תכנון אלגוריתמים 25.10.2020

הרצאה 3: אלגוריתמים חמדניים

Textbook: Cormen, Leiserson, Rivest, Stein. Introduction to Algorithms.

מרצה: אורי שטמר

<u>באופן אינטואיטיבי:</u> אלג' חמדן הוא אלג' אשר בכל צעד מבצע את הפעולה שנראית לו הכי טובה באותו הרגע, ולא מתחרט על בחירותיו.

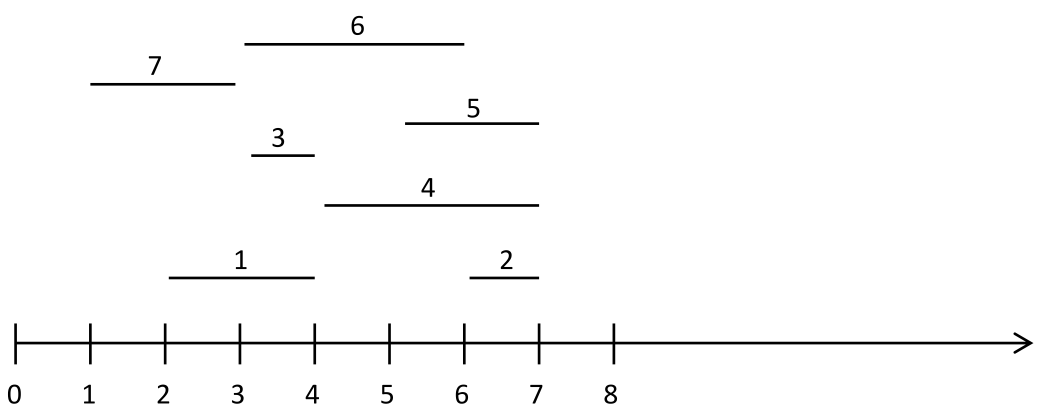
<u>בעיית הפעילויות</u>

<u>תאור לא פורמלי של הבעיה:</u> יש חדר הרצאות ויש פעילויות שרוצים לקיים בחדר. לכל פעילות יש זמן התחלה וזמן סיום. בחדר ההרצאות אפשר לקיים לכל היותר פעילות אחת בכל יחידת זמן.
<u>המטרה:</u> לשבץ מספר גדול ביותר של פעילויות.

<u>דוגמה:</u> נניח שיש לנו 7 פעילויות עם זמני התחלה וסיום כאלה:

i	1	2	3	4	5	6	7
זמן התחלה s_i	2	6	3	4	5	3	1
זמן סיום f_i	4	7	4	7	7	6	3

נניח שהפעילויות לא כוללות את קצוות אינטרוולי הזמן ונתאר אותם על ציר הזמן באופן הבא:



- 2,6,7 פתרון חוקי עם 3 פעילויות
- 7,3,4 פתרון חוקי עם 3 פעילויות
- אם היינו לוקחים את 1,4 לא היינו יכולים להרחיב אותו עם עוד פעילויות ולא היינו מגיעים למטרה

<u>הגדרה פורמלית של הבעיה:</u>
<u>קלט:</u> אוסף של n פעילויות. לכל פעילות i יש זמן התחלה s_i וזמן סיום f_i כך ש- $s_i < f_i$
<u>פתרון חוקי:</u> אוסף של פעילויות שלא חותכות זו את זו. כלומר, $I \subseteq \{1,2,\dots,n\}$ כך שלכל $i \neq j \in I$ מתקיים: או $f_i \leq s_j$ או $f_j \leq s_i$.
<u>יש למצוא:</u> פתרון חוקי I שמכיל מספר גדול ככל האפשר של פעילויות.

<u>אלגוריתם נאיבי לבעיה:</u> לכל תת קבוצה I נבדוק האם אוסף הפעילויות בה חוקי ונחזיר תת רבוצה חוקית גדולה ביותר.
<u>זמן ריצה:</u> האלגוריתם לא יעיל. עובר על 2^n קבוצות...

<u>תבנית לאלגוריתם חמדן לבעיה:</u>
<u>אתחול:</u> $G \leftarrow \emptyset$ (פתרון חלקי שיש לאלגוריתם בכל שלב)
$S \leftarrow \{1,2,\dots,n\}$ (קבוצת הקטעים שזרים לכל הקטעים ב- G)

<u>צעד:</u> כל עוד $S \neq \emptyset$:
- בחר $i \in S$ עפ"י <u>כלל בחירה כלשהו</u> ובצע $G \leftarrow G \cup \{i\}$
- הורד מ- S את כל הקטעים שנחתכים עם הקטע ה i (כולל הקטע ה i)

זוהי רק תבנית כמובן. מהו כלל הבחירה הנכון?
כלל זה יקבע האם האלגוריתם שלנו ימצא פתרון אופטימלי לכל קלט או שלא.
אבל עוד לפני שנחפש את כלל הבחירה, אנחנו כבר יכולים שים לב ש-

<u>אבחנה:</u> האלגוריתם מחזיר פתרון <u>חוקי</u> (לכל כלל בחירה)

<u>רעיונות לכלל הבחירה:</u>

1. **בחר קטע שנחתך עם מספר קטן ככל האפשר של קטעים**
האינטואיציה מאחורי הרעיון הזה: הוא פוסל מספר קטן ככל האפשר של קטעים

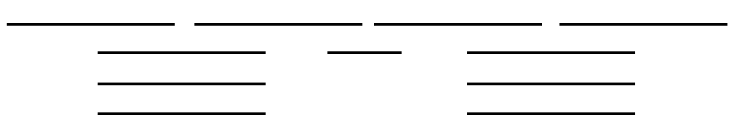
2. **בחר קטע קצר ביותר**
האינטואיציה: אינטואיטיבית זה יפסול לנו הכי מעט קטעים

3. **בחר את הקטע שמתחיל ראשון**
האינטואיציה: כך (אולי) ננצל הכי טוב את מרחב הזמן

4. **בחר את הקטע שמסתיים ראשון**
האינטואיציה: מפנים את החזר בזמן המינימלי האפשרי

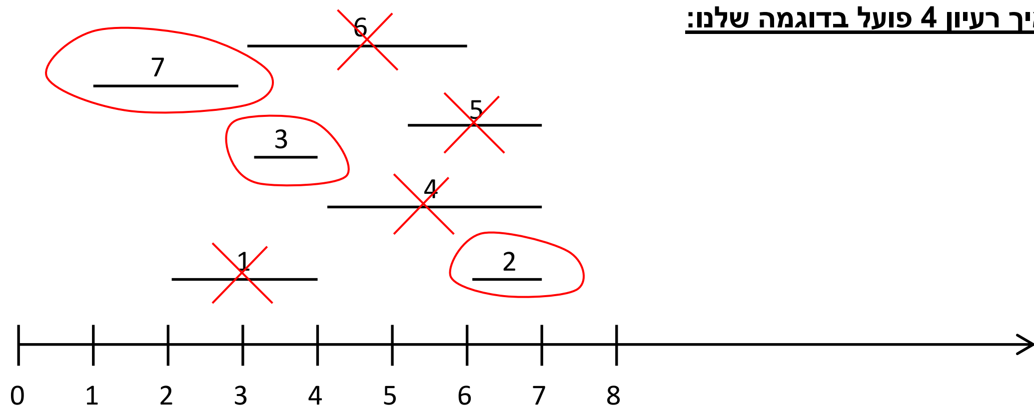
מכל ארבעת הרעיונות האלה, רק רעיון 4 יעבוד וכל שאר הרעיונות ייכשלו. לכן אנחנו נעבור עכשיו על הרעיונות, לשלושת הראשונים ניתן דוגמה נגדית ואת הרביעי נוכיח.

<u>דוגמה נגדית לרעיונות 1 ו- 2:</u>



<u>דוגמה נגדית לרעיון 3:</u>

נראה איך רעיון 4 פועל בדוגמה שלנו:



רעיון הוכחה של אלגוריתם חמדן:
אנחנו רוצים לומר ש- "בכל שלב קיים פתרון אופטימלי שמכיל את מה שעשינו".
כלומר, נוכיח שלא טעינו.

תהייה לנו טענה נשמרת:
בכל שלב של האלגוריתם אפשר להרחיב את הקבוצה G לפתרון אופטימלי

ההוכחה תהייה באינדוקציה

הערה: לא מספיק להראות שלא טועים משום שגם אלגוריתם שלא עושה כלום לא טועה...

אלגוריתם חמדן עם כלל הבחירה הנכון:
אתחול: $G \leftarrow \emptyset$
$S \leftarrow \{1,2,\dots,n\}$

צעד: כל עוד $S \neq \emptyset$:
- בחר את הפעילות $i \in S$ שמסתיימת ראשונה
- $G \leftarrow G \cup \{i\}$
- הורד מ- S את כל הפעילויות החותכות את פעילות i

בדוגמה שלנו:
- הפעילות ראשונה שנבחר תהייה 7 ואז S תהייה כל הפעילויות חוץ מ 1,7, כלומר $S = \{2,3,4,5,6\}$
- הפעילות השניה שתבחר תהייה 3 ואז $S = \{2,4,5\}$
- שלושת הפעילויות 2,4,5 מסתיימות באותו זמן ולכן תיבחר אחת מהן בצורה שרירותית, למשל 2, ואז $S = \emptyset$
- האלגוריתם יחזיר $S = \{7,3,2\}$

אז יש לנו אלגוריתם ועכשיו צריך להוכיח נכונות. היינו רוצים פשוט לטעון שבכל שלב אנחנו עושים את הדבר הטוב ביותר ולכן נמצא פתרון אופטימלי, אבל זאת לא הוכחה. גם ראינו רעיונות אחרים "שעושים את הדבר הכי טוב" אבל לא מובילים לפתרון אופטימלי...

משפט: האלגוריתם מחזיר פתרון חוקי בגודל מקסימלי

<u>אבחנה:</u> האלגוריתם מחזיר פתרון חוקי
(משום שבכל שלב כל הפעילויות ב- S זרות לפעילויות ב- G ולכן כאשר אנחנו מוסיפים פעילות מ- S ל- G תמיד כל הפעילויות ב- G זרות)

נשווה את גודל הפתרון שהאלגוריתם מחזיר לגודל פתרון אופטימלי (שימו לב: יתכנו הרבה פתרונות אופטימליים).

<u>טענה נשמרת:</u> בכל שלב באלגוריתם, קיים פתרון אופטימלי שמכיל את G.

<u>הוכחת הטענה הנשמרת (באינדוקציה):</u>

<u>סימונים:</u> נסמן את האיבר ה- ℓ שהאלגוריתם מכניס ל- G ב- i_ℓ, ונסמן את ℓ האיברים הראשונים שהאלגוריתם בחר כ-

$$G_\ell = \{i_1, i_2, \dots, i_\ell\}$$

<u>בסיס האינדוקציה:</u> $G_0 = \emptyset$ מוכלת בכל פתרון אופטימלי

<u>צעד:</u> נניח כי $G_{\ell-1} \subseteq O$ עבור פתרון אופטימלי O כלשהו.
נוכיח כי $G_\ell \subseteq O'$ עבור פתרון אופטימלי O' כלשהו.

מקרה פשוט: $i_\ell \in O$. ניקח $O' = O$.

מקרה פחות פשוט: $i_\ell \notin O$.

בדוגמה שלנו:
נניח אחרי השלב הראשון $G_1 = \{7\}$ וידוע כי $G_1 = \{7\} \subseteq \{7,6,2\} = O$.
בשלב השני נבחר את 3, אבל אז $G_2 = \{7,3\} \nsubseteq O$.

נראה ש- G_2 מוכלת בפתרון אופטימלי אחר: נראה שאנחנו יכולים לבצע החלפה ולקחת בפתרון האופטימלי את 3 במקום את 6 ולקבל פתרון אופטימלי $O' = \{7,3,2\}$ כך ש- $G_2 \subseteq O'$.

למה אפשר לבצע החלפה כזאת?
האלגוריתם בחר את הפעילות שמסתיימת ראשונה. לכן 3 מסתיימת לפני 6 ולכן אם 6 לא נחתכת עם 2 אז גם 3 לא...

בצורה ציורית, מה שיש לנו זה:

נסמן ב- $j_\ell, j_{\ell+1}, \dots, j_k$ את הקטעים ב- O שלא מוכלים ב- G_ℓ, ממויינים לפי זמני סיום. בפרט, j_ℓ הוא הקטע הראשון (מבחינת זמן סיום) ב- O שלא מוכל ב- G_ℓ.

נגדיר

$$O' = O \setminus \{j_\ell\} \cup \{i_\ell\}$$

ונשים לב כי $|O| = |O'|$.

מספיק להוכיח ש- O' פתרון חוקי (משום שאם הוא פתרון חוקי בגודל כמו O אז הוא אופטימלי)

כל הפעילויות ב- O זרות זו לזו ולכן צריך רק להוכיח שהפעילות i_ℓ זרה לשאר הפעילויות ב- O'.

- עפ"י האלגוריתם, הפעילות i_ℓ זרה לכל הפעילויות $i_1, \dots, i_{\ell-1}$
- מכיוון ש- O הוא פתרון חוקי, אנחנו יודעים שהפעילות j_ℓ זרה לפעילויות $i_1, \dots, i_{\ell-1}$. לכן, בשלב ℓ בריצת האלגוריתם מתקיים ש- $j_\ell \in S$. האלגוריתם בחר בשלב זה את i_ℓ ולכן אנחנו יודעים שהפעילות i_ℓ מסתיימת לפני/בזמן ש- j_ℓ מסתיימת.
- מכיוון ש- O חוקי, הפעילויות $j_{\ell+1}, .., j_k$ מתחילות אחרי סיום j_ℓ ובפרט אחרי סיום i_ℓ.
- לכן הפעילות i_ℓ זרה לכל הפעילויות $j_{\ell+1}, .., j_k$

כלומר, i_ℓ זרה לשאר הפעילויות ב- O' ולכן O' הוא פתרון חוקי שגודלו כפתרון O, כלומר O' פתרון אופטימלי.

מ.ש.ל.

מה שהוכחנו עד עכשיו זה שהאלגוריתם לא עושה טעויות. עכשיו אנחנו רוצים להשתמש בזה על מנת להוכיח את המשפט הבא:

<u>משפט:</u> בסוף ריצת האלגוריתם, הקבוצה G היא פתרון אופטימלי.

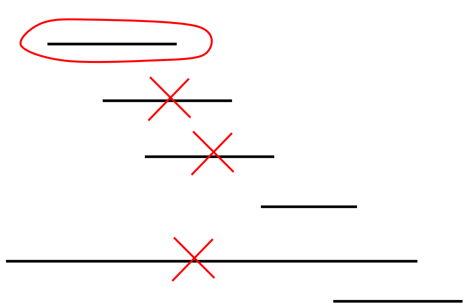
<u>הוכחה:</u>
מהטענה הנשמרת קיים O אופטימלי כך ש- $G \subseteq O$. בסוף הריצה $S = \emptyset$, כלומר כל פעילות שלא ב- G חותכת פעילות ב- G. לכן לא יתכן כי O מכיל פעילות שאיננה ב- G.
מסקנה: בסוף הריצה, הקבוצה O שמובטחת מהטענה הנשמרת מקיימת $O = G$.

<u>ניתוח זמן ריצה:</u>

- איך מוצאים את הפעילות שמסתיימת ראשונה ב- S?
  נמיין את הקטעים עפ"י זמני סיום $O(n \cdot \log n)$
- איך מעדכנים את S?

  בציור: מיינו את הפעילויות לפי זמני סיום. את הפעילות שמסתיימת ראשונה אנחנו לוקחים. אח"כ אנחנו צריכים לזרוק מ- S את כל הפעילויות שנחתכות איתה. איך נמצא אותם?

  מימוש נאיבי ידרוש זמן $O(n)$ לכל שלב ולכן זמן $O(n^2)$ בסה"כ.



נראה מימוש יעיל יותר שלא מתחזק את הקבוצה S באופן מפורש.

<u>מימוש יעיל:</u>

1. מיין את הקטעים עפ"י זמן סיום. נקבל מיון $i_1, i_2, \dots, i_n$ כך ש- $f_{i_1} \le f_{i_2} \le \dots \le f_{i_n}$.
2. אתחל $H \leftarrow \emptyset$ ו- $f = 0$
3. עבור $k = 1$ עד n:
   - אם $f \le s_{i_k}$ אזי בצע $H \leftarrow H \cup \{i_k\}$ ו- $f \leftarrow f_{i_k}$.

המיון מתבצע בזמן $O(n\cdot\log n)$ ואח"כ הלולאה מתבצעת בזמן $O(n)$.

שאלה: איך המימוש היעיל יעבוד בדוגמה האחרונה (הדוגמה מהניתוח הנאיבי של זמן הריצה)?

אנחנו רוצים להוכיח שהאלגוריתם הזה זהה לאלגוריתם הקודם שניסחנו.

סימון: עבור $1\le k\le n$ נסמן ב- H^k את הקבוצה H בתחילת השלב ה- k של האלגוריתם.

טענת עזר: אם בשלב ה- k של האלגוריתם מתקיים ש- i_k זרה לפעילויות ב- H^k אזי מוסיפים את i_k ל- H.
כלומר $i_k\in H^{k+1}$.

הוכחת טענת העזר:
כל הפעילויות ב- H^k מסתיימות לפני f_{i_k} (כי אנחנו עוברים על הפעילויות לפי זמני סיום). לכן, אם i_k זרה לפעילויות ב- H^k אז היא מתחילה אחרי שהם מסתיימות ולכן $f\le s_{i_k}$ ולכן האלגוריתם מכניס את i_k ל- H.

סימונים:
H_ℓ = ℓ הבחירות הראשונות של המימוש היעיל
G_ℓ = ℓ הבחירות הראשונות של המימוש המקורי

משפט: לכל ℓ מתקיים ש- $H_\ell=G_\ell$

הערה: לשם פשטות אנחנו נניח שזמני הסיום של הפעילויות שונים זה מזה (אחרת G_ℓ, H_ℓ לא מוגדרים הייטב).

הוכחה (באינדוקציה)
בסיס האינדוקציה טריוויאלי (כאשר הקבוצות ריקות).

נניח ש $H_{\ell-1}=G_{\ell-1}$ ונוכיח ש- $H_\ell=G_\ell$

נסמן $H_{\ell-1}=G_{\ell-1}=\{i_1,i_2,\dots,i_{\ell-1}\}$ ונניח שהמימוש היעיל בוחר עכשיו את הקטע i_ℓ.

(א) נשים לב ש- i_ℓ מתחילה אחרי שכל הפעילויות ב- $H_{\ell-1}=G_{\ell-1}$ נגמרות (לפי בחירת i_ℓ ע"י האלגוריתם) ולכן זרה להם. לכן $i\in S$.

(ב) לכן כדי להראות ש- i_ℓ היא גם הבחירה הבאה של המימוש המקורי, מספיק להראות ש- i_ℓ מסתיימת ראשונה מבין הפעילויות ב- S, כלומר מבין הפעילויות שזרות ל- $H_{\ell-1}=G_{\ell-1}$.

נניח בשלילה שלא. כלומר קיימת פעילות j_ℓ זרה לכל הפעילויות ב- $H_{\ell-1}$ שנגמרת לפני זמן הסיום של i_ℓ. אבל ז"א שבלולאה במימוש היעיל כבר עברנו על j_ℓ ואז היא היתה צריכה להופיע ב- $H_{\ell-1}$ לפי טענת העזר. סתירה.