Algorithmic Bioinformatics II

Homework Assignment I

Xiheng He

Januar 2022

Topic: Optimization

Exercise O1 (Linear Programming, 10P):

Calculate the maximum of the function

$$y = -1.5x_1 - x_2$$

given the following constraints:

$$x_1 + x_2 \geq 35 \tag{1}$$

$$x_1 \leq 109 \tag{2}$$

$$23 \leq x_2 \leq 42 \tag{3}$$

$$x_2 \leq 27 + 2x_1 \tag{4}$$

$$\forall i : x_i \geq 0 \tag{5}$$

Draft the solution space first by hand, then use, for instance, the R-package **Rglpk** for calculating the optimal values of x_1 and x_2.

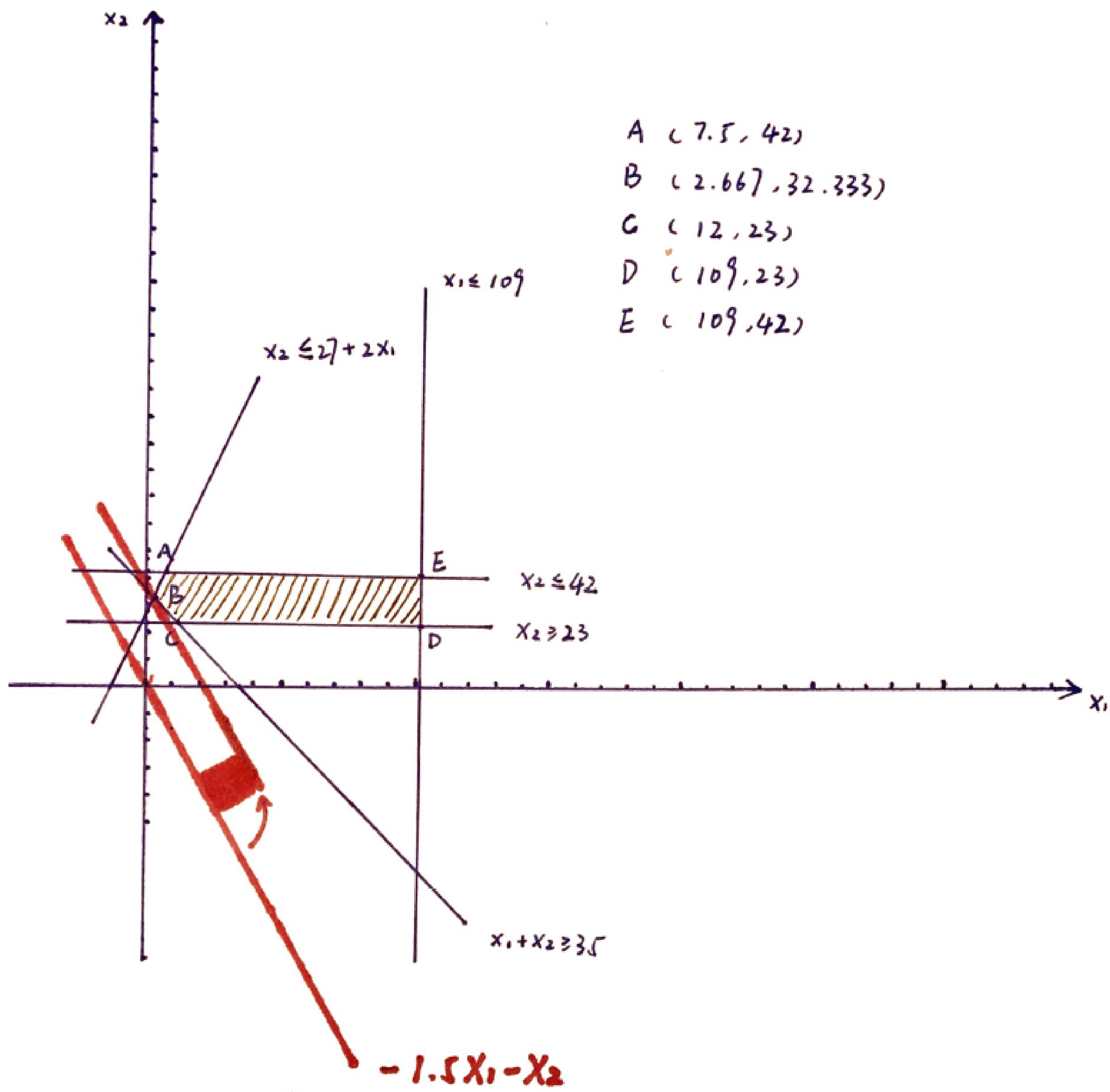

Figure 1: Solution space

As shown in Figure 1, the optimal values of x_1 and x_2 is given by point B and the schaded area which is enclosed by the constraints is our solution space. Red line is the function to be maximized which means that x_1 and x_2 should be minimized. Move the function line in parallel to the point where it first intersects with solution space, which is the optimal solution of the problem:

$B(2.667, 32.333) \Longrightarrow x_1 = 2.667, x_2 = 32.333 \Longrightarrow y = -36.3335$

```
library(slam)
library(Rglpk)

# Usage

# Rglpk_solve_LP(obj, mat, dir, rhs, bounds = NULL, types = NULL, max = FALSE, control
    = list(), ...)

obj <- c(-1.5, -1)
mat <- matrix(c(1, 1, 2, 0, 0, 1, 0, 1, 0, -1, 1, 1, 0, 1), nrow = 7)
dir <- c(">=", "<=", ">=", ">=", "<=", ">=", ">=")
rhs <- c(35, 109, -27, 23, 42, 0, 0)
Rglpk_solve_LP(obj, mat, dir, rhs, bound = NULL, max = TRUE)

# optimum

# [1] -36.33333

# solution

# [1]  2.666667 32.333333
```

Topic: Combinatorial Optimization, Matroids, Greedy-Algorithm

Exercise B1 (Matroids 1, 10P):

Let E be a set with $|E| = n$ elements. Prove or disprove that (E, F) is a matroid.

(a) For a given p in $0 \leq p \leq n$, let F be the set which contains all subsets of E with p or less elements.

Definition 0.1. (E, F) is a independent system, if:

- $(M1)$ $\emptyset \in F$
- $(M2)$ $x \subseteq y \in F \Longrightarrow x \in F$

Definition 0.2. (E, F) is a matroid, if:

- (E, F) is independent system and,
- (M_3) $x, y \in F \wedge |x| > |y| \Longrightarrow \exists \chi \in x \setminus y, y \cup \chi \in F$

$$M_1 : 0 \leq p \leq n \wedge \forall S \in F, |S| \leq p \Longrightarrow \emptyset \in F$$

$$M_2 : \forall x, y, x \subseteq y \in F \Longrightarrow 0 \leq |x| \leq |y| \leq p \Longrightarrow x \in F$$

$$M_1 \wedge M_2 \Longrightarrow (E, F) \text{ is a independent system}$$

$$M_3 : \exists x, y \in F \wedge \exists z, x = y \cup \{z\} \wedge 0 \leq |y + 1| \leq |x| \leq p \Longrightarrow x, y \in F \wedge |x| > |y|$$

$$\Longrightarrow \exists z \in x \setminus y, y \cup \{z\} \in F$$

$$M_1 \wedge M_2 \wedge M_3 \Longrightarrow (E, F) \text{ is a matroid}$$

(b) For a given $x, y \in E$ let F be the set of all subsets of E which do not contain x and y.

$$M_1 : x, y \notin \emptyset \Longrightarrow \emptyset \in F$$

$$M_2 : \forall m, x \notin m \wedge y \notin m \Longrightarrow \forall n, m \subseteq n \in F \Longrightarrow m \in F$$

$$M_1 \wedge M_2 \Longrightarrow (E, F) \text{ is a independent system}$$

For M_3 we have to discuss under following conditions:

- $$a \in x \wedge a \in y \vee b \in x \wedge b \in y \Longrightarrow \exists z \neq a, b \wedge z \in x \setminus y, y \cup \{z\} \in F \Longrightarrow M_3$$

- $$a \in x \wedge b \in y \vee a \in y \wedge b \in x \Longrightarrow \exists z \neq a, b \wedge z \in x \setminus y, y \cup \{z\} \in F \Longrightarrow M_3$$

- $$a \notin x \wedge b \notin x \Longrightarrow \exists z \neq a, b \wedge z \in x \setminus y, y \cup \{z\} \in F \Longrightarrow M_3$$

- $$a \in x \wedge a, b \notin y \Longrightarrow z := a, z \in x \setminus y, y \cup \{z\} \in F \Longrightarrow M_3$$

- $$b \in x \wedge a, b \notin y \Longrightarrow z := b, z \in x \setminus y, y \cup \{z\} \in F \Longrightarrow M_3$$

Thus, (E, F) is a matroid as M_1, M_2, M_3 are satisfied.

Exercise B2 (Combinatorial optimization problems, 5P):

Consider the following combinatorial optimization problems (kOPs from the lecture) and the corresponding set systems. Define the independent sets F formally and show, that these are independent systems, matroids or greedoids. Calculate the basis of these systems. For the following exercise let $G = (V, E)$ be a graph with nodes V and edges E.

(a) *Bipartite Matching:* $V = U \cup W, U = \{u_1, u_2, \dots, u_n\}$. Set of subsets X of W, which can be matched with $\{u_1, \dots, u_{|X|}\}$.

- obviously, F is indepent set of edges E and $\emptyset \in F$.
- There is a subset Y of W which can be machted with $\{u_1, \dots, u_{|Y|}\}$, and $\forall y \in F \wedge y \subset E$, st. its edges come from vertices $\{u_1, \dots, u_{|Y|}\} \cup Y$. As $x \subset y \in F$, we denote the vertices of x as X, X is a subset of W which can also be matched with $\{u_1, \dots, u_{|X|}\}$, thus $x \subseteq y \in F \Longrightarrow x \in F$. $G = (V, E)$ is a independent system.

(b) *Spanning tree:* Subset of edges without cycle.

- obviously, F is indepent set of all edges E and $\emptyset \in F$.
- $\forall x \subset y \in F \Longrightarrow x \in F$. If y is independent set and $y \in F$, then $x \in F$ as x is subset of y without cycle.
- $\exists x, y \in F \wedge |x| > |y| \Longrightarrow \exists z \in x \setminus y, y \cup \{z\} \in F$. If x has more edges than y, then there exists always a edge $z \in x \setminus y$ st. $y \cup \{z\}$ is subset of E without cycle as there is no cycle in both x and y. Thus, $G = (E, F)$ is a matroid.
- As $G = (E, F)$ is a matroid, By definition, the basis of a matroid is F.

(c) *Travelling-Salesman-Problem:* Subset of edges which are part of a hamiltonian cycle.

- obviously, F is indepent set of all edges E and $\emptyset \in F$.
- For set x and y: $x \subset y \subseteq E \wedge y \in F$ and x is a subset of y which is a part of a hamiltonian cycle st. $x \in F$. obviously that $x \subset y \wedge y \in F \Longrightarrow x \in F \Longrightarrow G = (E, F)$ is a independent system.

(d) *Matching:* subset of edges which are not pairwise adjacent.

- $\emptyset \in F$ as it is not pairwise adjacent.
- $y \in F$ is a subset of edges which are not pairwise adjacent, then for $x \subset y$, x is also a subset of edges which are not pairwise adjacent. Thus, $x \subseteq y \in F \implies x \in F$.
- M_3 for Matroid does not hold, as we can not ensure that a new edge $z \in x \setminus y$, all edges in $y \cup \{z\}$ are not pairwise adjacent. Thus, $G = (V, E)$ is a independent system.

Exercise B3 (0-1 Knapsack-Implementation, 5P):

Write programs to solve the 0-1 Knapsack-Problem with the optimal and greedy algorithm. Plot $\frac{C_G}{C_O}$ for a given amount of instances I against $|I|$ and n.

Please check my jupyter notebook.

Exercise B4 (Best-In Greedy, 5P):

The best-in greedy algorithm for (a) the 0-1 knapsack problem and (b) independence systems can result in arbitrarily bad solutions. Define *arbitrarily bad* and give an example while not forgetting to justify your answer.

Arbitrarily bad means the algorithm does not always find a good solution in a arbitrary case. In some cases, we could get a really absurd answer which leads to a false conclusion.

Consider the following example:

- $$\text{maximize } x_1 + kx_2$$
  $$\text{subject to } x_1 + kx_2 \leq k \text{ where } k \gg 1$$
  $$x_1, x_2 \in \{0, 1\}$$
- Greedy algorithm will calculate the value-weight ratio and get $\dfrac{c_1}{w_1} = 1 \leq \dfrac{c_2}{w_2} = 1$.
- Greedy alogrithm will give a solution $x_1 = 1, x_2 = 0$ which is not optimal $(x_1 = 0, x_2 = 1)$.

Exercise B5 (Matroide 2, 5P):

Let $M = (E, F)$ be a matroid.

(a) Let S be a subset of E. Show that $M' = (E \setminus S, F')$ with $F' = \{I | I \in F, I \subseteq E \setminus S\}$ is a matroid.

$$M_1 : M = (E, F) \text{ is a matroid} \implies M \text{ is a independent system} \implies \emptyset \in F \implies \emptyset \subset E \setminus S \implies \emptyset \in F'$$

$$M_2 : \exists x \subseteq y \in F \wedge x, y \subseteq E \setminus S \implies x \subseteq y \in F' \implies x \in F'$$

$$M_3 : \exists x, y \in F \wedge x \subseteq E \setminus S \wedge y \subset E \setminus S \implies x, y \in F' \wedge \exists z, z \in x \setminus y \wedge |x| > |y| \implies y \cup \{z\} \in F'$$

$$M_1 \wedge M_2 \wedge M_3 \implies M' \text{ is a matroid}$$

(b) The dual matroid $M^* = (E, F^*)$ is defined by $F^* = \{F \subseteq E | \exists_{BasisB} F \cap B = \emptyset\}$. Show that M^* is a matroid. Proof that $(E, F^{**}) = (E, F)$ holds.

Lemma 0.1 (**Steinitz Exchange Lemma**)**.** If A and B are distinct members of $\mathcal{B}$ and $a \in A \setminus B$, then there exists an element $b \in B \setminus A$ such that $(A \setminus \{a\}) \cup \{b\} \in \mathcal{B}$.

- $F \neq \emptyset \implies F^* = \{F \subseteq E | \exists_{BasisB} F \cap B = \emptyset\} \neq \emptyset$
- $\forall S_1, S_2 \in B \implies |S_1| = |S_2| \wedge \forall a \in S_1 \setminus S_2, \exists b \in S_2 \setminus S_1 \implies |S_1| = |S_2| = |S_1 \setminus \{a\} \cup \{b\}| \implies$ $(S_1 \setminus \{a\}) \cup \{b\} \in B$

Since $\forall S \in F, S \subseteq E$, and we know that $F^* = \{F \subseteq E | \exists_{BasisB} F \cap B = \emptyset\}$ and $\forall B \in F^*, B \subseteq E$ as M^* is defind as a matroid. Thus all elements from B and F compose set E st.

$F^* := \{E \setminus B', B' \in F\}$. This means that the basis of dual matroid M^* is the complement of the basis in the original matroid M.

Since complement is an involutotory function $(A^c)^c = A \implies (E, F^{**}) = (E, F)$ holds.

Exercise B6 (Integer-Knapsack, 5P):

n object types with values c_i and weights a_i are given for the Integer-Knapsack problem. The sum $\sum_{i=1}^{n} c_i x_i$ is to be maximized given the following constraint for the weight: $\sum_{n=1}^{n} a_i x_i \leq b$.

Consider now $n = 2$ object types with values $c = (2, 1)$ and weights $a = (1, 2)$. The given knapsack holds up to a weight of 10. As additional constraint a maximum of 6 pieces of object type 1 and at least one object of type 2 should be contained. Formulate the Integer Knapsack problem as an ILP problem. Define the matrix A and the vectors b and c. Draft the valid set as demonstrated in the lecture.

$$\mathcal{Z}_{ILP} = \max\{cx | ax \leq b, x \in Z_+^n\}$$

$$A = [1, 2] \in \mathcal{M}_{1\times n}$$

$$b \in \mathcal{M}_{n\times 1}, \sum_{i=1}^{n} b_i = 10$$

$$c = [2, 1] \in \mathcal{M}_{1\times n}$$

or:

Calculate the maximum of the function:

$$2x_1 + x_2$$

given the following constraints:

$$x_1 + 2x_2 \leq 10 \tag{6}$$

$$x_1 \leq 6 \tag{7}$$

$$x_2 \geq 1 \tag{8}$$

$$\forall i : x_i > 0 \tag{9}$$

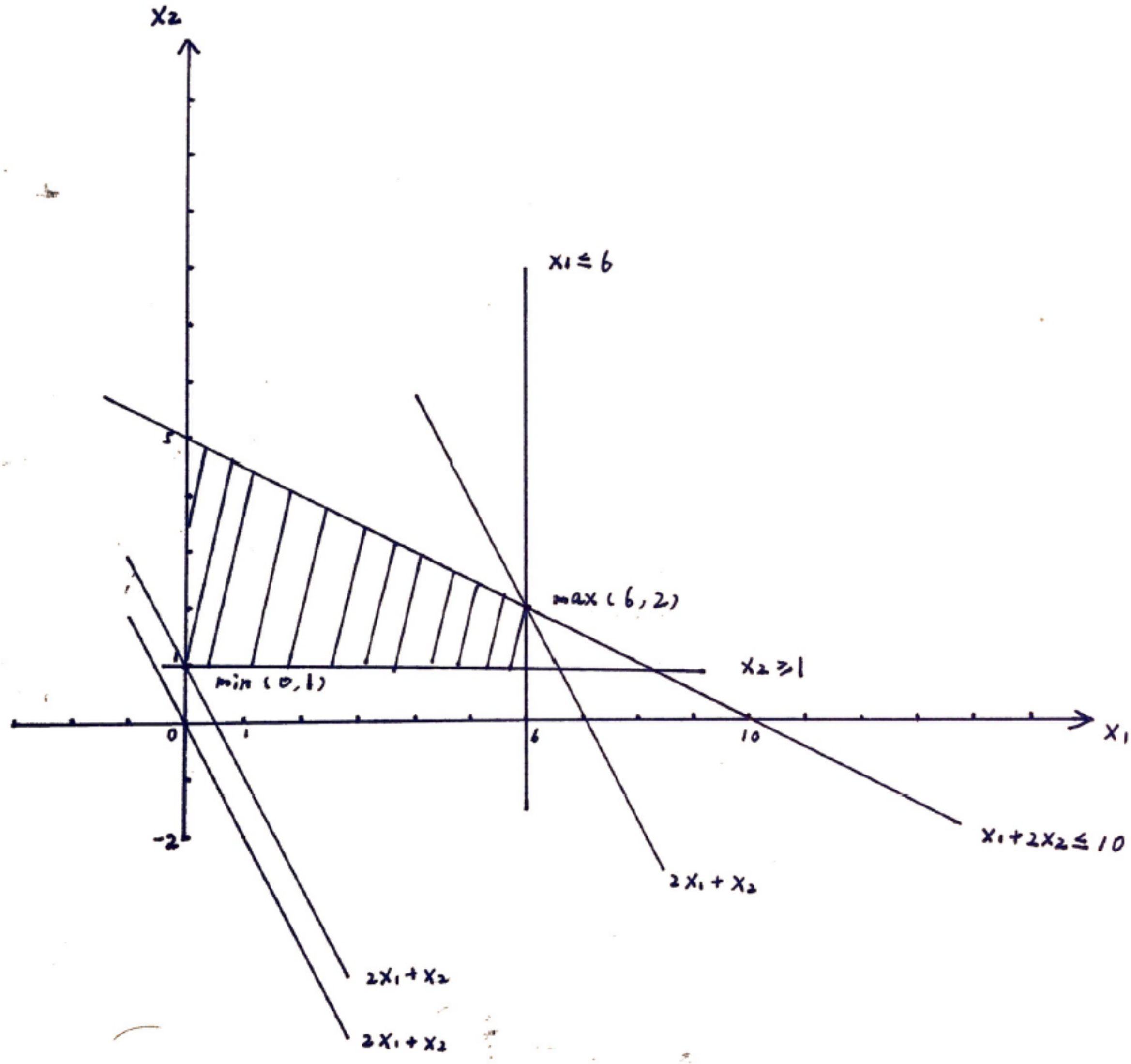

Figure 2: valid set for knapsack

Figure 2 shows optimal solution for Integer-Knapsack problem should be 14 with $x_1 = 6, x_2 = 2$.

Furthermore, if we convert Integer-Knapsack problem as a 0-1 knapsack problem as following:

$$\text{maximize} \sum_{i=1}^{n} c_i x_i \tag{10}$$

$$\text{subject to} \sum_{i} w_i x_i \leq 1 \tag{11}$$

$$\forall i, x_i \in \{0, 1\} \tag{12}$$

where we can take the (10) as maximize the indepedent number, (11) constraints that the independent set can and only contain one vertex from same edge and (12) indicates that a vertex can be placed in or not in the independent set. (Knapsack $\Longrightarrow$ Independent System $\Longrightarrow$ Indepedent Set)

Exercise B7 (Greedy Integer-Knapsack, 5P):

In the lecture the greedy algorithm for the knapsack problems has been introduced. For this the given object types (elements $e_i \in E, |E| = n$) are sorted by their values per weight (weight density), such that $\frac{c_1}{a_1} \geq \frac{c_2}{a_2} \geq \cdots \geq \frac{c_n}{a_n}$ holds.

let C_G be the greedy solution and C_{opt} the optimal solution. Show that the greedy algorithm of the Integer knapsack has an approximation factor of 2. In order to achieve this show that $C_G \geq \frac{1}{2} C_{opt}$ by finding an upper bound for C_{opt} and a lower bound for C_G.

Proof:

- C_G is the greedy solution for Integer-Knapsack, we split all elements according to their weights.

- let S be a set of elements where $S \subset E$. We denote w_i as the weight for every element in this set and W for total allowed weight where $\sum w_i \leq W$.

- assume there is a k-te element st. $e_k \in E \setminus S$ where $\sum_{e_i \in S} w_i + w_k > W$.

- We denote C_{opt} as optimal solution. As we know that greedy solution can not be better than optimal solution, we define $\sum_{i=1}^{j} c_i e_i$ is the greedy solution and $j = k - 1$, then:

$$\sum_{i=1}^{j} c_i e_i \leq C_{opt}$$

  .

- Since for k-te element we get $c_k e_k \leq C_{opt}$ and adding k-te element will exceed total allowed weight, thus:

$$C_{opt} \leq \sum_{i=1}^{j} c_i e_i \leq \sum_{e_i \in S} c_i e_i + c_k e_k \leq C_{opt} + C_{opt} = 2C_{opt} \quad \square$$

Therefore, the greedy algorithm of the Integer knapsack has an approximation factor of 2.