \{0, 1\}, \quad \text{for all } j \in N$$

Combinatorial optimization problems

Let's generalize such **Combinatorial Optimization Problems**.

Let

$$N = \{1, \dots, n\}$$

be a finite ground set of elements, and

$$c = (c_1, \dots, c_n)$$

an n-vector. For $S \subseteq N$ define

$$c(S) = \sum_{j \in S} c_j.$$

Suppose we are given a collection $\mathcal{S}$ of feasible subsets.

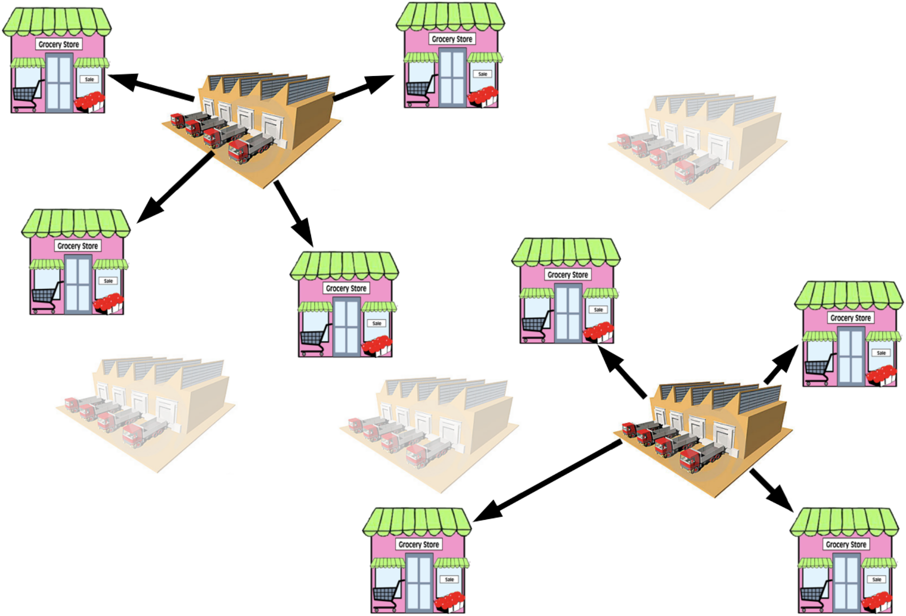
Problem

The following is a **Combinatorial Optimization Problem** *(COP): Find the highest(lowest)-value subset among all feasible subsets.*

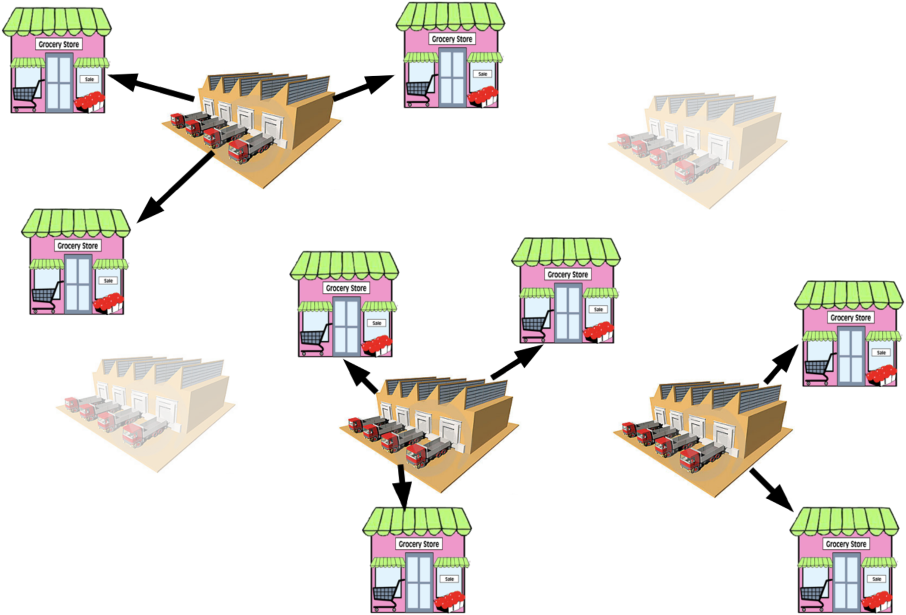
$$\max \mathit{or} \ \min\{c(S) \mid S \in \mathcal{S}\}$$

A more “complicated” COP: Facility Location

Formulating ILPs: Facility Location



A more "complicated" COP: Facility Location



A more "complicated" COP: Facility Location

Problem

- *n stores need to be supplied regularly.*
- *m locations are available where distribution centers (facilities) can be established.*
- *Select locations for the facilities such that the total costs are as small as possible:*
  - *f_i is the cost of establishing a facility at location i*
  - *c_{ij} is the cost of supplying store j from facility location i.*

Here, the ground set N of elements is the set of facilities.
The collection of feasible subsets $\mathcal{S}$ contains **all** subsets of N except the empty set.

Formulate the facility location problem as an ILP

- **Step 1: define the decision variables**

  Here we have **two types of decisions** that need to be made:

  - At which locations should facilities be established?
  - To which open facility should the various stores be assigned?

  Let

$$y_i = \begin{cases} 1, & \text{if facility at location } i \text{ is established} \\ 0, & \text{otherwise.} \end{cases}$$

$$x_{ij} = \begin{cases} 1, & \text{if a facility at location } i \text{ supplies store } j \\ 0, & \text{otherwise.} \end{cases}$$

- **Step 2: formulate the objective function**

$$\min z = \sum_{i=1}^{m} f_i y_i + \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

Formulate the facility location problem as an ILP

- **Step 3: formulate the constraints**

  First, make sure that each store is assigned to a facility:

  $$\sum_{i=1}^{m} x_{ij} = 1, \quad \text{for all } j = 1, \ldots, n.$$

  Next, make sure that the the **facility** that store j is assigned to is **established**!

  If we do not force this, all y_i-variables will take value 0, and we won't count the fixed costs f_i!

  $$x_{ij} - y_i \leq 0, \quad \text{for all } i = 1, \ldots, m, \; j = 1, \ldots, n.$$

Formulate the facility location problem as an ILP

- **Step 3: formulate the constraints**

$$
\begin{aligned}
\sum_{i=1}^{m} x_{ij} &= 1, \quad \text{for all } j = 1, \ldots, n, \quad (1) \\
x_{ij} - y_i &\leq 0, \quad \text{for all } i = 1, \ldots, m, \ j = 1, \ldots, n. \quad (2)
\end{aligned}
$$

And, then the "trivial constraints":

$$
\begin{aligned}
x_{ij} &\in \{0, 1\}, \quad \text{for all } i = 1, \ldots, m, \ j = 1, \ldots, n, \\
y_i &\in \{0, 1\}, \quad \text{for all } i = 1, \ldots, m.
\end{aligned}
$$

Notice!

- If x_{ij} takes value 1, then Constraint (2) "forces" y_i to take value 1 as well.
- If y_i takes value 0, x_{ij} is also forced to take value 0.
- If all $f_i > 0$ we will not establish a facility unless a store is assigned to it.

Formulate the facility location problem as a MILP

If we allow for a store to be supplied from more than one facility, we can make x_{ij} a real-valued variable. The model then becomes:

$$
\begin{aligned}
\min z = \sum_{i=1}^{m} f_i y_i \quad &+ \quad \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\
\text{s.t. } \sum_{i=1}^{m} x_{ij} \quad &= \quad 1, \quad \text{for all } j = 1, \ldots, n, \\
x_{ij} - y_i \quad &\leq \quad 0, \;\; \text{for all } i = 1, \ldots, m, \;\; j = 1, \ldots, n, \\
{\color{red} 0 \leq x_{ij}} \quad &{\color{red}\leq} \quad {\color{red}1}, \;\; \text{for all } i = 1, \ldots, m, \;\; j = 1, \ldots, n, \\
y_i \quad &\in \quad \{0, 1\}, \quad \text{for all } i = 1, \ldots, m.
\end{aligned}
$$

Then, x_{ij} models the **proportion** of store j's demand is supplied from facility i.

Box selection problem

Application at bol.com (MSc project)

- Determine the dimensions of boxes to be used in a warehouse.
- The number of different boxes is fixed as p (currently 15).
- Find an optimal solution on historical data.
- Minimize either:
  - the total volume of the boxes sent; or
  - the total amount of cardboard used.

Box selection problem: approach

Formulate as (a variant of) the **facility location** problem:

- define a large set of candidate boxes, these are the **locations**;
- the historical orders are the **stores**;
- the cost of supplying store (order) i from a location (candidate box) j is:
  - ∞ if the order does not fit in the box;
  - otherwise, the box volume, or cardboard usage.
- the cost c_{ij} of opening a facility is 0, but the number of facilities to open is fixed as p.

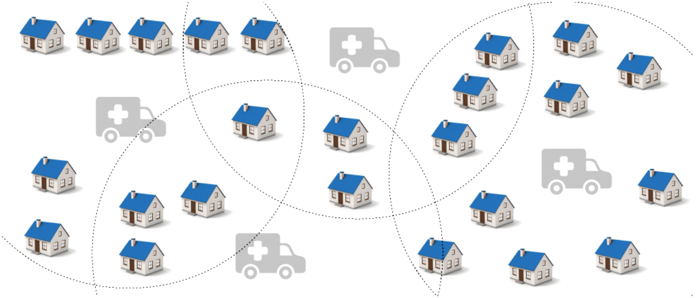
Box selection problem: ILP

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\
\text{s.t.} \ \sum_{i=1}^{m} x_{ij} &= 1, \quad \text{for all } j = 1, \ldots, n, \\
x_{ij} - y_i &\leq 0, \ \text{for all } i = 1, \ldots, m, \ j = 1, \ldots, n, \\
\sum_{i=1}^{m} y_i &= p \\
x_{ij} &\in \{0,1\}, \ \text{for all } i = 1, \ldots, m, \ j = 1, \ldots, n, \\
y_i &\in \{0,1\}, \quad \text{for all } i = 1, \ldots, m.
\end{aligned}
$$

The Ambulance Covering Problem: modeling Max Min

Problem

- *There are p ambulances, n locations (stations) and m houses.*
- *The travel time from location i to house j is t_{ij}.*
- *The* **coverage** *of a house is the number of ambulances that is located within r minutes travel time.*
- *Locate the ambulances such that the* **smallest** *coverage of a house is as* **large** *as possible.*



The Ambulance Covering Problem: modeling Max Min

Problem

- *There are p ambulances, n locations (stations) and m houses.*
- *The travel time from location i to house j is t_{ij}.*
- *The* **coverage** *of a house is the number of ambulances that is located within r minutes travel time.*
- *Locate the ambulances such that the* **smallest** *coverage of a house is as* **large** *as possible.*

INTERMEZZO

How to model Max Min problems

Question

*This is **not** a correct LP formulation. How can we correct it?*

$$\begin{aligned} \max \quad & \min\{x, y\} \\ s.t. \quad & 2x + 3y \geq 5 \\ & 3x - 4y \leq 7 \\ & x, y \geq 0 \end{aligned}$$

Answer

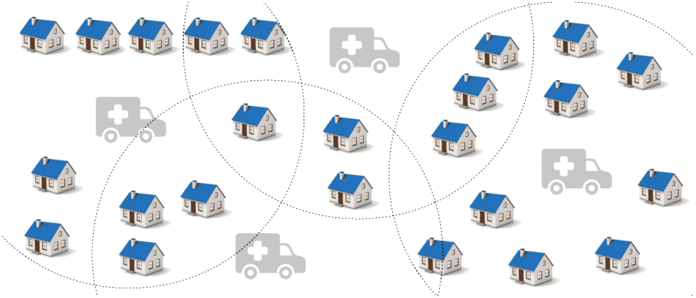
$$\begin{aligned} \max \quad & z \\ s.t. \quad & x \geq z \;\; (\text{ or } x - z \geq 0) \\ & y \geq z \;\; (\text{ or } y - z \geq 0) \\ & 2x + 3y \geq 5 \\ & 3x - 4y \leq 7 \\ & x, y, z \geq 0 \end{aligned}$$

with z a new decision variable.

Back to the Ambulance Covering Problem

Problem

- *There are p ambulances, n locations (stations) and m houses.*
- *The travel time from location i to house j is t_{ij}.*
- *The* **coverage** *of a house is the number of ambulances that is located within r minutes travel time.*
- *Locate the ambulances such that the* **smallest** *coverage of a house is as* **large** *as possible.*



Back to the Ambulance Covering Problem

Problem

- *There are p ambulances, n locations (stations) and m houses.*
- *The travel time from location i to house j is t_{ij}.*
- *The* **coverage** *of a house is the number of ambulances that is located within r minutes travel time.*
- *Locate the ambulances such that the* **smallest** *coverage of a house is as* **large** *as possible.*

- **Step 1: define the decision variables**

  $x_i =$ the number of ambulances placed at location i.
  $y_j =$ number of ambulances covering house j (coverage of j).
  $z =$ smallest coverage.

- **Step 2: formulate the objective function**

$$\max z$$

Back to the Ambulance Covering Problem

- **Step 2: formulate the objective function**

$$\max z$$

- **Step 3: formulate the constraints**

$$\begin{aligned}
y_j &\geq z \quad \text{for all } j = 1, \ldots, m \\
\sum_{\{i \mid t_{ij} \leq r\}} x_i &\geq y_j \quad \text{for all } j = 1, \ldots, m \\
\sum_{i=1}^{n} x_i &\leq p \\
x_i &\in \mathbb{Z}_{\geq 0} \quad \text{for all } i = 1, \ldots, n \\
y_j &\in \mathbb{Z}_{\geq 0} \quad \text{for all } j = 1, \ldots, m
\end{aligned}$$

The notation $x_i \in \mathbb{Z}_{\geq 0}$ means $x_i \geq 0,\ x_i \in \mathbb{Z}$

Point to ponder: Shouldn't the first two constraints be equalities?

Modeling **or**

Question

This is **not** *a correct ILP formulation. How can we correct it?*

$$
\begin{aligned}
\min \quad & 4x + 5y \\
s.t. \quad & 2x + 3y \leq 5 \quad \textbf{or} \quad 3x - 4y \leq 7 \\
& x, y \in \mathbb{Z}_{\geq 0}
\end{aligned}
$$

Answer

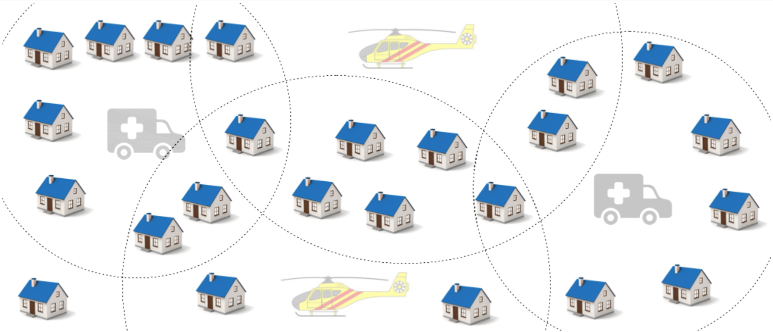
$$
\begin{aligned}
\min \quad & 4x + 5y \\
s.t. \quad & 2x + 3y \leq 5 + Mz \\
& 3x - 4y \leq 7 + M(1 - z) \\
& x, y \in \mathbb{Z}_{\geq 0} \\
& z \in \{0, 1\}
\end{aligned}
$$

with z a new binary decision variable and M a large constant.

Modeling **or**

Problem

- *An ambulance costs c_a and a helicopter c_h. There are S stations.*
- *The driving and flying time from stations i to house j are d_{ij} en f_{ij}.*
- *A house is* **covered** *if there are a ambulances within driving time D* **or** *h helicopters within flying time F.*
- *Make sure that all H houses are covered and minimize the total costs.*



Modeling **or**

Problem

- *An ambulance costs c_a and a helicopter c_h. There are S stations.*
- *The driving and flying time from stations i to house j are d_{ij} en f_{ij}.*
- *A house is* **covered** *if there are a ambulances within driving time D* **or** *h helicopters within flying time F.*
- *Make sure that all H houses are covered and minimize the total costs.*

Do as an exercise!

Modeling integers by binary variables

Problem

Suppose we want to produce an item in certain batch sizes, say 0, 5, 15, 30, and 50. How can we model that linearly? Or, equivalently, how can we model that $x \in \{0, 5, 10, 25, 50\}$ *with linear constraints?*

Here we have five possible integer values that x can take.

- **Step 1: define the decision variables**

  Let

  $$y_i = \begin{cases} 1, & \text{if positive batch size } i \text{ is chosen} \\ 0, & \text{otherwise.} \end{cases}$$

- **Step 2: formulate the constraints**

  $$\begin{aligned} \sum_{i=1}^{4} y_i &\leq 1 \\ y_i &\in \{0, 1\}, \quad \text{for all } i = 1, \ldots, 4. \end{aligned}$$

Modeling connectivity in a graph

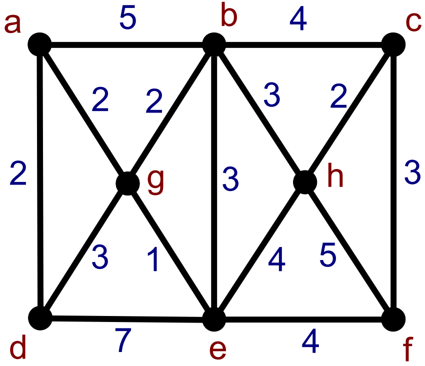
Back to TSP from the previous lecture!

Problem

Traveling Salesman Problem (TSP)

Given: *a graph* $G = (V, E)$ *and a length* $\ell(e)$ *for each edge* $e \in E$

Find: *a cycle that visits each vertex of the graph exactly once and is as short as possible.*



Modeling connectivity in a graph

Problem

Traveling Salesman Problem (TSP)
Given: *a graph* $G = (V, E)$ *and a length* $\ell(e)$ *for each edge* $e \in E$
Find: *a cycle that visits each vertex of the graph exactly once and is as short as possible.*

- **Step 1: define the decision variables**

Let

$$x_e = \begin{cases} 1, & \text{edge } e \in E \text{ is included in the TSP-tour,} \\ 0, & \text{otherwise.} \end{cases}$$

- **Step 2: formulate the objective function**

$$\min \sum_{e \in E} \ell(e) x_e$$

Modeling connectivity in a graph

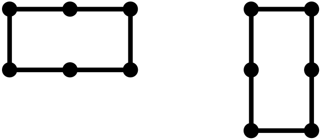
- **Step 3: formulate the constraints**

1. Make sure that the traveling salesman enters and leaves each city precisely once:

$$\sum_{\text{edges } e \text{ that are incident (connected) to } v} x_e = 2 \qquad \text{for each vertex } v \in V$$

2.
$$x_e \in \{0,1\}, \quad \text{for all } e \in E$$

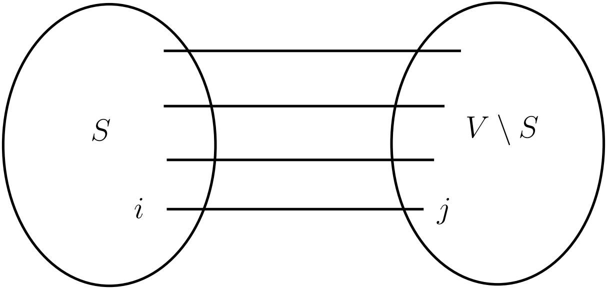
3. And, we need to exclude "subtours"!



Modeling connectivity in a graph

How to exclude subtours?
What we need to model:
If we take any subset of the vertices $S \subset V$ (not the empty set and not V itself), then this set should be connected to the vertices that are not in S.

$$\sum_{\{e=\{i,j\}\in E \mid i\in S,\ j\in V\setminus S\}} x_e \geq 2 \qquad \text{for all } S \subset V, S \neq \emptyset, S \neq V$$



The full ILP for TSP

Decision variables:

$$x_e = \begin{cases} 1, & \text{edge } e \in E \text{ is included in the TSP-tour,} \\ 0, & \text{otherwise.} \end{cases}$$

ILP:

$$\begin{aligned}
\min \quad & \sum_{e \in E} \ell(e) x_e \\
\text{s.t.} \quad & \sum_{\{e \in E | v \in E\}} x_e = 2 && \text{for each vertex } v \in V \\
& \sum_{\{e=\{i,j\} \in E | i \in S,\ j \in V \setminus S\}} x_e \geq 2 && \text{for all } S \subset V, S \neq \emptyset, S \neq V \\
& x_e \in \{0, 1\} && \text{for all } e \in E
\end{aligned}$$

Modeling connectivity in a graph

$$\sum_{\{e=\{i,j\}\in E \mid i\in S,\ j\in V\setminus S\}} x_e \geq 2 \qquad \text{for all } S \subset V, S \neq \emptyset, S \neq V$$

Question

What could be an issue with this formulation?

Answer

There are **exponentially many** *subsets $S \subset V$.*

Solution

Do not add all of them a priori! Add them sequentially when needed.
One can prove that this can be done efficiently!

Exercise

Give an ILP formulation of the following graph problem.

MINIMUM STEINER TREE
Given: a graph $G = (V, E)$, a length $\ell(e) > 0$ of each edge $e \in E$ and a set $T \subseteq V$ of *terminals*.
Find: a subgraph of G that is a tree and contains all terminals from T, such that the total length of all edges in the tree is minimized.

(A *tree* is a connected graph without cycles.)

To Do

Do exercises 2.1, 2.4, 2.6 in the Syllabus.
Read Chapter 3.1 in the Syllabus to catch up on definitions.