אלגוריתמים | 67504

הרצאות | פרופ׳ אלכס סמורודניצקי

כתיבה | נמרוד רק

תשפ״ב סמסטר א׳

כל תרגול מופיע בשבוע שרלוונטי לנושא שלו, ולכן ייתכן שהועבר בשבוע שלפני המקום בו הוא מופיע בסיכום. עד שבוע 10 התרגולים הם לפי ירדן יגיל, ולאחריו לפי אלעד גרנות.

נספחים

חלק א'

חלק ב'

החומר למבחן

הרשימה הזו אינה רשמית וייתכן שחסרים בה דברים (אפע"פ שרובה הוא עותק של מה שאלכס העלה), על כן היא אינה מחייבת ואני <u>לא לוקח אחריות בשום אופן על ההשלכות של למידה ממנה</u> ("נמרוד כתב בסיכום שזה לא במבחן" זה לא ערעור קביל).

חלק א'

- כל האלגוריתמים (חמדני, דינמי, קירוב, תכנון לינארי).
- כל ההוכחות ותתי-הלמות של נכונות האלגוריתמים, ובכלל זה כל הטענות וההוכחות שנלמדו.

חלק ב'

- בעיית הזרימה:
  - כל ההגדרות.
  - כל המשפטים (שטף והחתך) ותתי-הלמות על רשתות זרימה.
  - הוכחות הנכונות של FF ו-EK וההשלמות שלהם מחוץ להרצאה (זמינים בסיכום הזה).
  - הייצוג כבעיית תכנון לינארי.
- התמרות פורייה:
  - ייצוגי פולינמים (מקדמים וערכים).
  - פעולות על פולינומים בשני הייצוגים (בלי הצבת ערך בייצוג הערכים).
  - שורשי היחידה והטענות עליהם.
  - הגדרת DFT והוכחת משפט ה-FFT
  - הפרמול של האלג' לכפל מהיר של פולינומים בייצוג המקדמים.
- קריפטו ואלג' על מספרים:
  - אריתמטיקה מודולרית (הוכחות רק למה שהוכחנו)
  - הוכחת משפט אוילר ופרמה הקטן כמסקנה ממנו.
  - האלג' של RSA וההוכחה שפ' ההצפנה ופענוח שנובעות ממנו אכן הופכיות אחת לשנייה.
  - האלג' של מילר-רבין והוכחת הנכונות והקירוב שלו.

רשימת בעיות אלגוריתמיות

אלגוריתמים חמדניים

1. בעיית התרמיל השברי (אלגוריתם שבוחר לפי משקל סגולי).

הכי קל להוכיח בלי למת ההחלפה: חוקיות ברור ואילו אופטימליות מוצאים פתרון נוסף ומסתכלים על ההפרש שלהם. פותחים את הערך למשקל סגולי ומוציאים את הפריט המיוחד מהסכימה. עושים אריתמטיקה כדי להגיע להפרש המשקלים כפול המשקל הסגולי המיוחד והוא כבר אי-שלילי.

2. בעיית תא הדלק הקטן (אלגוריתם שמתקדם כמה שאפשר כל פעם).

בלמת ההחלפה בוחרים לאחר $k-1$ הסכמות את תא הדלק הכי רחוק שאפשר ומוכיחים שאפשר להגיע אליו מהתחנה לפני וממנו לתחנה אחרי.

3. מציאת MST (קרוסקל).

בלמת ההחלפה מניחים שיש עפ"מ שמסכים על $k-1$ הצלעות הראשונות, מוסיפים צלע ומורידים אחת כדי שיהיה שוב עץ ואז מקבלים עץ עם משקל לכל היותר מינימלי.

4. בעיית שיבוץ משימות (אלגוריתם שבוחר משימות לפי זמני סיום).

בלמת החלפה מסכימים על $k-1$ משימות ואז בוחרים את המשימה הלא חופפת המוקדמת ביותר ובאמצעות ההבחנה מוכיחים שזה שוב חוקי משני הכיוונים.

5. בעיית המטרואידים (אלגוריתם שבוחר בסדר יורד לפי פ' המשקל).

מנסחים למת החלפה שמגדילה את קבוצת הפתרון בהסתמך על תכונת ההחלפה ואז מסיקים שהפתרון החמדן באותו גודל כמו האופטימלי. באופטימליות מניחים בשלילה שיש איבר עם משקל קטן מהמקביל לו (כשהם מסודרים) באופטימלי ומגיעים לסתירה בדרך פעולת האלג' החמדן.

6. קבוצת וקטורים בת"ל בעלת משקל מקסימלי (אלגוריתם שבוחר בסדר יורד לפי פ' המשקל).

כנ"ל רק שמוכיחים שמתקיימת תכונת ההחלפה.

תכנון דינמי

1. בעיית ניתוב משימות (תתי הבעיות הן התחלה מנקודה באמצע ה"מפה").

2. לוח משימות תכנות (תתי הבעיות הן רווח אופטימלי עד לשבוע כלשהו).

3. תת-מחרוזת משותפת מקסימלית (תתי הבעיות הן התמ"א של כל שתי רישאות של המחרוזות).

4. בעיית כפל מטריצות (תתי הבעיות הן פיצול המכפלה איפה שהוא באמצע).

5. בעיית התרמיל השלם (תתי הבעיות הן התרמיל במשקלים שונים, עם ובלי כל פריט) + הנחה מקלה שהמשקלים טבעיים.

6. בעיית מסילת הרכבת (תתי הבעיות הן סיום המסילה בכל חיבור ובאורכים שונים).

7. מסלול קצר ביותר לכל הזוגות (תתי בעיות לפי אורך המסלול, או הגבלת קבוצת הקודקודים בתוך המסלול).

אלגוריתמי קירוב

1. חלוקת משימות (אלגוריתם חמדן $2-\frac{1}{k}$-מקרב שמעמיס איפה שהכי פחות עמוס).

מסתכלים על הפעם שבה הוסיפו למכונה הכי כבדה את המשימה האחרונה שלה ומוכיחים שסה"כ המשימות שלה, שהוא אורך הפתרון, הוא פחות מהקירוב הרצוי.

עושים זאת באמצעות פיצול למשימות עד כה והמשימה האחרונה וכל אחד בנפרד מפתחים. דוחפים את המשימה האחרונה אל תוך הסכום של המשימות לפני ומעדכנים את המשימה האחרונה ל-פחות מהמשימה הארוכה ביותר שהיא קטנה מהאופטימלי. משתמשים בכך שאורך המכונה הכבדה ביותר הוא גדול מהממוצע של המשימות עד כה ושהאחרון קטן מהאופטימלי.

2. בעיית כיסוי ע"י קבוצות (אלגוריתם חמדן $\lceil \ln n \rceil$ מקרב שבוחר בכל שלב את הקבוצה שחופפת כמה שיותר איברים שנותרו).

מוכיחים שבכל שלב קטנה הקבוצה שנותרה (X_t) בלפחות $\frac{1}{k}$ מערכה הנוכחי באמצעות טענת עזר על החפיפה של קבוצות מהפתרון האופטימלי עם X_t. משם, מניחים בשלילה שהפתרון החמדן גודלו פחות ממספר האיטרציות באלג' החמדן (שהוא גודל הפתרון החמדן). משרשרים אי שוויונות על X_t החל מ-X_u ועד $X_0=[n]$ יחד עם טענה מאינפי ומגיעים לסתירה.

3. בעיית התרמיל השברי כתכנון לינארי.

4. בעיית הסרת משולשים.

לכל משולש מגדירים שורה ב-A ודורשים שסכום הערכים על השורה יהיה קטן שווה ל-2.

5. בעיית כיסוי ע"י קודקודים (אלגוריתם שרירותי 2-מקרב שבוחר באקראי צלע, מוסיף את הקודקודים שלה ומסיר את הצלעות שנוגעות בהם).

האלג' מחזיר קודקודים שמגדירים שידוך, שחוסם מלמטה את כמות הקודקודים שיש בפתרון האופטימלי.

6. בעיית כיסוי ע"י קודקודים ממשוקלים (באמצעות סכימה לקירוב שנעזרת ב-ILP מוחלש).

הפתרון המוחזר קטן מפעמיים הפתרון לתכנון הלינארי הרגיל שעצמו קטן מהפתרון האופטימלי.

7. בעיית MAX-3SAT ששואלת מה ההשמה המספקת הכי הרבה פסוקיות ב-CNF-3 (אלג' אקראי דטרמיניסטי 2-מקרב ואלג' הסתברותי $\frac{7}{8}$-מקרב).

עבור ההסת' מוכיחים את ההסת' שהאלג' הבסיסי יחזיר פתרון חוקי באמצעות אש"מ ואז מוכיחים שהאלג' הכללי מחזיר פתרון בהסת' אקספוננציאלית על בסיס מספר האיטרציות שמבצעים.

8. בעיית MAX-LIN-2 ששואלת מה ההשמה המספקת הכי הרבה משוואות לינאריות מעל $\mathbb{F}_2$ (אלג' 2-מקרב שבוחר בכל שלב בחירה חמדנית למשתנה ספציפי על בסיס תוחלת סיפוק המשוואות עבור הבחירה).

מוכיחים שהתוחלת רק עולה על בסיס נוסחת ה-Av ומשם בטרנזיטיביות בגלל שהפתרון האופטימלי הוא פי שתיים מהתוחלת הראשונה אז התוחלת האחרונה שהיא מספר המשוואות המסופקות בפתרון הסופי גדול מחצי האופטימלי.

9. בעיית MAX-CUT ששואלת מה החתך המקסימלי בגרף (אלג' 2-מקרב שמעביר קודקודים מצד לצד אם הם מגדילים את החתך).

כל קודקוד מספק לפחות חצי מהצלעות שלו אחרת היה מועבר צד ומפרמלים את זה עם לחיצות הידיים והוא בטוח עוצר אחרי $|E|$ איטרציות.

10. בעיית M-TSP (הסוכן המטרי) ששואלת מה המעגל העובר בכל קודקודי הגרף בעל המשקל המינימלי בגרף מלא ממושקל (אלג' 2-מקרב שמריץ DFS על עפ"מ בגרף ואוגר את כל הצלעות בעץ ה-DFS).

מוכיחים שמסלול שהוא צמצום של אחר הוא במשקל יותר נמוך. לכן הפתרון שמוחזר הוא לכל היותר במשקל המעבר הכפול על כל עץ ה-DFS שהוא פעמיים העפ"מ שהוא קטן מהפתרון האופטימלי.

בעיית הזרימה

1. בעיית מציאת זיווג מקסימלי ששואלת מה הזיווג הגדול ביותר שיש בגרף דו"צ (אלג' רדוקציה לבעיית הזרימה שבונה רשת עם קיבולים 1 שבה מזרימים דרך צלע אם"ם היא נבחרה לזיווג).

מניחים בשלילה שיש זיווג יותר טוב, מוכיחים שהזיווג המקורי והחדש מתאימים לשטפים ברשת ומגיעים לסתירה למקסימליות השטף.

2. בעיית הזרימה ששואלת מה הזרימה עם שטף מקסימלי ברשת (אלג' FF שבוחר מסלול מ-s ל-t, מזרים כמה שאפשר ובונה צלעות חזור כך שיהיה ניתן לערוך את הזרם במקרה הצורך).

החוקיות נובעת מחוקיות רשת הזרימה המורחבת לאחר הזרמת f, הזרימה השיורית והזרימה המעדוכנת. האופטימליות נובעת ממשפט הלמה והחתך, שכן אין מסילת הרחבה אם האלג' עצר.

3. בעיית הזרימה ששואלת מה הזרימה עם שטף מקסימלי ברשת (אלג' EK שבוחר מסלול מינימלי מ-s ל-t, מזרים כמה שאפשר ובונה צלעות חזור כך שיהיה ניתן לערוך את הזרם במקרה הצורך).

החוקיות ואופטימליות נובעות מחוקיות ואופטימליות FF (מקרה פרטי בסך הכל). הוכחת העצירה נובעת מכך שהקודקודים מתרחקים מ-s, ואם קודקוד הוא על צלע קריטית פעמיים אז הוא מתרחק ממש מ-s ולכן יש חסם על מספר האיטרציות.

4. בעיית המשקיעות והשחקנים ששואלת מה הרווח המקסימלי ומהפקת סרט עם משקיעות שכל אחת דורשת שחקנים מסוימים (אלג' רדוקציה לבעיית הזרימה שבונה רשת עם קיבולי ∞ בין שחקנים של משקיעות ומוצא חתך מינימלי ומחזיר את S).

חוקיות נובעת מהעובדה שיש חתך בגודל סופי ולכן לא תבחר צלע אינסופית ואופטימליות נובעת מהעובדה שמינימום קיבול בחתך זה מקסימום רווח.

5. בעיית הזרימה כתכנון לינארי.

התמרת פורייה בדידה ומהירה

1. בעיית הפעולות על פולינומים בייצוג מקדמים (אלג' שמחבר, מציב ומכפיל נאיבית).

2. בעיית הפעולות על פולינומים בייצוג הערכים (אלג' שמחבר ומכפיל נאיבית, ומציב באמצעות פולינומיי לגראנז' כהעשרה).

3. בעיית כפל פולינומים מהיר במקדמים (אלג' שממיר לערכים עם FFT, מחשב שם כפל וממיר חזרה עם FFT^{-1}).

4. בעיית התאמת המחרוזות ששואלת כמה פעמים מופיעה מחרוזת עם ±1 במחרוזת אחרת עם ±1 (אלג' שמחשב קונבולוציה על היפוך המחרוזת הקצרה ובודק אילו אינדקסים מקבלים ערך מלא).

   נכונות מבוססת על כך שהמחרוזות מותאמות אם"ם מכפלת כל האיברים היא 1.

5. בעיית המרה בין ייצוג מקדמים לייצוג ערכים (אלג' הפרד ומשול שמפתח את הפולינום לשני פולינומים ומשתמש בזהויות של שורשי היחידה).

   נכונות מבוססת על הזהות הריקורסיבית.

קריפטוגרפיה ואלגוריתמים על מספרים

1. בעיית ההצפנה הפומבית (אלג' RSA שמגריל שני ראשוניים ומספר הזר לפ' אוילר על מכפלת הראשוניים).

   נכונות מבוססת על משפט אוילר ואריתמטיקה מודולרית.

2. בעיית הפעולות הבסיסיות (חיבור נאיבי, כפל וחילוק אורך, חישוב מודולו עם ערך שלם, העלה בחזקה באופן לוגריתמי).

3. בעיית המחלק המשותף המקסימלי (אלג' ריקורסיבי שמחלק בשארית ומשתמש בזהויות מודולריות).

   נכונות נובעת ישירות מתכונות ה-gcd.

4. בעיית בדיקת ראשוניות (אלג' רבין-מילר הסת' שמגריל מספרים ובודק האם הסובייקט מקיים איתם את תנאי משפט אוילר).

הוכחת הקירוב מתבססת על כך שיותר מחצי מהמספרים הקטנים ממספר שאינו Carmichael אינם מקיימים את התכונה של משפט אוילר. ההוכחה הזו מתבססת על שימוש בפ' הכפל המודולרי עם אחד מהמספרים שלא מקיים את התכונה (שקיים כי זה לא מספר CM) והוכחה שהיא מעבירה את המספרים שכן מקיימים את התכונה וזרים למספר למספרים שלא מקיימים את התכונה וזרים למספר.

סכמות לפתרון בעיות אלגוריתמיות

- סכימה להוכחת נכונות ואופטימליות של אלגוריתם חמדן.

1. הוכחת חוקיות.
2. הוכחת אופטימליות:

(א) נטען טענת עזר כי קיים פתרון אופטימלי המסכים עם הפתרון החמדן על k איבריו הראשונים, $\forall k \in [|B|]$.

(ב) נוכיח את טענת העזר באינדוקציה. בצעד האינדוקציה נסתכל על הפתרון C המסכים עם הפתרון B על $k-1$ איבריו הראשונים (מה"א הוא חוקי ואופטימלי) ועל הפתרון C' המסכים עם הפתרון B על k איבריו הראשונים ונוכיח כי C' הוא פתרון חוקי ואופטימלי.

(ג) נסיק כי עבור $k = m$ קיים פתרון אופטימלי המכיל את הפתרון החמדן B ולכן B אופטימלי.

- סכימה לפתירת בעיה באמצעות תכנון דינמי.

1. הגדרת תתי בעיות.
2. הגדרת נוסחאת ריקורסיה (המקשרת בין תת הבעיות לבעיה הגדולה).
3. בניית טבלה - תיאור, מילוי וחילוץ פתרון.
4. ניתוח זמן ריצה.
5. הוכחת אופטימליות (אינדוקציה על מילוי הטבלה).

- סכימה לפתירת בעיית קירוב באמצעות תכנון לינארי בשלמים.

1. נתרגם את הבעיה לבעיית מינימיזציה של פ' לינארית במשתנים שמקבלים ערכים שלמים ומקיימים אילוצים לינאריים (ניסוח אנליטי של הבעיה המקורית) - זו בעיית ILP.
2. נסיר את האילוץ של הערכים השלמים ונחליף אותו באי שוויונים לינאריים (צריכה להיות רלקסציה - הבעיה החדשה צריכה להיות קלה יותר).
3. נפתור את בעיית ה-LP שקיבלנו בשלב הקודם בעזרת אלג' הפותר בעיות תכנון לינארי (Simplex לדוגמה) ונקבל פתרון אופטימלי לבעיה זו.
4. נעגל את הפתרון האופטימלי לבעיית ה-LP לפתרון טוב לבעיה המקורית ונחזיר את הפתרון המקורב.

- סכימה לפתירת בעיות באמצעות רדוקציה לבעיית הזרימה.

1. נבנה רשת זרימה המתאימה לבעיה.
2. נריץ את אלג' EK על הרשת שבנינו ונקבל זרימה מקסימלית f.
3. נחזיר פתרון לבעיה המקורית על סמך f.

כדי להוכיח נכונות נוכיח את הדברים הבאים:

1. כל פתרון חוקי של הבעיה ניתן לתרגום לזרימה חוקית ברשת (מוודא שלא נפספס פתרונות חוקיים של הבעיה, ובפרט האופטימלי).
2. כל זרימה חוקית ברשת ניתנת לתרגום לפתרון חוקי של הבעיה (מוודא שהפתרון לבעיה חוקי).
3. המרה בין פתרונות משמרת את ערך הבעיה (משמש להוכחת אופטימליות).

סוף הוכחת הנכונות של אדמונדס וקארפ

להלן המשך הוכחת הנכונות של אלג' E & K.

הגדרה הצלע (x,y) תקרא קריטית באיטרציה $i+1$ אם היא שייכת למסילת ההרחבה p_i ואם הקיבול השיורי שלה הוא המינימלי בין כל הצלעות של p_i, כלומר $c_{f_i}(x,y)=c_{f_i}(p_i)$.

למה 2 יהיו $x,y\in V$ ו-$i<k$ כך ש-(x,y) היא צלע קריטית באיטרציות $i+1$ ו-$k+1$ של האלג'. אזי $\delta_k(x)\geq\delta_i(x)+2$. כלומר, כל בין כל פעמיים שצלע היא קריטית, היא מתרחקת מ-s באופן מונוטוני עולה ממש.

הוכחה: נוכיח כי מהיות (x,y) צלע קריטית באיטרציה ה-$i+1$, הרי ש-(x,y) לא שייכת ל-$G_{f_{i+1}}$. מספיק שנוכיח כי $c_{f_{i+1}}(x,y)=0$.

$$\begin{aligned}
f_{i+1}(x,y) &= f_i(x,y)+\Delta_{f_i,p_i}(x,y)\\
&\overset{\text{קריטית}}{=} f_i(x,y)+c_{f_i}(x,y)\\
&= f_i(x,y)+(c(x,y)-f_i(x,y))\\
&= c(x,y)
\end{aligned}$$

ולכן $c_{f_{i+1}}(x,y)=c(x,y)-f_{i+1}(x,y)=0$ ולכן (x,y) לא ב-G_{f_i}. מצד שני, מהיותה צלע קריטית באיטרציה $k+1$, הרי ש-(x,y) שייכת ל-p_k ולכן היא צלע ב-G_{f_k}. כלומר הזרימה דרך (x,y) חייבת לקטון באיטרציה $j+1$ עבור $i+1\leq j\leq k-1$ (לא באיטרציה ה-i כי שם הזרימה גדלה, לא עבור k והלאה כי $j+1$ כבר אחרי G_{f_k} וזה צריך להיות מוקטן כבר לפני G_{f_k}, כלומר לכל היותר $k-1$). הזרימה קטנה אם"ם $(y,x)\in p_j$ לכן

$$\delta_k(x)\overset{\text{למה 1}}{\geq}\delta_j(x)\overset{(y,x)\in p_j}{=}\delta_j(y)+1\overset{\text{למה 1}}{\geq}\delta_i(y)+1\overset{(x,y)\in p_i}{=}\delta_i(x)+2$$

■

למת עזר תהי (x,y) צלע קריטית באטירציה ה-$i+1$ של האלג'. אזי $(x,y)\in E$ או $(y,x)\in E$, כאשר E הוא אוסף הצלעות של הרשת המקורית $\tilde{N}$.

הוכחה: (x,y) היא צלע ב-G_{f_i} ולכן $0<c_{f_i}=c(x,y)-f_i(x,y)$.

אם $c(x,y)>0$ אז מהגדרת הקיבול המורחב $(x,y)\in E$.

אחרת $f_i(x,y)<0$ ומאנטי-סימטריה $f_i(y,x)>0$ ומאילוץ הקיבול (לרשתות מורחבות) $c(y,x)\geq f_i(y,x)>0$ ולכן מהגדרת הקיבול המורחב $(y,x)\in E$. ■

משפט (נכונות E & K) האלג' של אדמונדס וקארפ עוצר לאחר $\mathcal{O}(|V|\cdot|E|)$.

הוכחה: נוכיח כי לכל $x,y\in V$, הצלע יכולה להיות צלע קריטית לכל היותר $\frac{|V|+1}{2}$ איטרציות.

נסמן ב-t את מספר האיטרציות בהן (x,y) צלע קריטית ונניח שאלו האיטרציות $i_1+1,\ldots,i_t+1$. מלמה 2, לכל $1\leq j\leq t-1$, $\delta_j(x)\geq\delta_{j+1}(x)+2$ לכן מהיות $\delta_{i_1}\geq 0$ הרי ש-$\delta_{i_t}(x)\geq 2(t-1)$. מצד שני, ב-$G_{f_{i_t}}$ יש $|V|$ קודקודים ולכן $\delta_{i_t}(x)\leq|V|-1$ ולכן $2t-2\leq|V|-1$ אם"ם $t\leq\frac{|V|+1}{2}$ כרצוי.

מלמת העזר יש לכל היותר $2|E|$ צלעות שיכולות להיות צלעות קריטיות. לכן בכל איטרציה יש צלע קריטית מבין $2|E|$ אפשרויות, וזו מופיעה לכל היותר $\frac{|V|+1}{2}$, כלומר יש לכל היותר

$$\frac{|V|+1}{2}2|E|=\mathcal{O}(|V|\cdot|E|)$$

איטרציות. ■

סוף ההוכחה של מילר-רבין

לאלגוריתם והערות ראשונות לגביו

הגדרה מספר טבעי m יקרא מספר Carmichael אם m אינו ראשוני, אבל לכל a זר ל-m מתקיים $a^{m-1}\equiv 1 \mod m$.

הערה מספרי Charmichael נדירים מאוד.

משפט אם n אינו מספר Carmichael אזי:

1. אם n ראשוני, האלג' תמיד מחזיר "ראשוני"
2. אם n אינו ראשוני, האלג' יחזיר "לא ראשוני" בסיכוי לפחות $1-\frac{1}{2^t}$.

הוכחה:

1. אם n ראשוני אז כל מספר a זר ל-n מקיים $a^{n-1}\equiv 1 \mod n$ ממשפט פרמה הקטן. לכן בכל האיטרציות $f=1$ והאלג' יחזיר "ראשוני".

(א) נסמן $B=\{1\le a\le n-1 : a^n\not\equiv 1 \mod n\}$ כלומר כל המספרים שהיו "מכשילים" את n בבדיקת ראשוניות לפי משפט פרמה הקטן.

נוכיח כי $|B|\ge\frac{n-1}{2}$ ונקבל כי בכל איטרציה של האלג' $f\neq 1$ בסיכוי לפחות $\frac{1}{2}$ ולכן הסיכוי שבכל האיטרציות יצא $f=1$ הוא לכל היותר $\frac{1}{2^t}$.

עבור $Y=\{y\in[n-1] : \gcd(y,n)\neq 1\}$ מתקיים $[n-1]=\mathbb{Z}_n^*\dot{\cup}Y$.

לכל $y\in Y$ מתקיים $\gcd\left(y^{n-1},n\right)=\gcd\left(y^{n-1} \mod n,n\right)>1$ ולכן $y^{n-1} \mod n\neq 1$ (כי יש לו מחלק גדול מ-1) כלומר $y\in B$, לכן $Y\subseteq B$.

עתה מספיק שנוכיח ש-$|B\cap\mathbb{Z}_n^*|\ge\frac{|\mathbb{Z}_n^*|}{2}$ כי אם לפחות חצי מ-$\mathbb{Z}_n^*$ וכל Y ב-B אז לפחות חצי מ-$[n-1]$ ב-B, או באותיות

$$|B|=|Y|+|B\cap\mathbb{Z}_n^*|\ge|Y|+\frac{|Z_n^*|}{2}=((n-1)-|\mathbb{Z}_n^*|)+\frac{|\mathbb{Z}_n^*|}{2}=n-1-\frac{|\mathbb{Z}_n^*|}{2}\ge n-1-\frac{n-1}{2}=\frac{n-1}{2}$$

נוכיח כי עבור

$$C=B\cap\mathbb{Z}_n^*=\left\{a\in[n-1] : \gcd(a,n)=1, a^{n-1}\not\equiv 1 \mod n\right\}$$

מתקיים $|C|\ge\frac{|\mathbb{Z}_n^*|}{2}$.

מהיות n לא Carmichael, קיים a זר ל-n כך ש-$a^{n-1}\not\equiv 1 \mod n$. נגדיר העתקה $f_a:\mathbb{Z}_n^*\to\mathbb{Z}_n^*$ ע"י $f_a(x)=ax \mod n$. ראינו כי העתקה זו חחע"ל. נראה כי היא מעבירה כל $d\in\mathbb{Z}_n^*\backslash C$ ל-C.

יהי $d \in \mathbb{Z}_n^* \backslash C$. $a^{n-1} \equiv 1 \mod n$ (הוא לא ב-C) ולכן

$$(f_a(d))^{n-1} \equiv (ad)^{n-1} = a^{n-1} \cdot d^{n-1} \equiv a^{n-1} \cdot 1 \not\equiv 1 \mod n$$

ולכן f_a מעבירה מ-$\mathbb{Z}_n^* \backslash C$ ל-C. f_a חח"ע ולכן $|\mathbb{Z}_n^* \backslash C| \leq |C|$ ולכן $\frac{|\mathbb{Z}_n^*|}{2} \leq |C|$.

■

שבוע $\mathbb{O}$ | תזכורת מקורסים אחרים

תרגול

סקרנו מחדש אינדוקציה וחסמים אסימפטוטיים.

הערה כשאנחנו מנתחים זמן ריצה של אלג׳ אנחנו סופרים את כמות הפעולות האלמנטריות שהאלג׳ מבצע. בתורת המספרים (כשהקלט הוא למשל מספר או כמות קבועה של מספרים), הפעולות שנספור הן פעולות ביטיות.

הערה בקורס נאמר כי אלגוריתם הוא יעיל אם הוא רץ בזמן פולינומיאלי בגודל הקלט.

דוגמה פירוק לגורמים ראשוניים.

קלט $p_1^{m_1} \cdot \dots \cdot p_l^{m_l} = n \in \mathbb{N}$.

פלט $L = \{p_1, \dots, p_l\}$ המספרים הראשוניים שמרכיבים את n.

אלגוריתם

1. נאתחל $L = \varnothing$, $r = n$.
2. לכל $2 \le i \le \sqrt{n}$:
   אם i מחלק את r:
   (א) נוסיף את i ל-L.
   (ב) נמשיך לחלק את r ב-i כל עוד אפשר.
3. אם נשארנו עם $r \neq 1$, נוסיף את r ל-L.

זמן ריצה בגלל שאנחנו רצים על מספר, גודל הקלט הוא $k = \log k$. סיבוכיות האלג׳ היא $\Theta\left(\sqrt{n}\right)$ (שלב $1, 3$ לוקחים $\mathcal{O}\left(1\right)$ ואילו שלב 2 לוקח $\Theta\left(\sqrt{n}\right)$) אבל $\sqrt{n} = \sqrt{2^k} = \left(2^k\right)^{\frac{1}{2}} = 2^{\frac{k}{2}} = \sqrt{2}^k$ ולכן זמן הריצה הוא $\Omega\left(\sqrt{2}^k\right)$ שזה אקספוננציאלי ולכן הוא לא אופטימלי.

אלגו׳	גודל הקלט	זמן הריצה	האם יעיל?
Merge Sort	n	$\mathcal{O}\left(n \log n\right)$	כן - פולינומיאלי עם n^2
חיפוש בינארי	n	$\Theta\left(\log n\right)$	כן - תת לינארי
קרוסקל	$\|E\| + \|V\|$	$\mathcal{O}\left(\|E\| \log \|E\|\right)$	כן - $\|E\| \log \|E\| \le \|E\|^2 \le \left(\|E\| + \|V\|\right)^2$
פירוק לראשוניים	$k = \log n$	$\Omega\left(\sqrt{n}\right)$	לא, כי אקספוננציאלי

שבוע $\mathbb{I}$ | מבוא ואלגוריתמים חמדניים

הרצאה

חלק א' של ההרצאה

איך נפתור בעיות אלגוריתמיות?

- אנחנו כבר יודעים מה הפתרון.
- רידוקציה למשהו שאנחנו יודעים.
- פתרון איטרטיבי תוך מיקסום של רווח מקומי.
- חלוקת הבעיה לבעיות קטנות יותר (הפרד ומשול).
- שימוש באקראיות, לפעמים אנחנו יכולים לפתור באמצעות החלטה שרירותית באופן יעיל יותר מהחלטה מושכלת כלשהי.
- נחפש פתרון מקורב כשאי אפשר אחרת (בעיות NP-קשות).

דוגמה נתונים n מספרים שלמים $a_1, \ldots, a_n$. האם יש שניים שסכומם הוא 2021?

פתרון 1 נעבור על כל הזוגות $1 \leq i < j \leq n$ ונבדוק האם סכומם הוא 2021. יש $\binom{n}{2}$ זוגות ולכן זמן הריצה הוא $\Theta\left(n^2\right)$.

פתרון 2 נמיין את המספרים בסדר עולה $a_{i_1} \leq \cdots \leq a_{i_n}$ ולכל $1 \leq j \leq n$ נבדוק האם $2021 - a_{i_j}$ נמצא במערך באמצעות מיון בינארי. המיון יקח $\Theta\left(n \log n\right)$ ונבצע n פעולות חיפוש בינארי ($\mathcal{O}\left(\log n\right)$) ולכן סה"כ זמן הריצה הוא $\Theta\left(n \log n\right)$.

הגדרה <u>אלגוריתם חמדני</u> בונה פתרון באופן איטרטיבי כאשר בכל שלב הוא ממקסם פונקציית רווח מקומית.

הערה נבנה תאוריה כללית המאפשרת לזהות דוגמאות של בעיות אלגוריתמיות שניתן לפתור באמצעות אלגוריתם חמדן.

בעיית התרמיל השברי

סיפור מסגרת גנב נכנס לחנות שבה יש מספר פריטים לכל פריט ערך ומשקל משלו. לגנב יש תרמיל עם משקל מירבי שהוא יכול לשים בו. הגנב מעוניין להכניס לתרמיל פריטים עם ערך כולל מקסימלי. הפריטים ניתנים לחלוקה (אפשר לקחת חצי פסנתר).

קלט n הוא מספר הפריטים בחנות, $0 \leq W$ הוא המשקל המירבי של התרמיל, רשימה של n זוגות של מספרים אי שליליים $\left(v_1, w_1\right), \ldots, \left(v_n, w_n\right)$ המייצגים את הערך והמשקל של כל פריט בהתאמה.

פלט רשימה של n מספרים $0 \leq x_1, \ldots, x_n \leq 1$ כאשר x_i הוא חלקיות הפריט ה-i, המקיימת את האילוץ $\sum_{i=1}^{n} x_i w_i \leq W$ וכך ש-$\sum_{i=1}^{n} x_i v_i$ מקסימלי.

דוגמה $n=3$, $W=50$, $(60,10)$, $(100,20)$, $(120,30)$.

פתרון 1 $x_1=x_2=1, x_3=0$ הסטודנטית המשקיעה תבדוק שאכן זהו פתרון מקסימלי המקיים את אילוץ המשקל.

פתרון 2 נחשב לכל פריט את הערך הסגולי שלו (ערך ליחידת משקל) ומשם נבחר את אלו בעלי הערך הסגולי הגבוה ביותר עד שיגמר המקום (המשקל). במקרה הזה, $r_1=4, r_2=5, r_3=6$. ראשית נבחר $x_3=1$ וגם $x_2=1$ ואז נשאר לנו $\frac{2}{3}=\frac{50-20-10}{60}=x_1$ שזהו גם פתרון אופטימלי.

פתרון כללי

מקרה I - $\sum_{i=1}^{n} w_i \le W$. במקרה כזה $x_1=\cdots=x_n=1$ הוא פתרון חוקי ואופטימלי.

מקרה II - $\sum_{i=1}^{n} w_i > W$. נחשב לכל $1\le i\le n$ את הערך סגולי של הפריט ה-i, $r_i=\frac{v_i}{w_i}$ ונמיין את הפריטים בסדר יורד לפי הערכים הסגוליים ונניח בה"כ $r_1\ge r_2\ge\cdots\ge r_n$. קיים אינדקס $0\le t\le n-1$ כך ש- $\sum_{i=1}^{t} w_i\le W$ אבל $\sum_{i=1}^{t+1} w_i > W$.
נגדיר

$$x_1=\cdots=x_t=1,\quad x_{t+1}=\frac{W-\sum\limits_{i=1}^{t} w_i}{w_{t+1}},\quad x_{t+2}=\cdots=x_n=0$$

טענה (נכונות) האלגוריתם מחזיר פתרון חוקי ואופטימלי של הבעיה.

הוכחה: חוקיות: ברור כי $0\le x_1=\cdots=x_t=1\le 1$ וגם $0\le x_{t+2}=\cdots=x_n=0\le 1$ ואילו

$$0=\frac{\sum\limits_{i=1}^{t} w_i-\sum\limits_{i=1}^{t} w_i}{w_{t+1}}\le x_{t+1}=\frac{W-\sum\limits_{i=1}^{t} w_i}{w_{t+1}}<\frac{\sum\limits_{i=1}^{t+1} w_i-\sum\limits_{i=1}^{t} w_i}{w_{t+1}}=\frac{w_{t+1}}{w_{t+1}}=1$$

ומבחינת אילוץ המשקל,

$$\sum_{i=1}^{n} x_iw_i=\sum_{i=1}^{t} 1\cdot w_i+x_{t+1}w_{t+1}+\sum_{i=t+2}^{n} 0\cdot w_i=\sum_{I=1}^{t} w_i+\frac{W-\sum\limits_{i=1}^{t} w_i}{w_{t+1}}\cdot w_{t+1}=\sum_{I=1}^{t} w_i+\left(W-\sum_{I=1}^{t} w_i\right)=W$$

אופטימליות: יהי $y_1,\dots,y_n$ פתרון חוקי כלשהו לבעיה. נוכיח כי $\sum_{i=1}^{n} y_iv_i \geq \sum_{i=1}^{n} x_iv_i$. מספיק שנוכיח כי $\sum_{i=1}^{n} x_iv_i - \sum_{i=1}^{n} y_iv_i \geq 0$.

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i-y_i)v_i &= \sum_{i=1}^{n}(x_i-y_i)w_ir_i \\
&= \sum_{i=1}^{t}(x_i-y_I)w_ir_i + (x_{t+1}-y_{t+1})w_{t+1}r_{t+1} + \sum_{i=t+2}^{n}(x_i-y_i)w_ir_i \\
&\geq \sum_{i=1}^{t}\left(\underbrace{x_i}_{1}-\underbrace{y_i}_{\leq 1}\right)w_i\underbrace{r_{t+1}}_{\leq r_i} + (x_{t+1}-y_{t+1})w_{t+1}r_{t+1} + \sum_{j=t+2}^{n}\left(\underbrace{x_j}_{0}-\underbrace{y_j}_{\geq 0}\right)w_j\underbrace{r_{t+1}}_{\geq r_j} \\
&= r_{t+1}\sum_{i=1}^{n}(x_i-y_i)w_i \\
&= r_{t+1}\left(\sum_{i=1}^{n}x_iw_i - \sum_{i=1}^{n}y_iw_i\right) \\
&= r_{t+1}\left(W - \underbrace{\sum_{i=1}^{n}y_iw_i}_{\leq W}\right) \\
&\geq 0
\end{aligned}
$$

■

חלק ב' של ההרצאה

פסאודו קוד ננסח פורמלית פסאודו קוד לאלג' שלנו לפתרון בעיית התרמיל השברי.

0. עיבוד מידע מוקדם: נחשב את הערכים הסגוליים של הפריטים ונמיין אותם בסדר יורד של הערכים. ($\Theta(n\log n)$)
1. אתחול: $K=\varnothing$. ($\mathcal{O}(1)$)
2. איטרציה: נעבור על הפריטים לפי הסדר ובכל שלב נכניס כמה שיותר מהפריט שהגענו אליו. ($\mathcal{O}(n)$)
3. עצירה: כאשר התרמיל מתמלא (או כשנגמרים הפריטים) נעצור ונחזיר את K. ($\mathcal{O}(1)$)

זמן ריצה סה"כ $\Theta(n\log n+n)=\Theta(n\log n)$.

עקרונות פורמליים לאלגוריתמים חמדניים ולמת החלפה

נפתור בעיות אפוטימיזציה. כל בעיה כזו מגדירה את מרחב הפתרונות החוקיים של $\mathcal{S}$ שלה פונקציית ערך $f:\mathcal{S}\to\mathbb{R}_+$ ונרצה למצוא $s^*\in\mathcal{S}$ כך ש-$f(s^*)$ מקסימלית (מינימלית). נחפש דרך איטרטיבית לבניית פתרון אופטימלי כשבכל שלב האלג' בוחר בחירה חמדנית (למקסם רווח מקומי). כדי להוכיח שדרך כזו תביא אותנו לפתרון אופטימלי, נשתמש בעיקרון של שיפורים מקומיים, כלומר, נראה שאם באיטרציה מסויימת האלג' לא עושה בחירה חמדנית הוא טועה (למשל ניתן להחליף את בחירתו בבחירה חמדנית ורק לשפרו). הטענה שאומרת שבכל שלב כדאי לאלג' לבצע בחירה חמדנית נקראת למת החלפה. למת החלפה מפרמלת את העיקרון החמדני.

טענה (למת ההחלפה לבעית התרמיל השברי) יהי $y = (y_1, \ldots, y_n)$ פתרון חוקי לטענה (כאשר ה-n-יה כבר ממוינת בסדר יורד). נניח כי קיים אינדקס $1 \le j \le n$ כך ש-$y_j < 1$ (החלקיות לא מלאה) וכך ש-$\sum_{i=1}^{j} y_i w_i < W$ אז קיים פתרון חוקי $z = (z_1, \ldots, z_n)$ כך ש-$z_1 = y_1, \ldots, z_{j-1} = y_{j-1}$ כך ש-$z_j > y_j$ וכך ש-$\sum_{i=1}^{n} y_i v_i < \sum_{i=1}^{n} z_i v_i$ (במילים אחרות y לא אופטימלי).

הוכחה: נניח בה"כ כי הערכים הסגוליים של כל הפריטים שונים זה מזה, כי אם יש שני פריטים עם ערך סגולי נוכל לאחד אותם לפריט אחד ולקבל בעיה שקולה.

1. $\sum_{i=1}^{n} y_i w_i < W$. במקרה זה נוכל להגדיל את החלקיות של y_j באופן שלא יעבור את המשקל (הרי $y_j < 1$) ולקבל באופן טריוויאלי פתרון אופטימלי יותר.

2. $\sum_{i=1}^{n} y_i w_i = W$. לפי ההנחה $\sum_{i=1}^{j} y_i w_i < W$ ולכן קיים $k > j$ כך ש-$y_k > 0$.

נגדיר $d_j = (1 - y_j)\, w_j$ הוא המשקל המקסימלי שאפשר להכניס עדיין ל-j (הפריט היקר יותר ליחידת משקל) ו-$d_k = y_k w_k$ הוא המשקל המקסימלי שאפשר להוציא מ-k מהתרמיל (הפריט הזול יותר ליחידת משקל) ולבסוף $0 < d = \min\{d_j, d_k\}$ המשקל שלוקח לא יותר משאפשר מ-k ומכניס לא יותר משנשאר ל-j.

נגדיר פתרון חדש z באופן הבא.

$$z_k = y_k - \frac{d}{w_k}, \quad z_j = y_j + \frac{d}{w_j}, \quad \forall i \ne j,k, z_i = y_i$$

נוכיח כי z הוא פתרון חוקי וטוב יותר מ-y. $\forall i \ne j,k$ מתקיים $0 \le z_i = y_i \le 1$. עבור $i = j$,

$$0 \le y_j < y_j + \frac{d}{w_j} = z_j \le y_j + \frac{d_j}{w_j} = y_j + \frac{(1 - y_j)\, w_j}{w_j} = 1$$

בנוסף $0 \le z_k \le 1$ (מההגדרה).

$$\sum_{i=1}^{n} z_i w_i - \sum_{i=1}^{n} y_i w_i = \sum_{i=1}^{n} (z_i - y_i)\, w_i = (z_j - y_j)\, w_j + (z_k - y_k)\, w_k = \frac{d}{w_j} w_j - \frac{d}{w_k} w_k = d - d = 0$$

ולכן

$$\sum_{i=1}^{n} z_i w_i = \sum_{i=1}^{n} y_i w_i = W$$

נוכיח כי $\sum_{i=1}^{n} y_i v_i < \sum_{i=1}^{n} z_i v_i$.

$$\sum_{i=1}^{n} z_i v_i - \sum_{i=1}^{n} y_i v_i = \sum_{i=1}^{n} (z_i - y_i)\, v_i = (z_j - y_j)\, v_j + (z_k - y_k)\, v_k = \frac{d}{w_j} v_j - \frac{d}{w_k} v_k = dr_j - dr_k = d\,(r_j - r_k) \overset{k>j}{>} 0$$

■

תרגול

בעיית תא הדלק הקטן

סיפור רקע יש לנו כלי רכב עם תא דלק קטן. נרצה למצוא מסלול עם כמה שפחות עצירות מהמקור ליעד.

קלט $N \in \mathbb{N}$ שהוא מס' הק"מ שאפשר לנסוע עם מיכל דלק מלא, $a_1 < \cdots < a_n \in \mathbb{N}$ שהם תחנות הדלק לפי הסדר (a_1 נק' המוצא ו-a_n נקודת היעד) המקיימות $a_{i+1} - a_i \leq N$.

פלט $(b_1, \ldots, b_m)$ כך ש-$b_i \in \{a_j\}$ באורך מינימלי ותקיים את התנאים הבאים: $b_1 = a_1, b_m = a_n$ ו-$b_{i+1} - b_i \leq N$.

דוגמה $a = 0, 1, 7, 9, 16, 17, 20$, $N = 10$ במקרה זה $(0, 7, 16, 20)$, $(0, 9, 16, 20)$, $(0, 7, 17, 20)$, $b^* = (0, 9, 17, 20)$ הם פתרונות אופטימליים.

פסאודו-קוד

1. <u>אתחול</u>: $B = (a_1)$.
2. <u>איטרציה</u>: נעבור על כל התחנות ונוסיף את a_i ל-B אם מתקיים אחד מהתנאים הבאים:
   (א) $a_i = a_n$.
   (ב) אם אין מספיק דלק כדי להגיע לתחנה a_{i+1}.
3. <u>עצירה</u>: אם $a_i = a_n$ נעצור ונחזיר את B.

טענה האלג' שהצגנו מחזיר פתרון חוקי.

הוכחה: אתחלנו $B = (a_1)$ ולכן $b_1 = a_1$ כנדרש. עברנו על כל התחנות והוספנו את a_i אם $a_i = a_n$ וסיימנו ולכן $b_m = a_n$.

אנחנו מוסיפים תחנה a_i ל-B רק אם אנחנו יכולים להגיע אליה מהתחנה האחרונה בה עצרנו ולכן מתקיים $b_i - b_{i-1} \leq N$. ■

טענת עזר $\forall k \in [n]$ קיים פתרון אופטימלי C שמסכים עם הפתרון החמדן על k האיברים הראשונים, כלומר קיים פתרון חוקי מהצורה $C = (b_1, \ldots, b_k, c_{k+1}, \ldots, c_{m'})$.

הוכחה: נוכיח באינדוקציה על k.

בסיס ($k = 1$): כל פתרון חוקי מתחיל ב-$a_1 = b_1$ ובפרט פתרון אופטימלי.

צעד ($k - 1 \to k$): נוכיח כי $C' = (b_1, \ldots, b_k, c_{k+1}, \ldots, c_{m'})$ הוא פתרון חוקי ואופטימלי.

חוקיות: $b_1 = a_1$ כי B חוקי ואילו $c_{m'} = a_n$ מה"א על C. מחוקיות B מתקיים $b_i - b_{i-1} \leq N$ עבור $1 \leq i \leq k$ ואילו מחוקיות C (ה"א) מתקיים $c_i - c_{i-1} \leq N$ עבור $k+1 \leq i \leq m$. נשאר להראות כי $c_{k+1} - b_k \leq N$. מאופן פעולת האלג' החמדן מתקיים $b_k \geq c_k$ (כי אם עצרנו ב-b_{k-1} האלג' החמדן יבחר את התחנה הרחוקה ביותר שהוא עדיין יכול ליסוע אליה) ולכן $c_{k+1} - b_k \leq c_{k+1} - c_k \leq N$.

אופטימליות: $|C'| = |C|$ ולכן בעזרת ה"א C אופטימלי ולכן גם C'. ■

טענה האלג' שהצגנו מחזיר פתרון אופטימלי, כלומר, לכל פתרון חוקי B' מתקיים $|B| \leq |B'|$.

הוכחה: עבור $k = m$ נקבל כי קיים פתרון אופטימלי $C = (b_1, \ldots, b_m, c_{m+1}, \ldots, c_{m'})$. אבל בגלל שהוספת איברים לפתרון בהכרח פוגעת באופטימליות נקבל כי C בהכרח מסתיים ב-b_m, ולכן $C = B$ הוא פתרון אופטימלי. ■

סכימת הוכחת נכונות אלג' חמדן

1. הוכחת חוקיות.
2. הוכחת אופטימליות:

(א) נטען טענת עזר כי קיים פתרון אופטימלי המסכים עם הפתרון החמדן על k איבריו הראשונים, $\forall k \in [|B|]$.

(ב) נוכיח את טענת העזר באינדוקציה. בצעד האינדוקציה נסתכל על הפתרון C המסכים עם הפתרון B על $k-1$ איבריו הראשונים (מה"א הוא חוקי ואופטימלי) ועל הפתרון C' המסכים עם הפתרון B על k איבריו הראשונים ונוכיח כי C' הוא פתרון חוקי ואופטימלי.

(ג) נסיק כי עבור $k = m$ קיים פתרון אופטימלי המכיל את הפתרון החמדן B ולכן B אופטימלי.

מציאת MST

קלט גרף ממושקל לא מכוון.

פלט MST על הגרף.

פסאודו-קוד האלג' של קרוסקל.

1. עיבוד מוקדם: נמיין את צלעות הגרף בסדר עולה של משקלן.
2. אתחול: $T = \varnothing$.
3. איטרציה: נעבור על הצלעות ונוסיף את הצלע רק אם היא לא סוגרת מעגל ב-T.
4. עצירה: כשעברנו על כל הצלעות נעצור ונחזיר את T.

טענה האלג' של קרוסקל מחזיר עץ פורש (הוכחה בתרגיל בית).

טענת עזר נסמן $t_1, \ldots, t_{|V|-1}$ הצלעות ב-T לפי סדר ההכנסה. $\forall k \in |V|-1$ קיים פתרון אופטימלי S שמקיים $T_k = \{t_1, \ldots, t_k\} \subseteq S$.

הוכחה: נוכיח באינדוקציה על k הצלעות הראשונות.

בסיס $(k=0)$: כל פתרון חוקי בהכרח מכיל את $T_9=\varnothing$ ובפרט פתרון אופטימלי.

צעד $(k-1\to k)$: נסתכל על הגרף $S\cup\{t_k\}$, אם $t_k\in S$ סיימנו. אחרת, בגלל ש-S הוא עץ פורש, בהכרח קיים מעגל ב-$S\cup\{t_k\}$ ובהכרח אחת צלעות המעגל אינה נמצאת ב-T כי אחרת היינו מקבלים מעגל ב-T סתירה לחוקיות שלו. נסמן צלע זו ב-e. נסתכל על $S'=S\cup\{t_k\}\setminus\{e\}$. נוכיח כי S' הוא פתרון חוקי ואופטימלי.

חוקיות: הוספנו צלע לעץ פורש ואז הסרנו צלע מהמעגל שנוצר ולכן אנחנו שוב עם עץ פורש.

אופטימליות: נראה כי $w(e)=w(t_k)$ ומכך נסיק כי $w(S')=w(S)$ ולכן S' יהיה אופטימלי גם הוא.

נניח בשלילה כי $w(t_k)>w(e)$. לפי אופי פעולת האלג' החמדן, e הייתה אמורה להתווסף ל-T אך זה לא קרה כי היא סגרה מעגל עם $\{t_1,\dots,t_{k-1}\}$ אבל $\{t_1,\dots,t_{k-1},e\}\subseteq S$ פתרון חוקי ולכן עץ סתירה.

נניח בשלילה כי $w(t_k)<w(e)$ לכן $w(S')=w(S)+w(t_k)-w(e)<w(S)$ ולכן S לא אופטימלי סתירה.

לכן $w(t_k)=w(e)$ ולכן S' אופטימלי. ■

מסקנה קיים פתרון אופטימלי S שמקיים $T\subseteq S$. אם נוסיף צלעות ל-T אז יסגר מעגל כי T הוא עץ פורש ולכן $S=T$ הוא פתרון אופטימלי.

בעיית החזרת העודף

קלט $k\in\mathbb{N}$ סכום שאנחנו רוצים לפרוט ו-$c=(c_1,\dots,c_n)\in\mathbb{N}$ מטבעות באמצעותם נפרוט.

פלט פריטה של k עם מס' מינימלי של מטבעות.

דוגמה $k=23$, $c=\{1,2,5,10\}$, $(10,10,2,1)$.

פסאודו-קוד נציע אלג' חמדן: בכל שלב ניקח את המטבע הכי גדול שקטן מהשארית ונוסיף אותו לפריטה.

דוגמה נגדית עבור $k=6$, $c=\{1,3,4\}$, האלג' יפרוט ל-$\{4,1,1\}$ ואילו הפתרון האופטימלי הוא $\{3,3\}$.

מסקנה לא כל בעיה ניתן לפתור באמצעות אלג' חמדן!

שבוע $\mathbb{III}$ | עוד דוגמאות לאלגוריתמים חמדניים

הרצאה

חלק א' של ההרצאה

משפט הפתרון המוחזר ע"י האלג' החמדן לבעיית התרמיל השברי הינו פתרון אופטימלי.

הוכחה: לבעיה זו יש פתרון אופטימלי. אנחנו מחפשים מקסימום של פ' הערך $f\left(x\right)=\sum_{i=1}^{n}x_iv_i$ במרחב הפתרונות החוקיים לבעייה, $\mathcal{S}=\left\{x\in\left[0,1\right]^n:\sum_{i=1}^{n}x_iw_i\le W\right\}$. מכיוון ש-$\mathcal{S}$ קבוצה קומפקטית (אינפי מתקדם, לא ממש חשוב לנו מה זה אומר) ו-f היא רציפה, f מקבלת מקסימום.

נשתמש בלמת ההחלפה כדי להראות שאם y הינו פתרון חוקי לבעיה השונה מפתרון החמדן אז y אינו אופטימלי. מספיק להוכיח כי y מקיים את התנאים של למת ההחלפה. מכך נסיק שאם כל פתרון שאינו החמדני הוא לא אופטימלי אבל לפ' הערך יש מקסימום, אז הפתרון החמדני הוא אכן המקסימום.

לפתרון החמדני נגיע באופן הבא: יהי $0\le t\le n-1$ כך ש-$\sum_{i=1}^{t}w_i\le W$, $\sum_{i=1}^{t+1}w_i>W$, אזי $x_1=\ldots=x_t=1$. יהי j המקום הראשון עבורו $x_j\neq y_j$. נראה כי $y_j<x_j$. נניח בשלילה כי $y_j>x_j$. מתקיים בהכרח כי $y_i\le x_i=1$ עבור $1\le i\le t$. לכן $j\ge t+1$ ומתקיים $W=\sum_{i=1}^{t+1}x_iw_i\ge\sum_{i=1}^{j}x_iw_i>\sum_{i=1}^{j}y_iw_i$ סתירה.

לכן $y_j<x_j\le 1$ ובנוסף $\sum_{i=1}^{j}y_iw_i<\sum_{i=1}^{t}x_iw_i\le W$ ולכן y מקיים את תנאי למת ההחלפה ולכן לא אופטימלי. ■

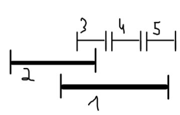
בעיית שיבוץ משימות

סיפור מסגרת נתונות n משימות כך שלא ניתן לבצע שתי משימות. במקביל המטרה היא לבחור כמה שיותר מתוך המשימות האלה שנוכל לבצען ללא התנגשויות.

קלט n קטעים סגורים על ציר זמן הניתנים ע"י נקודות ההתחלה והסיום שלהם $\left[s_1,f_1\right],\ldots,\left[s_n,f_n\right]$.

פלט תת קבוצה $S\subseteq\left[n\right]$ של אינדקסים כך שהקטעים עם אינדקסים ב-S זרים זה לזה וכך ש-$\left|S\right|$ מקסימלי.

דוגמה $s_1^*=\left\{3,4,5\right\}$, $s_2^*=\left\{2,4,5\right\}$ הם פתרונות אופטימליים לבעיה הבאה



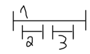
מרחב הפתרונות החוקיים לבעיה הוא $\mathcal{S}=\left\{S\subseteq\left[n\right]:\right.$ הקטעים עם אינדקסים ב-S זרים זה לזה$\left.\right\}$. נגדיר פ' ערך על הקבוצה $\mathcal{S}$ באופן הבא: $q\left(s\right)=\left|S\right|$, נחפש מקסימום s^* של q על S.

פתרון נאיבי נעבור על כל $\mathcal{S}$ ובאופן כזה נמצא את המקסימום של q. זמן הריצה הוא $\Omega\left(\left|\mathcal{S}\right|\right)$, שיכול להיות $\Omega\left(2^n\right)$.

הצעות לפתרונות חמדניים

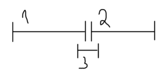
1. בכל שלב נבחר קטע שנחתך עם הכי פחות קטעים שנשארו (ולא מתנגש עם הקטעים שכבר בחרנו).

   לא אופטימלי, $G=\left\{3\right\}$, $s^*=\left\{1,2\right\}$.



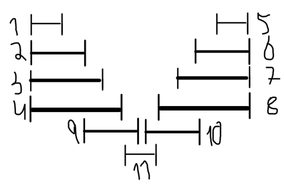
2. בכל שלב נבחר את הקטע הקצר ביותר.

לא אופטימלי, $G=\{3\}$, $s^*=\{1,2\}$.



3. בכל שלב נבחר את הקטע שזמן ההתחלה שלו הוא הקרוב ביותר לזמן הסיום של המשימה האחרונה שכבר בחרנו.

לא אופטימלי, $G=\{11,1,5\}$, $s^*=\{1,5,9,10\}$.



4. בכל שלב נבחר משימה שזמן הסיום שלה מינימלי.

זהו הפתרון הנכון (מבין רבים אחרים)!

פסאודו-קוד האלג' החמדן ע"פ ההצעה הרביעית.

0. עיבוד מידע מוקדם: נמיין את הקטעים בסדר עולה לפי זמני הסיום שלהם. $\Theta\left(n\log n\right)$.
1. אתחול: $A=[n]$, $G=\varnothing$. A יכיל את האינדקסים של המשימות שעוד אפשר להכניס ל-G (הפתרון החמדן). $\mathcal{O}\left(1\right)$.
2. איטרציה: בכל שלב נעביר מ-A ל-G את האינדקס הנמוך ביותר (הקטע המאוחר ביותר). נמחק מ-A את כל הקטעים הנחתכים עם הקטע שכרגע בחרנו. $\mathcal{O}\left(n^2\right)$ (באמצעות מימוש נאיבי של מבנה נתונים התומך ב-A - מערך. כאן המחיקה תדרוש $\Theta\left(n\right)$ - מעבר נאיבי על כל האיברים).
3. סיום: כאשר $A=\varnothing$ נעצר ונחזיר את G. $\mathcal{O}\left(1\right)$.

זמן ריצה סה"כ $\mathcal{O}\left(n^2\right)$.

חלק ב' של ההרצאה

למת החלפה יהי $s=(i_1,\dots,i_m)$ פתרון חוקי לבעיה, $0\le j\le m-1$ ו-t האינדקס המינימלי של קטע שלא חותך את הקטעים $i_1,\dots,i_j$ אזי גם $s'=(i_1,\dots,i_j,t,i_{j+2},\dots,i_m)$ הוא פתרון חוקי.

הוכחה: עלינו להראות שהקטעים ב-s' זרים זה לזה. מכיוון ש-s הוא פתרון חוקי, הקטעים $i_1,\dots,i_j,i_{j+2},\dots,i_m$ זרים זה לזה. בנוסף, לפי בחירת t, הקטע t זר לקטעים $i_1,\dots,i_j$. נוודא כי לכל $j+2\le k\le m$ מתקיים כי t,i_k זרים זה לזה.

נבחין כי אם $[s,f],[s',f']$ שני קטעים זרים כך ש-$f' > f$ אזי $s' > f$ כי אחרת הנקודה f הייתה בשני הקטעים בסתירה להנחה שהם זרים.

נשים לב שהקטעים i_k, i_{j+1} שייכים לפתרון חוקי ולכן זרים לזה לזה. בנוסף, $i_k < i_{j+1}$ ולכן $f_{i_k} < f_{i_{j+1}}$. לפי ההבחנה הנ"ל זה גורר כי $s_{i_k} < f_{i_{j+1}}$. בנוסף, הקטע i_{j+1} זר לקטעים $i_1,\ldots,i_j$ (כי כל הקטעים האלה נמצאים איתו בפתרון החוקי s). נזכור כי הקטע t נבחר כקטע בעל אינדקס מינימלי שלא חותך את הקטעים $i_1,\ldots,i_j$ ולכן מתקיים $t \le i_{j+1}$ ולכן $s_{f_t} \le f_{i_{j+1}}$ ולכן הקטעים t ו-i_j זרים לזה לזה. ■

משפט הפתרון החמדן הינו פתרון חוקי ואופטימלי לבעיית שיבוץ המשימות.

הוכחה: חוקיות: נשים לב שלפי דרך פעולות של האלג' החמדני, לפני תחילת כל איטרציה, הקבוצה A מכילה רק קטעים שזרים לכל הקטעים ב-G ולכן בסוף האיטרציה הקבוצה G מכילה קטעים זרים (הינה פתרון חוקי). זה המצב גם בסיום הריצה של האלג' ולכן הפתרון G המוחזר ע"י האלג' הוא חוקי.

אופטימליות: יהי $G=(g_1,\ldots,g_r)$ הפתרון החמדן כך ש-$g_1<\ldots<g_r$. עבור פתרון חוקי $s=(i_1,\ldots,i_k)$ לבעיה (אינדקסים ממוינים גם כן). נגדיר את גודל הרישא המקסימלית של x ו-G באופן הבא (סתם פורמליקה, בעצם מדובר בכמות האיברים המשותפים לשני הפתרונות ברציפות מההתחלה):

1. אם $k \ge r$ ומתקיים $i_1=g_1,\ldots,i_r=g_r$ אז במקרה זה נאמר כי לשני הפתרונות רישא משותפת בגודל r.
2. אם קיים $0 \le l < r$ כך שמתקיים $i_1=g_1,\ldots,i_l=g_l$ אבל $i_{l+1}\neq g_{l+1}$ נאמר כי לשני הפתרונות רישא משותפת בגודל l.

יהי s^* פתרון אופטימלי שלו רישא משותפת מקסימלית (מבין כל הפתרונות האופטימליים) עם G. נוכיח כי $s^*=G$ ובכך נוכיח כי G פתרון אופטימלי.

נראה כי הרישא המשותפת של s^* ו-G היא בגודל r. נניח בשלילה, לכן קיים $l<r$ כך ש-$i_1=g_1,\ldots,i_{l,}=g_l, i_{l+1}\neq g_{l+1}$. נפעיל על s^* את למת ההחלפה עם $j=l$, לכן ניתן להחליף את i_{l+1} ב-t שהוא האינדקס המינימלי של קטע שלא חותך את $g_1,\ldots,g_l$. מדרך פעולותו של האלג' החמדן, $t=g_{l+1}$ ולכן קיבלנו פתרון חוקי חדש $s=(g_1,\ldots,g_l,g_{l+1},i_{l+2},\ldots,i_m)$. מכיוון ש-$|s|=|s^*|=m$, s הוא גם פתרון אופטימלי אבל עם רישא משותפת עם G הגדולה מזו של s^* בסתירה להגדרת s^*.

נוכיח כי $G=s^*$. הראינו כי $G\subseteq s^*$. נניח בשלילה כי קיים קטע ב-s^* שלא נמצא ב-G, לכן $i_{r+1}\notin G$. נשים לב כי לפי דרך פעולות של האלג' החמדן, הקטע i_{r+1} שייך לקבוצה A בסוף כל איטרציה של האלג' ולכן $i_{r+1}\in A$ גם בסוף הריצה של האלג' החמדני אבל האלג' החמדני עוצר כאשר $A\neq\varnothing$ והיה אמור להוסיף את הקטע הזה סתירה. ■

תרגול

הגדרה מטרואיד הוא זוג $M=\langle S,I\rangle$ כאשר S היא קבוצה סופית ו- $I\subseteq 2^S$ כך שמתקיים:

1. I לא ריקה.

2. (תורשתיות) $\forall A \in I$ ו-$B \subseteq A$ מתקיים $B \in I$.

3. (החלפה) עבור $A, B \in I$ כך ש-$|B| > |A|$ אז קיים $a \in A \backslash B$ כך ש-$B \cup \{a\} \in I$.

דוגמאות למטרואידים

1. $S = [2]$, $I = \{\{1\}, \{2\}, \{1,2\}\}$. $M = \langle S, T \rangle$. לא מטרואיד, כי $\{2\} \subseteq \varnothing$ ו-$\varnothing \notin I$.

2. $S = [4]$, $I = \{\varnothing, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{3,4\}\}$ לא מטרואיד כי עבור $A = \{3,4\}$ ו-$B = \{2\}$ לא קיים $a \in A \backslash B$ כך ש-$B \cup \{a\} \in I$.

3. המטרואיד הוקטורי $M_V = \langle S_V, I_V \rangle$ מוגדר כך ש-S היא קבוצה סופית של וקטורים במ"ו V. I קבוצה של תתי קבוצות וקטורים ב-S כך שכל תת קבוצה מורכבת מוקטורים בת"ל.

4. המטרואיד הגרפי $M_G = \langle S_G, I_G \rangle$ עבור הגרף $G = \langle E, V \rangle$ כך ש-S הוא אוסף הצלעות בגרף ו-

$$I = \{E' \subseteq E : \langle V, E' \rangle \text{ is a forest } \}$$

טענה המטרואיד הגרפי הוא מטרואיד.

הוכחה: I לא ריקה כי $\langle V, \varnothing \rangle \in I$ כי אין בו מעגלים.

$\forall A \in I$ מתקיים שאין מעגלים ב-A כי הוא יער ולכן $\forall B \subseteq I$ הוא גם כן חסר מעגלים ולכן $B \in I$.

עבור $A, B \in I$ כך ש-$|B| > |A|$ נראה כי קיים $a \in A \backslash B$ כך ש-$B \cup \{a\} \in I$. כלומר ש-$B \cup \{a\}$ חסר מעגלים. הוספת צלע לגרף סוגרת מעגל או מקטינה את מס' רכיבי הקשירות. ■

טענת עזר קיים רכיב קשירות ב-G_A שחותך (שיש צלע בין שני רכיבים הקשירות) לפחות 2 רכיבי קשירות ב-G_B.

הוכחה: נניח בשלילה שלא קיים קיים רכיב קשירות כזה. אז כל רכיב קשירות ב-G_A חותך לכל היותר רכיב קישרות אחד ב-G_B, כלומר הוא מוכל ברכיב קשירות ב-G_B. מכאן שכל רכיב קשירות ב-G_B מכיל לפחות רכיב קשירות אחד של G_A. נקבל שמספר רכיבי ב-G_A גדול ממספר רכיבי הקשירות ב-G_B שזו סתירה לכך שב-G_B יש יותר רכיבי קשירות. ■

מסקנה קיים רכיב קשירות ב-G_A שחותך 2 רכיבי קשירות ב-G_B ולכן קיימת צלע ב-A שמחברת שני רכיבי קשירות ב-G_B ובפרט לא סוגרת מעגל. נסמן את הצלע ב-a ונקבל כי $B \cup \{a\}$ לא מכיל מעגל ולכן נמצא ב-I.

בעיית המטרואידים

קלט מטרואיד $M = \langle S, I \rangle$ ופ' משקל $w : S \to \mathbb{R}^+$.

פלט $A \in I$ כך ש-A בעלת משקל מקסימלי.

פסאודו-קוד

1. עיבוד מקדים: נמיין את האיברים ב-S ע"פ משקלם בסדר יורד.
2. אתחול: $A = \varnothing$.
3. איטרציה: נעבור על איבריי S ונוסיף את האיבר $x \in S$ ל-A אם $A \cup \{x\} \in I$.

הערות

1. הפלט של האלג' בגודל מקסימלי (אחרת היינו יכולים להוסיף עוד את אחד האיברים).
2. בעית המינימום שקולה אם נמיין את איברי S בסדר עולה והפלט יהיה קבוצה בגודל מקסימלי שהיא בעלת משקל מינימלי.
3. לא תמיד חייבים לעבור על כל האיברים ב-S (לדוגמה אם נתון אילוץ נוסף על האיברים).
4. אם בעיה שקולה לבעיה שראינו, נוכל להשתמש באלג' החמדן שראינו מבלי להוכיח את נכונותו

הגדרה שידוך בגרף דו"צ הוא קבוצת צלעות כך שאף שתי צלעות לא נפגשות באותו קודקוד. שידוך מושלם הוא שידוך על כל קודקודי הגרף.

הגדרה עבור $G = \langle L, R, E\rangle$ גרף דו"צ, מטרואיד השידוכים הוא $M = \langle S, I\rangle$ כך ש-$S = L$ ו-

$$I = \{L' \subseteq L : \exists R' \subseteq R, E' \subseteq E \ s.t.\ G' = \langle L', R', E'\rangle \text{ הוא שידוך מושלם ב- } E'\}$$

טענה מטרואיד השידוכים הוא מטרואיד.

הוכחה: ברור כי I לא ריקה כל עוד יש שידוך כלשהו ב-G.

תכונת התורשתיות גם היא ברורה, אם $B \subseteq A \in I$ אז קיים R', E' כך שקיים ב-$\langle A, R', E'\rangle$ שידוך מושלם. B מתקבל מ-A ע"י הסרה של קודקודים מאחד מצדי הגרף. לכן נסיר מ-R' את כל הקודקודים שמחוברים בצלעות לקודקודים שהסרנו, וכמובן נסיר את הצלעות לקודקודים שהסרנו, ונקבל גרף חדש, $\langle B, R'', E''\rangle$ שבו יש שידוך מושלם.

תכונת ההחלפה היא יותר מורכבת. תהיינה $A, B \in I$, $|A| > |B|$ ונוכיח כי קיים $a \in A \backslash B$ כך ש-$B \cup \{a\} \in I$. נביט ב-$G = G_A \cup G_B$ (הגרף שמוגדר ע"י איחוד אוספי הצלעות E'_A ו-E'_B יחד עם כל קודקודי G. נשים לב כי הדרגה של כל קודקוד היא לכל היותר 2 (הקודקוד מופיע לכל בשידוך גם של A וגם של B שאלו שתי צלעות) ולכן מסלול מקודקוד מסויים יכול להיות אחד ורק אחד מהבאים:

1. מעגל: במצב זה נקבל צלעות שמתחלפות לסירגוין בין צלעות ב-G_A ולבין צלעות ב-G_B.
2. מסלול באורך זוגי: במקרה זה הצלע הראשונה והאחרונה בהכרח יהיו מגרפים שונים (אחת מ-G_A ואחת מ-G_B).

3. מסלול באורך אי זוגי: במצב זה הצלע הראשונה והאחרונה בהכרח יהיו מאותו הגרף (G_A או G_B).

נשים לב כי בכל גרף חייב להיות מסלול באורך אי זוגי שמתחיל (ונגמר) בצלעות מ-G_A, אחרת נקבל שכל המעגלים והמסלולים הם עם כמות שווה של צלעות משתי הקבוצות ולכן סה"כ נקבל שוויון בכמות הצלעות ב-G_A ו-G_B וזאת בסתירה לכך שב-A יש יותר קודקודים (ולכן יותר צלעות בשידוך שהוא הגרף G_A). בפרט, המסלול הזה מתחיל (או נגמר, תלוי איך מסתכלים על זה) בקודקוד ב-L. נסמן את הקודקוד ב-$a \in A$. נשים לב כי $a \notin B$ כי אז $\deg a = 2$ ואז a הוא לא סוף או התחלה של מסלול אלא אמצע שלו. לכן $a \in A \backslash B$. נותר להוכיח כי $B \cup \{a\} \in I$. השידוך המושלם על הקבוצה החדשה יוגדר באופן הבא: נשאיר את כל הצלעות של השידוך של B ורק את אלה ששיכות למסלול האי זוגי הנ"ל נחליף. שם, במקום להשתמש בצלעות של השידוך של B, נשתמש בצלעות של A שמרכיבות את המסלול, וזה איכשהו יסתדר. ■

שבוע $\mathbb{III}$ | הכללות לאלגוריתם החמדן

הרצאה

חלק א' של ההרצאה

קבוצת וקטורים בת"ל בעלת משקל מקסימלי

קלט קבוצה סופית $X = \{v_1, \ldots, v_n\} \subseteq V$ כאשר V מ"ו ממימד t ופ' משקל $\mu : X \to \mathbb{R}_+$ חיובית.

פלט $S \subseteq [n]$ כך שהוקטורים עם אינדקסים ב-S וכך ש-$\mu(S)$ מירבי כאשר $\mu(S) = \sum_{i \in S} \mu(v_i)$.

הערה בעיה זו היא הכללה של קרוסקל כאשר וקטורים הם בת"ל אם"ם הם סוגרים מעגל.

דוגמה $n = 5, V = \mathbb{R}^4$.

$$\begin{pmatrix} 1 & 1 & 1 & 3 & e \\ 2 & -1 & \sqrt{2} & 2+2\sqrt{2} & \pi \\ 3 & 1 & \sqrt{3} & 3+2\sqrt{3} & -\sqrt{e} \\ 4 & -1 & 2 & 8 & -\pi \end{pmatrix}$$

עם משקלים $5, 3, 7, 4, 2$ בהתאמה, $S^* = \{3, 1, 2, 5\}$ כי $v_4 = v_1 + 2v_3$.

פסאודו-קוד

0. <u>עיבוד מידע מוקדם</u>: נמיין את הוקטורים על פי משקלם בסדר יורד (מעתה נניח בה"כ כי $\mu(v_1) \geq \cdots \geq \mu(v_n)$). $\Theta(n \log n)$.
1. <u>אתחול</u>: $A = [n]$, $G = \varnothing$. $\mathcal{O}(1)$.
2. <u>איטרציה</u>: בכל שלב נעביר מ-A ל-G את האידנקס המינימלי ב-A ונמחק מ-A את כל הוקטורים התלויים לינארית בוקטורים ב-G.
n כפול זמן הריצה להקביעה האם וקטור מסויים ת"ל בקבוצה נתונה של וקטורים. אפשר לעשות זאת ב-$\Theta(n^2)$.
3. <u>עצירה</u>: כאשר $A = \varnothing$ נעצור ונחזיר את G. $\mathcal{O}(1)$.

זמן ריצה סה"כ $\mathcal{O}\left(n^3\right)$ כאשר חישוב הת"ל הוא ב-$\Theta\left(n^2\right)$.

למת החלפה (למת החלפה מטרואידלית) יהי V מ"ו ותהיינה $X, Y \subseteq V$ קבוצות סופיות בת"ל כך ש-$|X| < |Y|$. אזי קיים $y \in Y \backslash X$ כך ש-$X \cup \{y\}$ בת"ל.

הוכחה: נסמן $U = \mathrm{sp}X$. X בת"ל ולכן $\dim U = |X|$. מכיוון ש-$|X| < |Y|$ ו-Y בת"ל, מתקיים $Y \nsubseteq U$ ולכן קיים $y \in Y$ כך ש-$y \notin U$, כלומר y אינו ק"לשל וקטורים ב-X. לכן $y \notin X$ ו-$X \cup \{y\}$ בת"ל. ■

משפט האלג' החמדן מחזיר פתרון אופטימלי לבעיה.

הוכחה: יהי G הפתרון החמדן. יהי s^* פתרון אופטימלי. נראה כי $|G| = |s^*|$. נניח בשלילה.

אם $|G| < |s^*|$ אז מלמת ההחלפה קיים $s \in s^* \backslash G$ כך ש-$G \cup \{s\}$ בת"ל ולכן s אינו ת"ל ב-G, לכן לפי פעולתו של האלג' החמדן, הוקטור נשאר בקבוצה A לאחר כל איטרציה של האלג' ולכן גם בסיום הריצה של האלג' והאלג' עצר כאשר $A \neq \varnothing$ אבל $s \in A$ סתירה.

אם $|S^*| < |G|$ אז מלמת ההחלפה קיים $g \in G \backslash s^*$ כך ש-$s^* \cup \{g\}$ בת"ל ולכן $s_1 = s^* \cup \{g\}$ הוא פתרון חוקי שמשקלו

$$\mu\left(s_1\right) = \mu\left(s^*\right) + \mu\left(g\right) > \mu\left(s^*\right)$$

בסתירה לאופטימליות s^*. נסמן $k = |G| = |s^*|$, $G = \left(g_1, \ldots, g_k\right)$, $s^* = \left(i_1, \ldots, i_k\right)$ ממוינים בסדר יורד לפי משקל.

נניח בשלילה כי $\mu\left(s^*\right) > \mu\left(g^*\right)$. לכן, מתקיים

$$\mu\left(s^*\right) = \sum_{j=1}^{k} \mu\left(i_j\right) > \sum_{j=1}^{k} \mu\left(g_j\right) = \mu\left(G\right)$$

ולכן קיים $1 \leq j \leq k$ כך ש-$\mu\left(i_j\right) > \mu\left(g_j\right)$.

נסמן $X = \{g_1, \ldots, g_{j-1}\}$, $Y = \{i_1, \ldots, i_j\}$. X, Y בת"ל כי הן תת קבוצות של פתרונות חוקיים לבעיה, ו-$|X| < |Y|$ ולכן מלמת ההחלפה, קיים $y \in Y \backslash X$ כך ש-$X \cup \{y\}$ בת"ל. כלומר, קיים $1 \leq t \leq j$ כך שהוקטור $i_t \in s^*$ אינו ת"ל ב-$\{g_1, \ldots, g_{j-1}\}$ ולכן i_t היה ניתן לבחירה באיטרציה ה-j של האלג' החמדן.

בנוסף, מכיוון ש- $t \leq j$ מתקיים $\mu\left(i_t\right) \geq \mu\left(i_j\right) > \mu\left(g_j\right)$ ולכן האלג' החמדן באיטרציה ה-j שלו בחר וקטור שמשקלו אינו מקסימלי מבין כל הוקטרים ב-A, בסתירה לדרך פעולותו. ■

חלק ב' של ההרצאה

בעיה אלגוריתמית גנרית - מטרואידים

קלט קבוצה סופית של אובייקטים $B = \{x_1, \ldots, x_n\}$, פ' משקל $\mu : B \to \mathbb{R}_+$, משפחה I של תת קבוצות של B (שמייצגת את מרחב הפתרונות החוקיים לבעיה).

פלט תת קבוצה s של B, $s \in I$ כך ש-$\mu(s)$ מקסימלי.

דוגמאות למטרואידים

1. קבוצת וקטורים בת"ל בעלת משקל מקסימלי. $B = \{v_1, \dots, v_n\}$, $I = \{S \subseteq [n] :$ הוקטורים עם אינדקסים ב-S הם בת"ל$\}$, פ' המשקל μ כפי שמוגדרת בבעיה.

2. שיבוץ משימות. $B = \{[s_1, f_1], \dots, [s_n, f_n]\}$, $I = \{S \subseteq [n] :$ הוקטורים עם אינדקסים ב-S זרים זה לזה$\}$ ונגדיר פ' משקל $\mu \equiv 1$.

וריאציה על שיבוץ משימות: אותם נתונים, הפעם עם $\forall i \in [n]$, $\mu(i) = f_i - s_i$, תהיה הבעיה שממקסמת את סה"כ זמן ביצוע המשימות לעומת מספר משימות.

שבוע $\mathbb{IV}$ | תכנון דינמי

הרצאה

חלק א' של ההרצאה

דוגמה בעיית התרמיל השלם: $(v_1, w_1), \dots, (v_n, w_n)$, W המשקל המירבי של התרמיל, $B = [n]$, $I = \left\{S \subseteq [n] : \sum_{i \in S} w_i \in W\right\}$, $\mu(x_i) = v_i$. (בניגוד לשברי, נכניס את כל האיבר או שלא נכניס אותו בכלל).

אלגוריתם חמדן גנרי לפתרון הבעיה

פסאודו-קוד

0. עיבוד מידע מקדים: נמיין את הפריטים לפי משקלם בסדר יורד.
1. אתחול: $A = [n]$, $G = \varnothing$.
2. איטרציה: בכל שלב נעביר מ-A ל-G את האינדקס הנמוך ביותר (בחירה חמדנית) ונמחק מ-A את כל האינדקסים x כך ש-$I \not\ni G \cup \{x\}$ (חוקיות).
3. סיום: כאשר $A = \varnothing$ נעצור ונחזיר את G.

זמן ריצה נסמן T זמן ריצה לפתרון של הבעיה האלגוריתמית "בהינתן תת קבוצה $S \subseteq B$, האם $S \in I$?". לכן זמן הריצה של האלג' הגנרי הוא $\Theta(n \log n + n \cdot T)$.

משפט אם (B, I) הוא מטרואיד, אזי האלג' החמדן הגנרי מחזיר פתרון אופטימלי לבעיה לכל פ' משקל.

הוכחה: בתרגיל. ■

משפט תהי I משפחה של תת קבוצות של B הינה לא ריקה ותורשתית אבל לא מקיימת את תכונת ההחלפה אזי קיימת פ' משקל $\mu : B \to \mathbb{R}_+$, כך שהאלג' החמדן הגנרי לא מחזיר פתרון אופטימלי לבעיה המתאימה.

הוכחה: מהנתון, קיימות $S, T \in I$ כך ש-$|S| < |T|$ אבל $\forall t \in T \backslash S$ מתקיים $S \cup \{t\} \notin I$. נבנה פ' משקל μ ע"י

$$\mu(x) = \begin{cases} 1 & x \in S \\ 1-\epsilon & x \in T \backslash S \\ \epsilon & x \notin S \cup T \end{cases}$$

עבור $\epsilon = \frac{|T|-|S|}{2|B|}$ (בעיקרון פשוט ממש קטן). רעיון ההוכחה הוא שהאלג' החמדן יבחר את כל S ואז לא יוכל לקחת מ-T איברים כי תכונת ההחלפה לא מתקיימת, ואז יאלץ לבחור לכל היותר איברים עם משקל ϵ (קטן מאוד). לעומת זאת, פתרון אופטימלי יקח את כל האיברים מ-T ולא שום דבר אחר מ-S או שאר האיברים ויקבל משקל יותר גבוה.

לפי דרך פעולותו של האלג' החמדן הגנרי, מכיוון ש-$S \in I$ ו-I היא משפחה תורשתית, ב-$|S|$ האיטרציות הראשונות שלו האלג' החמדן יכניס ל-G את כל אברי הקבוצה S. בשלב העדכון של האיטרציה ה-$|S|$, האלג' ימחק מ-A את כל האיברים ב-$T \backslash S$ (כי $\forall t \in T \backslash S$, $S \cup \{t\} \notni I$) ולכן

$$\mu(G) \leq \mu(S) + \mu\left((S \cup T)^C\right) = |S| + \epsilon\left|(S \cup T)^C\right| < |S| + \epsilon|B| = |S| + \frac{|T|-|S|}{2|B|}|B| = \frac{|S|+|T|}{2}$$

מצד שני,

$$\mu(T) \geq (1-\epsilon)|T| = |T| - \epsilon|T| \geq |T| - \epsilon|B| = |T| - \frac{|T|-|S|}{2|B|}|B| = |S| + |T| > \mu(G)$$

ולכן קיים פתרון חוקי לבעיה (הקבוצה T) שמשקלו גדול יותר מזה של הפתרון החמדן. כלומר, האלג' החמדן הגנרי לא מחזיר פתרון אפטימלי.

■

הערה מהמשפט הראשון ניתן להסיק שהמשפחות I בבעית שיבוץ משימות ובעית התרמיל השלם הן (כנראה) אינן מטרואידים. אכן, עבור קבוצה שמכילה שני קטעים זרים וקבוצה נוספת שמכילה איבר אחד שחותכת את שני הקטעים הנ"ל, נקבל שתכונת ההחלפה של מטרואידים לא מתקיימת.

אלגוריתמי תכנון דינמי

עקרון אלגוריתמי חדש: נחלק את הבעיה הנתונה לתת בעיות. נפתור את תת הבעיות ונמזג את הפתרונות לפתרון הבעיה כולה. בשפת רחוב "הפרד ומשול".

דוגמה חישוב פ' המוגדרת באופן ריקורסיבי. נתונה פ' $T\left(a,b\right)$ של שני פרמטרים טבעיים המוגדרת ע"י

$$T\left(a,b\right)=\begin{cases}1 & a=0\ \vee\ b=0\\ T\left(a-1,b\right)+T\left(a,b-1\right) & \text{otherwise}\end{cases}$$

פתרון נאיבי ריקורסיה, $T\left(2021,2021\right)$ מתפצל ל-$T\left(2021,2020\right)$ ו-$T\left(2020,2021\right)$ וכן הלאה. עומק העץ הוא 2021 ולכן זמן הריצה הוא $\Omega\left(2^{2021}\right)$.

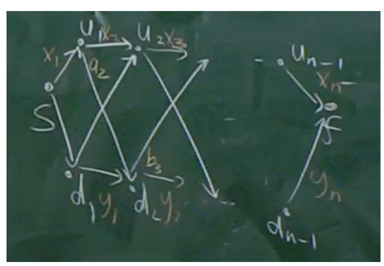
מה הן תת הבעיות השונות שנצטרך לפתור לאורך הריצה של הריקורסיה? נשים לב כי כל תתי הבעיות הללו הן חישוב של $T\left(a,b\right)$, עבור $0\leq a,b\leq 2021$ ומספרן הוא $\left(2021\right)^{2}<\left(2048\right)^{2}=2^{22}$ שזה כבר הרבה פחות.

לחלופין, כדי לפתור כל בעיה חלקית רק פעם אחת, נבנה טבלה בגודל n^2 עם $n=2021$, בה במקום ה-ij נשבץ את הפתרון לבעיה $T\left(i,j\right)$. במצב זה, זמן הריצה של האלג' יורד ל-$\Theta\left(n^{2}\right)$ (נצטרך לפתור רק n^2 חיושבים יחודיים).

בעית ניתוב משימות

סיפור מסגרת פריט עובר תהליך ייצור. יש שני פסי ייצור זהים, על כל פס מספר תחנות עבודה. המחירים של מעבר מתחנה לתחנה שונים זה מזה. נרצה למצוא מסלול לפריט בין תחנות העבודה כדי למזער את המחיר הכולל של תהליך הייצור.

קלט $x_1,\ldots,x_n$, $y_1,\ldots,y_n$, $a_2,\ldots,a_{n-1}$, $b_2,\ldots,b_{n-1}$ כאשר הנתונים הנ"ל מתאימים לגרף הבא



כלומר x_i המחיר המעבר מהתחנה ה-u_{i-1} לתחנה ה-u_i, y_i המעבר מ-d_{i-1} ל-d_i, a_i המעבר בין u_{i-1} ל-d_i ו-b_i המעבר מ-d_{i-1} ל-u_i.

פלט מסלול מ-s ל-f בעל מחיר מינימלי.

- אלג' חמדן טבעי שבכל פיצול מעדיף את הכיוון הזול יותר - לא יעבוד (הסטודנטית המשקיעה תבין למה).
- הפתרון הנאיבי הוא לעבור על כל המסלולים. גודל מרחב הפתרונות החוקיים פה הוא 2^{n-1} ולכן המעבר על כל האפשרויות יקר מדי ולא משתלם.
- האלג' של Dijkstra פותר את הבעיה הזו. במקרה זה, האלג' רץ בזמן $\mathcal{O}\left(n\log n\right)$.

נוכל לתכנן אלג' שרץ בזמן לינארי באמצעות תכנון דינמי. נבחין הבחנה קריטית - כל תת מסלול של מסלול במחיר מינימלי הוא בהכרח אופטימלי בעצמו. ככלל, הבחנה זו נקראת עקרון בלמן: תת פתרון של פתרון אופטימלי הוא אופטימלי בעצמו.

פתרון דינאמי

1. הגדרת תתי בעיות: נתבונן בפיצול של הבעיה הגדולה לשתי תתי בעיות לפי הצעד הראשון במסלול. אם בחרנו ללכת מ-s ל-u_1, נגיע לתת בעיה אנלוגית, שהפעם הצעד הבא יהיה או ל-u_2 או ל-d_2. נרצה לגלות את התלות של המחיר האופטימלי של הבעיה הגדולה במחירים של תת הבעיות האלה. נסמן ב-p^* את המחיר של הבעיה הגדולה, ב-$p_u\left[k\right]$ את המחיר האופטימלי להגיע מ-u_k ל-f וב-$p_d\left[k\right]$ להגיע מ-d_k ל-f עבור $k\in\left[n-1\right]$. לכן מתקיים

$$p^*=\min\left\{x_1+p_u\left[1\right],y_1+p_d\left[1\right]\right\}$$

2. נוסחאת ריקורסיה: במהלך הפיצולים הבאים נגיע לכל תתי הבעיות שהן הגעה מכל אחת התחנות $d_1,\ldots,d_{n-1},u_1,\ldots,u_{n-1}$ ל-f. נוסחת הנסיגה למחיר אופטימלי לפתרון שמתחיל ב-u_k הוא

$$p_u\left[k\right]=\begin{cases}x_n & k=n-1\\ \min\left\{x_{k+1}+p_u\left[k+1\right],a_{k+1}+p_d\left[k+1\right]\right\} & k<n-1\end{cases}$$

והסטודנטית המשקיעה תנסח את הנוחסה האנלוגית ל-$p_d\left[k\right]$.

3. בניית טבלה: נבנה טבלה $n\times 2$ שבשורה עליונה נשבץ את $p_u\left[i\right]$ ובתחתונה את $p_d\left[i\right]$ (ובעמודה ה-0 את p^*).

מילוי הטבלה: נמלא את הטבלה בעמודות מימין לשמאל (i יורד), כאשר בעמודה ה-$(n-1)$ נרשום $\binom{x_n}{y_n}$. עבור $k<n-1$, בהינתן העמודה העמודה ה-$k+1$, נוכל למלא את העמודה ה-k לפי נוסחת הריקורסיה הכללית בזמן $\mathcal{O}\left(1\right)$.

חילוץ הפתרון: כדי למצוא את המסלול האופטימלי, נזכור בזמן המילוי בכל תא בטבלה מהו הכיוון הנכון לבחירה בשלב זה.

4. ניתוח זמן ריצה: סה"כ נמלא כל תא בזמן $\mathcal{O}\left(1\right)$ ויש בטבלה $\mathcal{O}\left(n\right)$ תאים ולכן נוכל למלא את הטבלה כולה בזמן $\mathcal{O}\left(n\right)$. כשנמלא את הטבלה כולה, נחזיר את העמודה ה-0.

תרגול

דוגמה חישוב פיבונאצ'י. $a_n=a_{n-1}+a_{n-2}$. בתכנון דינמי, נשמור מערך עם ערכי פיבונאצ'י לכל אינדקס. זמן הריצה יהיה $\mathcal{O}\left(n\right)$ כי נפתור כל בעיה יחודית פעם אחת בלבד. למרות שרשמית, הקלט הוא מספר ולכן זמן הריצה הוא על $k=\log n$ ולכן לכאורה זמן הריצה הוא $\mathcal{O}\left(2^k\right)$, וזה שיפור לעומת הפתרון הנאיבי שהוא $\mathcal{O}\left(2^{2^k}\right)$ (שזה עדיין שיפור). לפיבונאצ'י יש נוסחה ולכן ישנו פתרון גם ב-$\mathcal{O}\left(1\right)$.

סכימה לתכנות דינאמי

1. הגדרת תתי בעיות.
2. הגדרת נוסחאת ריקורסיה (המקשרת בין תת הבעיות לבעיה הגדולה).
3. בניית טבלה - תיאור, מילוי וחילוץ פתרון.
4. ניתוח זמן ריצה.
5. הוכחת אופטימליות (אינדוקציה על מילוי הטבלה).

לוח משימות תכנות

סיפור מסגרת נרצה לסדר לוח זמנים לביצוע משימות קלות וקשות. אם מבצעים משימה קשה בשבוע כלשהו, חייבים לנוח בשבוע שלפניו. נרצה למקסם את הרווח על ביצוע המשימות.

קלט $\{l_i, h_i\}_{i=1}^{n}$ כאשר l_i הוא הרווח מביצוע משימה קלה בשבוע ה-i ו-h_i הרווח מביצוע משימה קשה בשבוע ה-i.

פלט רשימת משימות שמניבות את הרווח האופטימלי ל-n שבועות.

דוגמה השורה התחתונה היא פתרון אוטפימלי לבעיה (השבוע הראשון מתחיל במשימה קשה כי לא ביצענו משימה בשבוע שלפניו).

	1	2	3	4	5
l	10	10	15	15	5
h	5	5	15	20	10
	h	l	$\backslash$	h	l

פתרון דינאמי

1. תתי בעיות: מציאת רווח אופטימלי לשבוע ה-i, $i \in [n]$.
2. נוסחת ריקורסיה: נסמן $M[i]$ הפתרון האופטימלי לשבוע ה-i (סה"כ הרווח עד ל-i כולל). אם נבצע משימה קשה אז $M[i] = h_i + M[i-2]$ ואם נבצע קלה אז $M[i] = l_i + M[i-1]$. לכן נוסחת הנסיגה היא

$$M[i] = \begin{cases} 0 & i = 0 \\ \max\{h_1, l_1\} & i = 1 \\ \max\{h_i + M[i-2], l_i + M[i-1]\} & i > 1 \end{cases}$$

3. הגדרת טבלה: נבנה טבלה בגודל $1\times(n+1)$. נמלא את הטבלה משמאל לימין ונמלא כל תא בעזרת נוסחת הריקורסיה.

חילוץ הפתרון: בכל תא בטבלה נשמור בנוסף את סוג המשימה שהובילה לערך האופטימלי לאותו שבוע (אם עכשיו בחרנו h אז לפניכן נשמור $\varnothing$ ואם בחרנו l אז נשמור את המשימה של השבוע הקודם). בסיום מילוי הבטלה נחלץ מ-$i=n$ עד $i=0$ ונכניס את המשימה שביצענו באופן הבא:

- אם ביצענו משימה קלה בשבוע ה-i, נוסיף אותה ונמשיך לשבוע $i-1$.
- אם ביצענו משימה קשה בשבוע ה-i, נוסיף אותה ונמשיך לשובע ה-$i-2$.

4. זמן ריצה: מילוי כל תא בטבלה לוקח $\mathcal{O}\left(1\right)$ ויש $\mathcal{O}\left(n\right)$ תאים ולכן זמן הריצה הכולל הוא $\mathcal{O}\left(n\right)$.

5. הוכחת אופט': נוכיח באינדוקציה על מילוי התא ה-i, כי התא ה-i מכיל את הרווח האופטימלי לשבוע ה-i.

בסיס ($i=0$): אין משימות לבצע ולכן הערך האופט' הוא 0.

בסיס* ($i=1$): אין אילוצי מנוחה ולכן הרווח האופטימלי הוא המקסימום בין האפשרויות.

צעד ($1,\ldots,i-1\to i$): בשבוע ה-i אפשר לבצע משימה קשה ואז הרווח יהיה $l_i+M\left[i-2\right]$ או משימה קלה ואז הרווח יהיה $l_i+M\left[i-1\right]$. מה"א התאים $M\left[i-2\right]$ ו-$M\left[i-1\right]$ מכילים ערך אופט' ולכן לקיחת המקסימום מבין 2 האפשרויות תחזיר רווח אופט'.

הערה בבעיות של תכנון דינאמי לרוב לא נתעסק יותר מדי בהוכחת הנכונות כי רוב ההוכחה היא "זה ככה כי זה ככה" ~ירדן יגיל 2021.

תת-מחרוזת משותפת מקסימלית

קלט 2 מחרוזות, $x=x_1,\ldots,x_n$, $y_1=y_1,\ldots,y_m$, $n,m\in\mathbb{N}$.

פלט תמ"א - **ת**ת **מ**חרוזת משותפת **א**רוכה ביותר (לא תכנית מתאר ארצית).

הערה נסמן x^i הרישא באורך i של x (ובאותו האופן על y).

פתרון דינאמי

1. תתי בעיות: מציאת תמ"א של x^i,y^i, $\forall i,\in j\in[n]$.
2. נוסחת הריקורסיה: נסמן $f\left(i,j\right)$ אורך התמ"א של x^i,y^i.

$$f\left(i,j\right)=\begin{cases}0 & i=0\lor j=0\\ f\left(i-1,j-1\right)+1 & x_i=y_j\\ \max\left\{f\left(i,j-1\right),f\left(i-1,j\right)\right\} & x_i\neq y_j\end{cases}$$

3. בניית טבלה: נבנה טבלה בגודל $\left(m+1\right)\left(n+1\right)$. נמלא את הטבלה לפי שורות מהתחתונה לעליונה (כאשר השורה התחתונה היא $j=0$ והעליונה $j=n$) וכל שורה משמאל לימין (שמאל $i=0$ וימין $i=n$) כאשר נמלא כל תא באמצעות נוסחת הריקורסיה.

חילוץ הפתרון: בכל תא נשמור גם מצביע אל התא שממנו קיבלנו את הערך לאותו התא. בסיום מילוי הטבלה נתחיל מהתא (m,n) ונתקדם לפי המצביעים ששמרנו - אם הגענו למצביע שהולך באלכסון אז נוסיף את האות לתמ"א (כי זה אומר שהנוסחה הריקורסיבי הלכה ל-$(i-1,j-1)$ ולכן $x_i=y_j$ ולכן היא חלק מהתמ"א).

4. זמן ריצה: מילוי כל תא לוקח $\mathcal{O}(1)$ ויש $\mathcal{O}(mn)$ תאים תאים, לכן זמן מילוי הטבלה לוקח $\mathcal{O}(mn)$. חילוץ הפתרון לוקח $\mathcal{O}(m+n)$ ולכן סה"כ זמן הריצה הוא $\mathcal{O}(mn)$.

5. הוכחת אופט': נוכיח באינדוקציה כי התא $M[i,j]$ מכיל את אורך התמ"א עבור x^i,y^i.

בסיס ($i=0$ או $j=0$): תמ"א עם מחרוזת ריקה היא תמיד באורך 0.

צעד (נניח שמילאנו עד כה באופן אופט'): תהי $s=s_1,\ldots,s_r$ תמ"א של x^i,y^i.

- $x_i=y_j$ ולכן $x_i=y_j=s_r$ כי אחרת היינו מקבלים סתירה לאופט' של s. נותר להסתכל על x^{i-1},y^{j-1}, ומה"א התא $M[i-1,j-1]$ מכיל את אורך התמ"א של x^{i-1},y^{j-1} ולכן ערכו $r-1$. קיבלנו כי

$$M[i,j]=M[i-1,j-1]+1=r-1+1$$

- $x_i\neq y_j$ ולכן $x_i\neq s_r$ או (לא בלעדי) $y_j\neq s_r$, לכן נסתכל על x^{i-1},y^j ועל x^i,y^{j-1}. מה"א התאים $M[i,j-1]$ ו-$M[i-1,j]$ מכילים את אורך התמ"א בהתאמה. לכן אם ניקח את הערך המקסימלי מבין השניים, נקבל את אורך התמ"א של x^i,y^j.

דוגמה $X=ABCD,Y=BDC$ כאן התמ"א היא BC.

C	0	$\leftarrow 0$	$1\ \downarrow$	$2\ \swarrow$	$\leftarrow 2$
D	0	$\leftarrow 0$	$1\ \downarrow$	$\leftarrow 1$	$2\ \swarrow$
B	0	$\leftarrow 0$	$1\ \swarrow$	$\leftarrow 1$	$\leftarrow 1$
$\varnothing$	0	0	0	0	0
	$\varnothing$	A	B	C	D

נעקוב אחרי החצים החל מהערך בפינה הימנית עליונה ונוסיף את האותיות שמגיעים אליהן דרך אלכסון, במקרה זה C ולפניו B.

שבוע $\mathbb{V}$ | תכנון דינמי גו נאטס

הרצאה

חלק א' של ההרצאה

בעית כפל מטריצות

תהי $A\in M_{n\times t}\left(\mathbb{R}\right)$ ו-$B\in M_{t\times m}\left(\mathbb{R}\right)$. נרצה לחשב את $AB=C\in M_{n\times m}\left(\mathbb{R}\right)$, כמה עולה לבצע זאת? כמה כפלים של מספרים לוקח כדי לחשב את C? $\forall i\in[n]$ ו-$\forall j\in[m]$, נחשב $C_{ij}=\sum\limits_{k=1}^{t}\left[A\right]_{ik}\left[B\right]_{kj}$ ע"י t כפלים. סה"כ כדי לחשב את כל C נצטרך $n\cdot t\cdot m$ כפלים.

הערה אם $t=m=n$ אז נצטרך לחישוב נאיבי n^{3} כפלים וכיום הכי מהיר שהוכח זה $n^{2.37\ldots}$.

עבור 3 מטריצות, $D=ABC$, מתקיים $D=\left(AB\right)C=A\left(BC\right)$ אבל מתברר שלסדר פעולות הכפל יש משמעות מבחינת המחיר.

דוגמה עבור $A\in M_{10\times50}\left(\mathbb{R}\right),B\in M_{50\times20}\left(\mathbb{R}\right),C\in M_{20\times100}\left(\mathbb{R}\right)$, אם נחשב $D=\left(AB\right)C$ נקבל

$$10\cdot50\cdot20+10\cdot20\cdot100=3\cdot10^{4}$$

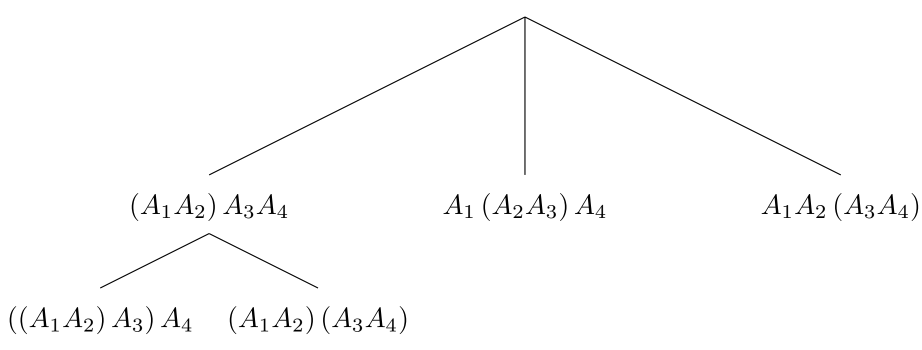
אם נחשב $D=A\left(BC\right)$ אז החישוב BC דורש $50\cdot20\cdot100=10^{5}$ שזה יותר מהחישוב הנ"ל.

קלט $n+1$ מספרים טבעיים $p_{0},\ldots,p_{n}$ המסמנים מימדים של n מטריצות $A_{1},\ldots,A_{n}$ כאשר $A_{i}\in M_{p_{i-1}\times p_{i}}\left(\mathbb{R}\right)$.

פלט סדר הכפלות מטריצות (חלוקת סוגריים) המשיג מחיר מינימלי לחישוב המכפלה $B=A_{1}\cdot\ldots\cdot A_{n}$.

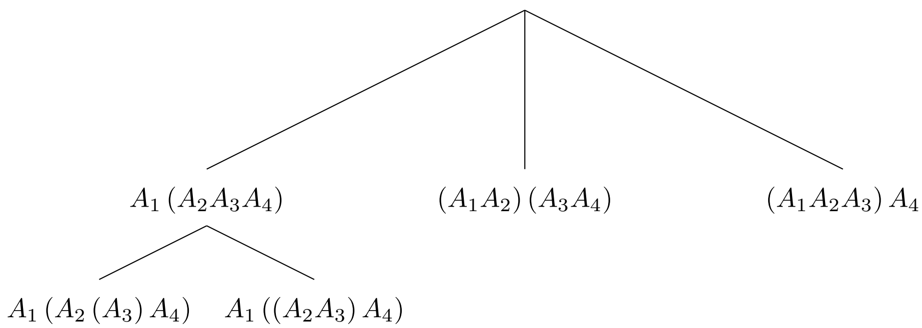
פתרון

- נאיבי - מעבר על כל האפשרויות הוא לא יעיל כי מרחב הפתרונות החוקיים הוא 4^{n} (מספר קטלן).
- חמדני - גם כנראה שלא.
- דינמי - כיצד נחלק את הבעיה הגדולה לתת בעיות?
  - לפי המכפלה הראשונה שמבצעים.



בעץ זה יש 5 קודקודים, ואילו הוא גדל באופן אקספוננציאלי בגלל שאנחנו כל פעם מכפילים מטריצות חדשות. נסביר (העשרה) למה זה גדל כל כך מהר. יש לנו $\frac{n}{2}$ זוגות לבצע בהם את המכפלה הראשונה. לאחר $\frac{n}{4}$ פיצולים, נקבל כל פעם בעיה חדשה שכן כל מכפלה יוצרת מטריצה חדשה ולכן סדרת מטריצות שונה. נוכל לבחור $\frac{n}{4}$ מתוך $\frac{n}{2}$ הזוגות הנ"ל ואלו יהיו זוגות יחודיים, כלומר $\binom{\frac{n}{2}}{\frac{n}{4}}\sim\Theta\left(\frac{2^{\frac{n}{2}}}{\sqrt{n}}\right)$ ולכן פתרון זה לא יעזור לנו.

  - לפי המכפלה האחרונה, עבור $1\le k\le n-1$, נחשב את תתי הבעיות הנובעות מ-$\left(A_{1}\ldots A_{k}\right)\left(A_{k+1}\ldots A_{n}\right)$.



בעץ זה יש 10 קודקודים (לא כל העץ מצויר) והוא גדל פולינומיאלית ומכיל מכפלת מטריצות ישנות בכל ירידה בעומק, וזהו הפתרון הרצוי.

פתרון דינאמי

1. הגדרת תתי בעיות: כל תת בעיה היא מהצורה לחשב $A_i\cdot\ldots\cdot A_j$ עבור $1\le i\le j\le n$ (ברגע שנפצל על הפעולה האחרונה, נקבל שני "קטעי" מטריצות לכפול, כל אחד מאלו הוא תת בעיה של רצף מטריצות, לדוגמה בפיצול הראשון תתי הבעיות יהיו $A_1\cdot\ldots\cdot A_k$ ו-$A_{k+1}\cdot\ldots\cdot A_n$). סה"כ הרצפים הללו (שכל אחד מתאים חחע"ל לתת בעיה) הוא $\binom{n}{2}+n=\Theta\left(n^2\right)$ (התוספת על העומק הרדוד ביותר - צריך לבחור איפה לחצות לראשונה את את ה-$n+1$ מספרים).
2. נוסחת ריקורסיה: נסמן ב-$P\left[i,j\right]$ את המחיר האופטימלי לחישוב $A_i\ldots A_j$. בפרט $P\left[1,n\right]$ הוא המחיר האופטימלי לחישוב $B=A_1\ldots A_n$. נוסחת הריקורסיה הכללית היא

$$P\left[i,j\right]=\begin{cases}0 & i=j\\ \min_{i\le k\le j-1}\left\{P\left[i,k\right]+P\left[k+1,j\right]+p_{i-1}p_kp_j\right\} & i<j\end{cases}$$

3. בניית טבלה: נגדיר טבלה T בגודל $n\times n$ ונרשום בכל תא $T\left[i,j\right]$, $1\le i,j\le n$ את $P\left[i,j\right]$.

מילוי הטבלה: נמלא $T\left[i,j\right]=0$, $\forall i>j$. נמלא את שאר הטבלה ב-n-איטרציות. באיטרציה $0\le d\le n-1$, נמלא את התאים $T\left[i,j\right]$ עבורם $j-i=d$ (כלומר האלכסון המרוחק ב-$n-d-1$ מהפינה השמאלית העליונה, ונמלא מהאלכסון המרכזי לכיוון צפון מערב - ראו דוגמה להמחשה) לפי נוסחת הריקורסיה.

חילוץ פתרון: נחזיר את $T\left[1,n\right]$. כדי לחלץ את חלוקת הסוגריים האופטימלית, נשמור בכל תא $T\left[i,j\right]$, $i<j$ בעת מילויו את ערך ה-k המשיג מינימום בנוסחת הריקורסיה.

4. ניתוח זמן ריצה: כל תא בטבלה מתמלא ב-$\mathcal{O}\left(n\right)$, כי כדי למלא תא נצטרך למצוא מינימום בין n איברים, שערכם חושב באיטרציות קודמות. סה"כ יש $\mathcal{O}\left(n^2\right)$ תאים ב-T ולכן מילוי כל הטבלה עולה $\mathcal{O}\left(n^3\right)$.
5. הוכחת נכונות:

טענה לכל $1\le i\le j\le n$ מתקיים $T\left[i,j\right]=P\left[i,j\right]$.

הוכחה: נוכיח באינדוקציה על $d=j-i$.

בסיס ($d=0$): כלומר $i=j$ ובמקרה זה $T\left[i,j\right]=0=P\left[i,j\right]$.

צעד $(d-1 \to d)$: מההגדרה,

$$T\left[i,j\right]=\min_{1\le k\le j-1}\left\{T\left[i,k\right]+T\left[k+1,j\right]+p_{i-1}p_kp_j\right\}$$

ראינו כי לכל $i\le k\le j-1$ מתקיים $d>j-(k+1)$, $d>k-i$ ולכן מה"א

$$T\left[i,j\right]=\min_{1\le k\le j-1}\left\{P\left[i,k\right]+P\left[k+1,j\right]+p_{i-1}p_kp_j\right\}=P\left[i,j\right]$$

■

דוגמה עבור $n=3$, $10,50,20,100$

3	$3\cdot10^4$	10^5	0
2	10^4	0	0
1	0	0	0
j/i	1	2	3

נחשב

$$T\left[1,3\right]=\min_{1\le k\le 2}\left\{T\left[1,k\right]+P\left[k+1,3\right]+10\cdot p_k\cdot 100\right\}=\min\left\{15\cdot10^4,3\cdot10^4\right\}=3\cdot10^4$$

חלק ב' של ההרצאה

בעיית התרמיל השלם

קלט W המשקל המירבי של התרמיל ו-$(v_1,w_1),\dots,(v_n,w_n)$ שהם זוגות של מחיר ומשקל בהתאמה. הפריטים אינם ניתנים לחלוקה.

פלט תת קבוצה $S\subseteq[n]$ כך ש-$\sum\limits_{i\in S}w_i\le W$ ו-$\sum\limits_{i\in S}v_I$ מירבי.

פתרון

- נאיבי - מעבר על כל האפשרויות, שהם $\Omega\left(2^n\right)$, שזה יקר מדי.
- חמדני -
  - נמיין את הפריטים בסדר יורד לפי ערכם. בכל שלב נכניס לתרמיל את הפריט היקר ביותר שניתן להכניס.
  - נמיין את הפריטים בסדר יורד על פי ערכם הסגולי $r_i=\frac{v_i}{w_i}$ ובכל שלב נכניס את הפריט עם הערך הסגולי הגדול ביותר שניתן להכניס.

אלו לא יעבדו. עבור $n=3$, $W=50$, $(100,25)\,,(100,25)\,,(150,30)$. לפי שני האלג' הנ"ל, נכניס את הפריט הראשון ונסיים ונרוויח 150. עם זאת, הפתרון האופטימלי הוא להכניס את הפריט השני והשלישי ונקבל 200.

- דינמי - כיצד נחלק לתת-בעיות?

–

תתי בעיות נביט בפריט הראשון ונבחר אם להכניס אותו או לא. נסדר את הפריטים בסדר מסויים $1,\dots,n$. בכל פיצול של עץ הריקורסיה, נפצל לשתי תתי אפשרויות - אחת בה הכנסנו את האיבר הנוכחי ואחת אם לא. לדוגמה, עבור $k=1$ מדובר בבעיות $W-w_1\,,\{2,\dots,n\}$ ו-$W\,,\{2,\dots,n\}$. סה"כ, ניתקל בכל הבעיות $u\,,\{i,\dots,n\}$ כאשר $0\le u\le W$ המשקל המירבי.

נוסחת ריקורסיה נסמן $K\left[i,u\right]$ את הערך של הפתרון האופטימלי לבעיית התרמיל השלם הנ"ל.

$$K\left[i,u\right]=\begin{cases}0 & i=n\wedge w_n>u\\ v_n & i=n\wedge w_n\le u\\ K\left[i+1,u\right] & i<n\wedge w_i>u\\ \max\left\{\underbrace{K\left[i+1,u-w_i\right]+v_i}_{\text{נכניס את הנוכחי}},\underbrace{K\left[i+1,u\right]}_{\text{לא נכניס}}\right\} & i<n\wedge w_i\le u\end{cases}$$

אלגוריתם נבנה טבלה T ונרשום בתא $T\left[i,u\right]$ את הערך $K\left[i,u\right]$. אחרי מילוי הטבלה נחזיר את $K\left[1,W\right]$.

ניתוח בהינתן $1\le i\le n$, מהם המשקלים השונים u בבעיות התרמיל עם קבוצת הפריטים $\{i,\dots,n\}$? נסמן $R\subseteq\{1,\dots,i-1\}$ הקבוצה שמכילה את הפריטים שהכסנו לתרמיל עד עכשיו. אזי $u=W-\sum\limits_{i\in R}w_i$. עתה הסיבוכיות תיקבע ע"י גודל הקבוצה

$$\left\{W-\sum_{i\in R}w_i\right\}_{R\subseteq\{1,\dots,i-1\}}$$

נביט בדוגמה הפרטית $w_i=2^i$ ולכן עבור $R_1\neq R_2$ מתקיים $\sum\limits_{i\in R_1}w_i\neq\sum\limits_{i\in R_2}w_i$ (זוהי סכימה של חזקות של 2 ולכן יחידה על אינדקסים שונים). לכן עבור הקלט הזה מספר הערכים השונים של u יהיה 2^{i-1}. עבור $i=n-1$ נקבל $\Omega\left(2^n\right)$ תתי בעיות וזה יקר (לפחות כמו הנאיבי, ואז לא התקדמנו בכלל).

– עתה נניח הנחה מקלה: המשקל המירבי של התרמיל W וכל המשקלים $w_1,\dots,w_n$ הם מספרים טבעיים. במקרה זה הערכים האפשריים הם מספרים טבעיים u בין 0 ל-W, $W+1$ במספר. לכן נוכל לבנות טבלה T עם n שורות ו-$W+1$ עמודות. סה"כ יהיו ב-T $\mathcal{O}\left(nW\right)$ תאים ונוכל למלא כל אחד ב-$\mathcal{O}\left(1\right)$.

תרגול

בעיית מסילת הרכבת

קלט $L\in\mathbb{N}$ אורך המסילה, $\{1,\dots,K\}$ סוגי חיבורים. $N\in\mathbb{N}$ סוגי חלקים, כל אחד מיוצג ע"י (s_i,e_i,d_i,p_i) שהם (בהתאמה) סוג החיבור בהתחלת החלק, סוג החיבור בסוף הקטע, אורך החלק ומחיר החלק.

פלט המחיר המינימלי עבור מסילה חוקית שהיא מסילה באורך L שבה החלק ה-i מופיע מיד אחרי חלק ה-j אם"ם $e_j=s_i$.

דוגמה $L=3$, החיבורים הם $\{(>,<),(\otimes,\otimes),(\ni,\in),(],[)\}$ והחלקים הם

$$\{(\ni,\otimes,1,30),(>,\in,1,10),(],[,1,30),(\otimes,\otimes,2,40),(],\in,3,100)\}$$

פתרונות חוקיים הם $[---\in$ במחיר 100, $]-[]-[]-[$ במחיר 90 והמינימלי הוא $\ni-\otimes--\otimes$ במחיר 70 (כאשר מספר המקפים מסמל את אורך החלק).

3	100	70	∞	90
2	∞	40	∞	60
1	10	30	∞	30
0	0	0	0	0
	$\in$	$\otimes$	$<$	$[$

פתרון דינמי

1. תתי בעיות: לכל $0\le l\le L$ ולכל $k\in\{1,\dots,K\}$, נמצא את המחיר המינימלי של מסילה באורך l שמסתיימת בחיבור k, $f(l,k)$.

2. נוסחת הריקורסיה: $f(l,k)=\begin{cases}0 & l=0\\ \min\limits_{1\le i\le l:e_i=k}\{p_i+f(l-d_i,s_i)\} & l\neq 0\end{cases}$ כלומר המינימום על הקטע ה-i שנגמר נכון עם הקדמה אליו שמחירם יחד מינימלי.

3. בניית הטבלה: נבנה טבלה בגודל $(L+1)\times K$ כך שבתא ה-(lk) נשמור את הערך ה-$f(l,k)$.

   מילוי הטבלה: נמלא לפי שורות, מהתחתונה לעליונה (כאשר $l=0$ היא התחתונה ו-$l=L$ העליונה) כאשר סדר המילוי בתוך השורות חסר משמעות.

   חילוץ הפתרון: ניקח מינימום על השורה העליונה, בה מתקיים $l=L$.

4. זמן ריצה: גודל הטבלה הוא $\mathcal{O}(L\cdot N\cdot K)$. מילוי כל תא לוקח $O(N)$ (מציאת מינימום על ערכים שכבר חושבו) וחילוץ הפתרון הוא $\Theta(K)$ ולכן סה"כ זמן הריצה הוא $\Theta(L\cdot N\cdot K)$.

All Pairs Shortest Path

קלט גרף מכוון $G=(V,E)$ ופ' משקל $w:E\to\mathbb{R}$.

פלט לכל זו קודקודים $1\leq i,j\leq|V|$ נחזיר את משקל המסלול המינימלי מ-v_i ל-v_j (מטריצה $|V|\times|V|$).

הערה נניח כי אין מעגל בגרף שמשקלו שלילי, כי במקרה זה התשובה היא $-\infty$.

פתרון דינמי ננסה כמה הצעות.

1.

- תתי בעיות: לכל זוג קודקודים, נחשב את המסלול המינימלי מ-v_i ל-v_j.

2.

- תתי בעיות: לכל $1\leq i,j<|V|$, $0\leq m\leq|V|-1$ נחזיר את הערך המינימלי למסלול מ-v_i ל-v_j שאורכו לכל היותר m, $f(i,j,m)$.
- נוסחת הריקורסיה:

$$f(i,j,m)=\begin{cases}0 & i=j\wedge m=0\\ \infty & i\neq j\wedge m=0\\ \min_{(v_k,v_j)\in E}\left\{w\left(\left(v_k,v_j\right)\right)+f\left(v_i,v_k,m-1\right)\right\} & \text{otherwise}\end{cases}$$

כלומר נפצל את המסלול לצלע הראשונה מ-v_i, ועוד מסלול באורך לכל היותר $m-1$ מיעד הצלע הראשונה ל-v_j.

- בניית הטבלה: נבנה טבלה $|V|\times|V|\times|V|$. כל תא נמלא בעזרת נוסחאת הריקורסיה, כאשר התא ה-(i,j,m) יכיל את $f(i,j,m)$.

  מילוי הטבלה: נמלא מ-$m=0$ עד $m=|V|-1$, כאשר את תת המטריצה ה-m-ית נמלא בלי חשיבות לסדר.

  חילוץ פתרון: נחזיר את המטריצה ה-$|V|-1$-ית.

- זמן ריצה: $\mathcal{O}\left(|V|^4\right)$ ($|V|$ תת מטריצות, כל אחת $|V|^2$ איברים שכל אחד דורש $\mathcal{O}(|V|)$ למילוי).

3. Warshall Floyd

- הגדרת תתי בעיות: נגדיר סדר על הקודקודים, $v_1,\ldots,v_R$, $R=|V|$. לכל $1\leq i,j\leq|V|$ ולכל $0\leq k\leq|V|$ נחזיר את משקל המסלול המינימלי מ-v_i ל-v_j שעובר (לכל היותר) דרך הקודקודים $\{v_1,\ldots,v_k\}$.

- נוסחת הריקורסיה:

$$f(i,j,k)=\begin{cases}0 & i=j\wedge k=0\\ \infty & (v_i,v_j)\notin E\wedge k=0\\ w(v_i,v_j) & (v_i,v_j)\in E\wedge k=0\\ \min\left\{\underbrace{f(i,k,k-1)+f(k,j,k-1)}_{k\text{ נמצא במסלול}},f(i,j,k-1)\right\} & \text{otherwise}\end{cases}$$

- בניית הטבלה, מילוי הטבלה, והחזרת הטבלה, זהים לחלוטין להצעה הקודמת.
- זמן הריצה ממלאים סה"כ n^3 תאים, כל אחד דורש $\mathcal{O}(1)$ כי הוא מחשב מינימום על שני איברים שחושבו כבר לפני.

שבוע $\mathbb{VI}$ | אלגוריתמי קירוב ותכנון לינארי

הרצאה

חלק א' של ההרצאה

נמשיך את פתרון בעית התרמיל השלם עם ההנחה המקלה שהמשקלים הם טבעיים.

פתרון דינמי

1. תתי בעיות: כבר ראינו - אם $K[i,u]$ הוא המשקל האופטימלי לבעית התרמיל השלם עם הפריטים $\{i,\dots,n\}$ ומשקל מירבי u, אז ניתקל בכל הבעיות $K[i,u]$.
2. נוסחת ריקורסיה: נשתמש בזו שמצאנו בהרצאה הקודמת:

$$K[i,u]=\begin{cases}0 & i=n\wedge w_n>u\\ v_n & i=n\wedge w_n\le u\\ K[i+1,u] & i<n\wedge w_i>u\\ \max\left\{\underbrace{K[i+1,u-w_i]}_{\text{נכניס את הנוכחי}},\underbrace{K[i+1,u]}_{\text{לא נכניס}}\right\} & i<n\wedge w_i\le u\end{cases}$$

3. בניית הטבלה: נבנה טבלה עם n שורות ו-$W+1$ עמודות ובתא ה-$T[i,u]$ נציב את $K[i,u]$, $\forall i\in[n],u\in\{0,\dots,W\}$.

מילוי הטבלה: נמלא שורה-שורה מ-$i = n$ בכיוון מטה עד ל-$i = 1$. את השורה ה-n נמלא לפי תנאי השפה המתאים בנוסחת הריקוריסה ואת השורות הבאות נמלא ב-$n-1$ איטרציות כאשר באיטרציה $1 \leq j \leq n-1$ נמלא את השורה $i = n - j$ ע"י

$$T\left[i,u\right] = \begin{cases} K\left[i+1,u\right] & w_i > u \\ \max\left\{\underbrace{K\left[i+1,u-w_i\right]}_{\text{נכניס את הנוכחי}}, \underbrace{K\left[i+1,u\right]}_{\text{לא נכניס}}\right\} & w_i \leq u \end{cases}$$

(נוסחת הריקורסיה שכבר ראינו רק בלי מקרי הקצה). נשים לב כי אין חשיבות לסדר המילוי מבחינת עמודות בתוך כל שורה.

חילוץ הפתרון: נחזיר את $T\left[1,W\right]$. כדי לחלץ אילו פריטים יש להכניס לתרמיל כדי להגיע למשקל האופטימלי הנ"ל, נזכור בעת מילוי כל תא בטבלה את ההחלטה הנכונה (להכניס או לא להכניס את הפריט ה-i המשיגה מקסימום בנוסחת הריקורסיה).

4. זמן ריצה: מילוי תא דורש $\mathcal{O}\left(1\right)$ וגודל הטבלה הוא $\mathcal{O}\left(nW\right)$ וסה"כ זמן הריצה הוא $\mathcal{O}\left(nW\right)$.

האם זמן הריצה $\mathcal{O}\left(nW\right)$ יעיל? אם W פולינומי ב-n אז זמן הריצה הוא פולינומי ב-n וזה יעיל מבחינתנו. אם למשל $W = 3^n$, זמן הריצה יהיה גדול יותר מזה של הפתרון הנאיבי.

אלגוריתמי קירוב

חלק גדול מהבעיות האלגוריתמיות המעניינות כנראה לא ניתנות לפתרון יעיל. נחפש פתרונות מקורבים לבעיות כאלה, שחישובם כן יעיל.

חלוקת משימות בין מכונות (load balancing)

קלט k מספר של מכונות זהות ו-n מספרים חיוביים $t_1, \ldots, t_n$ המסמנים את זמני הריצה של n משימות.

פלט חלוקה מאוזנת כמה שאפשר (הממזערת את זמן העבודה של המכונה העמוסה ביותר) של המשימות בין המכונות.

דוגמה $k = 2$, $n = 4$, $t = \left(1, \frac{1}{2}, \frac{1}{2}, 2\right)$. חלוקה אופטימלית היא לתת את המשימה האחרונה (עם זמן 2) למכונה הראשונה ואת השאר למכונה השנייה. באופן נאיבי, נוכל לעבור על כל משימה ולתת אותה למכונה עם זמן ריצת משימות מינימלי באותו הרגע. אלג' זה ייתן לנו במקרה הזה את המשימות עם זמנים $2, 1$ במכונה אחת ו-$\frac{1}{2}, \frac{1}{2}$ במכונה השנייה. זה לא פתרון אופטימלי, כי עיבוד המשימות לוקח 3 (יחידות זמן) ומנת הזמן עם הזמן האופטימלי היא $\frac{q(s)}{q(s^*)} = \frac{3}{2}$. נרצה לקרב פתרון עם מנה קטנה ככל האפשר.

מרחב הפתרונות החוקיים לבעיה זו הוא $S = \left[k\right]^n$ המתאימה כל משימה למכונה. בפרט, $|S| = k^n$.

פתרון נאיבי מעבר על כל אפשרויות הוא לא יעיל ובעצם אקספוננציאלי.

$\forall j \in \left[k\right]$ נגדיר $T_j\left(s\right)$ להיות זמן הריצה של המכונה ה-j לפי הפתרון s, מתקיים $T_j\left(s\right) = \sum_{s(i)=j} t_i$ ונגדיר $q\left(s\right) = \max_{1 \leq j \leq k} T_j\left(s\right)$. נרצה את $s^* \in S$ המקיים $q\left(s^*\right) = \min_{s \in S}\left\{q\left(s\right)\right\}$.

פתרון חמדן נחלק את המשימות בסדר הגעתן כאשר המשימה מגיעה נשלח אותה למכונה הכי פחות עמוסה ברגע זה.

למה נסמן ב-t_{max} את זמן הריצה של המשימה הארוכה ביותר, אזי $q\left(s^*\right) \geq t_{max}$.

הוכחה: ברור. ■

למה $q\left(s^{*}\right)\geq\frac{1}{k}\sum\limits_{i=1}^{n}t_{i}$.

הוכחה:

$$q\left(s^{*}\right)=\max_{1\leq j\leq k}T_{j}\left(s^{*}\right)\overset{(*)}{\geq}\frac{1}{k}\sum_{j=1}^{k}T_{j}\left(s^{*}\right)=\frac{1}{k}\sum_{j=1}^{k}\sum_{s^{*}(i)=j}t_{i}\overset{(**)}{=}\frac{1}{k}\sum_{i=1}^{n}t_{i}$$

$(*)$ פשוט ככה מקסימום עובד, הוא גדול מהממוצע.

$(**)$ הסכום הכפול מונה את כל המשימות בסופו של דבר. ■

משפט האלג' החמדן שתיארנו הינו אלג' $\left(2-\frac{1}{k}\right)$-מקרב לבעיה זו (הגדרה פורמלית בהמשך), כלומר ש-$\frac{q(s)}{q(s^{*})}\leq2-\frac{1}{k}$ כאשר s הפתרון החמדן ו-s^{*} פתרון אופטימלי.

הוכחה: יהי s הפתרון החמדן ו-s^{*} פתרון אופטימלי. יהי $j_{0}\in[k]$ האינדקס של המכונה העמוסה ביותר ב-s. יהי $l\in[n]$ האינדקס של המשימה האחרונה שנשלחה למכונה ה-j_{0}. לפי דרך פעולות של האלג' החמדן, המכונה ה-j_{0} היא המכונה הכי פחות עמוסה אחרי החלוקה של $l-1$ המשימות הראשונות. עבור $j\in[k]$, נסמן $F_{j}\left(s\right)$ את זמן הריצה של המכונה ה-j לאחר חלוקת $l-1$ המשימות הראשונות, מתקיים $F_{j}\left(s\right)=\sum\limits_{i\in[l-1]:s(i)=j}t_{i}$. לכן $F_{j_{0}}\left(s\right)=\min\limits_{1\leq j\leq k}F_{j}\left(s\right)$.

$$\begin{aligned}q\left(s\right) & =T_{j_{0}}\left(s\right)=t_{l}+F_{j_{0}}\left(s\right)=t_{l}+\min_{1\leq j\leq k}F_{j}\left(s\right)\leq t_{l}+\frac{1}{k}\sum_{j=1}^{k}F_{j}\left(s\right)\\
 & =t_{l}+\frac{1}{k}\sum_{j=1}^{k}\sum_{i\in[l-1]:s(i)=j}t_{i}=t_{l}+\frac{1}{k}\sum_{i=1}^{l-1}t_{i}=\left(1-\frac{1}{k}\right)t_{l}+\frac{1}{k}\sum_{i=1}^{l}t_{i}\\
 & \leq\left(1-\frac{1}{k}\right)t_{max}+\frac{1}{k}\sum_{i=1}^{l}t_{i}\leq\left(1-\frac{1}{k}\right)\underbrace{q\left(s^{*}\right)}_{\text{מלמה 1}}+\underbrace{q\left(s^{*}\right)}_{\text{מלמה 2}}\\
 & =\left(2-\frac{1}{k}\right)q\left(s^{*}\right)
\end{aligned}$$

■

הערה אם נמיין את המשימות בסדר יורד לפי זמני ריצה $t_{1}\geq\ldots\geq t_{n}$ ונשתמש באותו האלג', נגיע ל-$\frac{3}{2}$ קירוב.

הערה ניתן להשיג קירוב $1+\epsilon$ בזמן ריצה $\mathcal{O}\left(n^{\left(\frac{1}{\epsilon}\right)^{\frac{3}{2}}}\right)$, שזה מאוד גדול - לדוגמה עבור $\epsilon=\frac{1}{100}$, מדובר ב-$\mathcal{O}\left(n^{1000}\right)$.

בבעית אופטימיזציה נתון מרחב פתרונות חוקיים S לבעיה אלג' נתונה ופ' איכות $q:S\rightarrow\mathbb{R}_{+}$. בבעית מקסימיזציה נחפש $s^{*}\in S$ כך ש-$q\left(s^{*}\right)=\max\limits_{s\in S}q\left(s\right)$ ובבעית מינימיזציה נחפש $s^{*}\in S$ כך ש-$q\left(s^{*}\right)=\min\limits_{s\in S}q\left(s\right)$.

הגדרה נאמר כי אלג' הינו c-מקרב לבעיית מקסימיזציה נתונה עם פתרון אופטימלי s^* אם האלג' מחזיר פתרון s כך ש-$c \le \frac{q(s^*)}{q(s)}$. נאמר כי אלג' הינו c-מקרב לבעית מינימיזציה עם פתרון אופטימלי s^* אם האלג' מחזיר פתרון s כך ש-$c \le \frac{q(s)}{q(s^*)}$.

חלק ב' של ההרצאה

בעיית כיסוי ע"י קבוצות (Set Cover)

קלט $n \in \mathbb{N}$, $A_1, \ldots, A_r \subseteq [n]$ כך ש-$\bigcup_{i=1}^{r} A_i = [n]$.

פלט תת קבוצה $S \subseteq [r]$ כך ש-$\bigcup_{i \in S} A_i = [n]$ וגם $|S|$ מינימלי.

זאת בעיה NP קשה ולכן נחפש לה אלג' מקרב.

דוגמה $n = 10$, $r = 6$,

$$A_1 = \{1, \ldots, 5\}$$
$$A_2 = \{6, \ldots, 10\}$$
$$A_3 = \{1, 2, 3\}$$
$$A_4 = \{6, 7, 8\}$$
$$A_5 = \{1, \ldots, 8\}$$
$$A_6 = \{3, 4, 5\}$$

במקרה זה, $s^* = \{1, 2\}$.

ננסח אלג' מקרב שיעבוד לפי העיקרון החמדני לפיו בכל שלב נוסיף קבוצה המכסה הכי הרבה ממה שנשאר.

פסאודו-קוד

1. אתחול: $G = \varnothing$ (הפתרון החמדן), $X = [n]$ (מה שנותר לכסות).
2. איטרציה: נבחר $1 \le i^* \le r$ כך ש-$|A_i \cap X| = \max_{1 \le i \le r} |A_i \cap X|$ ונעדכן $G = G \cup \{i^*\}$, $X = X \backslash A_{i^*}$.

נסמן ב-s^* פתרון אופטימלי, $k = |s^*|$ וגם G הפתרון החמדן ו-$t = |G|$. נסמן X_j את הקבוצה שנותר לכסות אחרי j האיטרציות הראשונות של האלג', לכל $1 \le j \le t$ ונשים לב כי מתקיים $[n] = X_0 \supseteq \ldots \supseteq X_t = \varnothing$. נשים לב כי t הוא מספר האיטרציות של האלג' עד עצירה.

למה $\forall x \in (0, 1]$ ו-$\forall y \ge 0$ מתקיים $(1 - x)^y < e^{-xy}$.

למה לכל $0 \le j \le t - 1$ מתקיים $\max_{i \in s^*} |A_i \cap X_j| \ge \frac{1}{k} |X_j|$, כלומר שקיימת קבוצה שנבחרה בפתרון האופטימלי שמכסה לפחות $\frac{1}{k}$ מהאיברים הנותרים אחרי האיטרציה ה-j (עקרון שובך היונים בערך).

הוכחה: נראה כי $\bigcup_{i\in s^*}(A_i\cap X_j)=X_j$. מכיוון ש-$s^*$ הוא פתרון חוקי, מתקיים $\bigcup_{i\in s^*}A_i=[n]$ ולכן

$$\bigcup_{i\in s^*}(A_i\cap X_j)=\left(\bigcup_{i\in s^*}A_i\right)\cap X_j=[n]\cap X_j=X_j$$

ולכן

$$\max_{i\in s^*}|A_i\cap X_j|\geq\frac{1}{k}\sum_{i\in s^*}|A_i\cap X_j|\geq\frac{1}{k}\left|\bigcup_{i\in s^*}(A_i\cap X_j)\right|=\frac{1}{k}|X_j|$$

■

משפט האלג' החמדני משיג $\lceil\ln n\rceil$-קירוב לבעיה.

הוכחה: צריך להראות כי

$$t=|G|\leq|s^*|\lceil\ln n\rceil=k\lceil\ln n\rceil\overset{\text{סימון}}{=}u$$

נוכיח כי $t\leq u$. יהי $0\leq j\leq t-1$, נוכיח כי $|X_{j+1}|\leq\left(1-\frac{1}{k}\right)|X_j|$. יהי i^* האינדקס שנבחר באיטרציה $j+1$. לפי עקרון הפעולה של האלג' החמדן, מתקיים

$$|X_j\cap A_{i^*}|=\max_{1\leq i\leq r}|X_j\cap A_i|\geq\max_{i\in s^*}|X_j\cap A_i|\overset{\text{למה }2}{\geq}\frac{1}{k}|X_j|$$

לכן

$$|X_{j+1}|=|X_j\backslash A_{i^*}|=|X_j|-|X_j\cap A_{i^*}|\leq|X_j|-\frac{1}{k}|X_j|=\left(1-\frac{1}{k}\right)|X_j|$$

נניח בשלילה כי $t>u$ ולכן $X_u\neq 0$ ולכן

$$1\leq|X_u|\overset{\text{למה }2}{\leq}\left(1-\frac{1}{k}\right)|X_{u-1}|\overset{\text{למה }2}{\leq}\left(1-\frac{1}{k}\right)^2|X_{u-2}|\leq\ldots\overset{\text{למה }2}{\leq}\left(1-\frac{1}{k}\right)^u|X_0|\overset{\text{למה }1}{<}e^{-\frac{u}{k}}|X_0|=e^{-\lceil\ln n\rceil}n\leq\frac{1}{n}\cdot n=1$$

סתירה. ■

תרגול

הגדרה יהיו $a\in\mathbb{R}^n,b\in\mathbb{R}$ על-מישור על a,b מוגדר ע"י $H_{a,b}=\left\{x\in\mathbb{R}^n:a^Tx=b\right\}$.

חצי מרחב הוא $\widetilde{H}_{a,b}=\left\{x\in\mathbb{R}^n : a^Tx\le b\right\}$.

פוליהידרון מוגדר ע"י $P=\{x\in\mathbb{R}^n : Ax\le b\}$, עבור $A\in M_n(\mathbb{R}), b\in\mathbb{R}^n$.

הערה על מישור הוא דומה למשוואת מישור סתומה. חצי מרחב הוא בעצם איחוד של הרבה על מישורים ומוגדר ע"י n אי שוויונים לינאריים. $P=\bigcap_{i=1}^{n}\widetilde{H}_{A_i,b}$ כאשר A_i היא השורה ה-i ב-A. לאינטואיציה נוספת ראו אתר הקורס.

דוגמה $a=\begin{pmatrix}1\\-1\end{pmatrix}$, $b=0$. $\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=x-y=0$.

הגדרה בעיית אופטימיזציה תקרא בעית תכנון לינארי אם ניתן לתאר אותה בצורה הבאה: בהינתן $A\in M_{m\times n}(\mathbb{R}), b\in\mathbb{R}^m, c\in\mathbb{R}^n$, מציאת $x\in\mathbb{R}^n$ כך שמתקבל הערך המקסימלי $\max c^Tx$ כאשר $Ax\le b$ ו-$x\ge 0$.

דוגמה נרצה את הערך המקסימלי $\max x_1+2x_2$ כך ש-$4x_1-x_2\le 3$ ו-$2x_1+2x_3\ge -100$ וכמובן $x_1,x_2\ge 0$. זוהי בעיית תכנון לינארי כאשר $c=\begin{pmatrix}1\\2\end{pmatrix}$, $A=\begin{pmatrix}4 & -1\\-2 & -2\end{pmatrix}$ ו-$\begin{pmatrix}3\\100\end{pmatrix}$.

דוגמה נרצה את הערך המקסימלי $\max x_1+8x_2$ כך ש-

$$x_1\ge 3 \qquad (1)$$

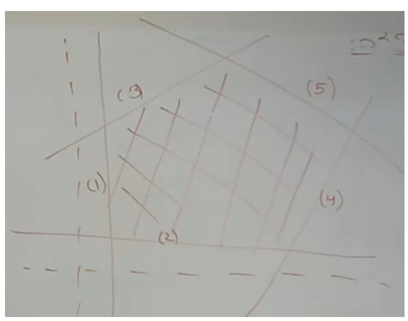
$$x_2\ge 2 \qquad (2)$$

$$-3x_1+4x_2\le 14 \qquad (3)$$

$$4x_1-4x_2\le 25 \qquad (4)$$

$$x_1+x_2\le 15 \qquad (5)$$

נרצה את $\begin{pmatrix}x_1\\x_2\end{pmatrix}$ בתחום בתמונה שעבורם x_1+8x_2 מקסימלי.



הערה מסתבר שהפתרון האופטימלי תמיד נמצא על אחד הקודקודים בהם נפגשים שני על מישורים (במקרה הדו ממדי, העל מישורים הם גרפים של פ' לינאריות). אלג' מאוד פופולרי Simplex שמתבסס על העובדה הזו רץ בזמן פולינומיאלי ב-$m+n$.

בעיית התרמיל השברי

היזכרו בבעיית התרמיל השברי (והשלם) משבוע $\mathbb{I}$.

נוכל להמיר את הבעיה הזו גם כן לבעית תכנון לינארי. נגדיר $c=\begin{pmatrix}v_1\\ \vdots\\ v_n\end{pmatrix}$, $\underbrace{\begin{pmatrix}1\\ \vdots\\ 1\\ W\end{pmatrix}}_{b} \le \begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\underbrace{\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ & & \ddots & \\ 0 & \cdots & 0 & 1\\ w_1 & w_2 & \cdots & w_n\end{pmatrix}}_{A}$.

הגדרה בעיית אופטימיזציה תקרא בעית תכנון לינארי בשלמים (ILP) אם ניתן לתאר אותה בצורה הבאה: בהינתן

$$c\in\mathbb{R}^n, b\in\mathbb{R}^m, A\in M_{m\times n}(\mathbb{R})$$

מציאת $x\in\mathbb{Z}^n$ כך שמתקבל הערך המקסימלי $\max c^Tx$ כאשר $Ax\le b$ ו-$x\ge 0$.

את בעית התרמיל השלם נוכל גם כן לייצג כבעית תכנון לינארי בשלמים, כאשר נדרוש ש-$x_i\le 1$ בנוסף לדרישה שהוא אי-שלילי.

בעית הסרת המשולשים

קלט גרף לא מכוון $G=(V,E)$.

פלט קבוצה $S\subseteq E$ בגודל מינימלי כך שבגרף $G=(V,E\backslash S)$ אין משולשים. כלומר לכל $u,v,w\in V$ מתקיים

$$\{\{u,v\},\{v,w\},\{w,u\}\}\nsubseteq E\backslash S$$

בעיה זו היא NP קשה (נלמד בחישוביות) ולכן אין לה פתרון יעיל. ננסח אלג' קירוב המבוסס על תכנון לינארי בשלמים (ייתכן שהוא מסיר יותר מדי צלעות, אבל הוא בטוח חוקי). נשים לב כי נצטרך להוריד לפחות צלע אחת מכל משולש בגרף.

נגדיר משתנים מציינים $X_j=\begin{cases}1 & e_j\in S\\ 0 & e_j\notin S\end{cases}$, $\forall j\in[|E|]$. נרצה ש-$\forall u,v,w\in V$ כך ש-$\{u,v\},\{v,w\},\{w,u\}\in E$ יתקיים $X_{u,v}+X_{v,w}+X_{w,u}\ge 1$ (כלומר אם יש משולש, לפחות צלע אחת ממנו נמצאת בקבוצת הצלעות שיוסרו).

ננסח את בעית התכנון הלינארי בשלמים. נרצה את $\min|S|$ כאשר $|S|=\sum_{i=1}^{|E|}X_i$, כלומר נרצה בעצם את $\min\sum_{i=1}^{|E|}X_i$. נוכל לחשוב על האופטימיזציה בבעיה כמציאת S כך ש-$|E|-\sum_{i=1}^{n}X_i$ מקסימלי, כלומר $\max|E|-\sum_{i=1}^{n}X_i$, אבל בגלל ש-$|E|$ קבוע אז מציאת מקסימום על הערך הנ"ל שקול על מציאת $\max\left(-\sum_{i=1}^{n}X_i\right)$, שזה שקול ל-$\min\sum_{i=1}^{n}X_i$.

נדרוש כי $0\le X_j$ וגם $X_j\le 1$ ו-$X\in\mathbb{Z}^n$. נדרוש בנוסף כי $\forall u,v,w\in V$ כך ש-$\{u,v\},\{v,w\},\{w,u\}\in E$,

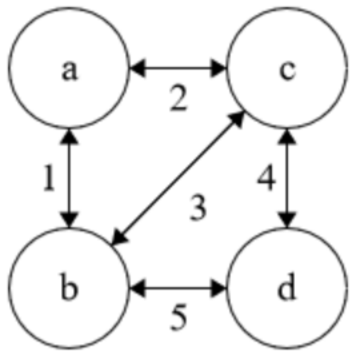
$$X_{u,v}+X_{v,w}+X_{w,u}\ge 1$$

או במקרה שלנו $-X_{u,v} - X_{v,w} - X_{w,u} \le -1$. לכן נגדיר

$$\underbrace{\begin{pmatrix} \overbrace{\begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \ddots & \\ 0 & \cdots & 0 & 1 \end{matrix}}^{|E|} \end{pmatrix}}_{M} \begin{pmatrix} x_1 \\ \vdots \\ x_{|E|} \end{pmatrix} \le \begin{pmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{pmatrix}$$

כאשר שורה ב-M היא מהצורה $\begin{pmatrix} 0 \cdots \underbrace{-1}_{X_{u,v}} \cdots \underbrace{-1}_{X_{v,w}} \cdots \underbrace{-1}_{X_{w,u}} \cdots 0 \end{pmatrix}$ כאשר $\{u,v\}, \{v,w\}, \{w,u\} \in E$, $u,v,w \in V$.

דוגמה נביט בגרף הבא



כאן, בהינתן מספור הצלעות, $M = \begin{pmatrix} -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \end{pmatrix}$.

שבוע $\mathbb{VIII}$ | אלגוריתמי קירוב נוספים

הרצאה

חלק א' של ההרצאה

סיימנו את ההוכחה מההרצאה הקודמת.

בעיית כיסוי ע"י קודקודים (Vertex Cover)

סיפור מסגרת נרצה לשלוח פייק ניוז, כל צלע היא יחס חברות בפייסבוק, כל קודקוד משתמש, נרצה לדעת מה הקבוצה המינימלי כך שכל אחד יוכל להפיץ לחברים שלו (לכל זוג יהיה מישהו שהפיצו לו).

קלט גרף לא מכוון $G = (V, E)$.

פלט קבוצה $S \subseteq V$ כך שלכל $e = (x,y) \in E$ מתקיים $x \in S$ או $y \in S$ וכך ש-$|S|$ מינימלי.

דוגמה בגרף כוכב (קודקוד במרכז וצלעות מחוברות רק אליו מסביב), $|s^*|=1$ הוא בחירת הקודקוד המרכזי.

דוגמה $G=K_n$, $|s^*|=n-1$ כי אם יש שני קודקודים שלא נבחרו אז הצלע של שניהם לא תקיים את התנאי הרצוי.

דוגמה $G=K_{a,b}$ הגרף המלא הדו צדדי, $a\le b$. כאן $|s^*|=a$ כי נוכל לבחור את כל הצד השמאלי וכל הצלעות בהכרח יכילו קודקוד אחד ממנו.

הערה זו בעיה NP-קשה ולכן נחפש אלג' קירוב.

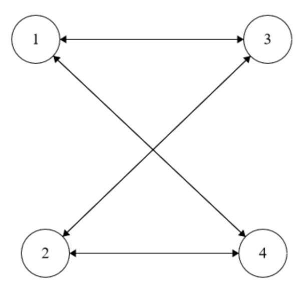
הערה נשים לב כי נוכל לעשות רידוקציה לבעיית כיסוי ע"י קבוצות: לכל קודקוד $x\in V$, נתאים את קבוצת הצלעות בגרף G העוברות דרך x, כלומר נקבל $|V|$ תת קבוצות של E ונרצה לכסות את E עם כמה שפחות קבוצות כאלה (כלומר כמה שפחות קודקודים). לכן נוכל להשתמש באלג' שלנו לכיסוי ע"י קבוצות כדי לקבל אלג' $\lceil\ln|E|\rceil$-מקרב לבעיית כיסוי ע"י קודקודים.

עם זאת, בגלל שגרף הוא מבנה מיוחד, נוכל לנצל זאת לטובתנו למציאת אלג' קירוב יותר טוב.

פסאודו-קוד

1. אתחול: $G=\varnothing$, $X=E$.
2. איטרציה: נבחר צלע $X\ni e=(x,y)$ (שרירותית!) ונעדכן $G=G\cup\{x\}\cup\{y\}$ ונמחק מ-X את כל הצלעות העוברות דרך x או דרך y.
3. סיום: כאשר $X=\varnothing$ נעצור ונחזיר את G.

דוגמה עבור הגרף הבא, $s^*=\{1,2\}$ אבל $s=\{1,2,3,4\}$ עבור סדר בחירת הצלעות $\{1,3\}$ ואז $\{2,4\}$.



הערה זהו האלג' הנאיבי ביותר אבל גם מקרב הכי טוב שידוע לנו. בנוסף, הוכח כי לא ניתן לקרב יותר טוב מ-$\sqrt{2}$ ורבים מאמינים כי אי אפשר לקרב יותר טוב מ-2.

משפט האלג' מחזיר 2-קירוב לבעיה.

הגדרה (מדיסקרטית) <u>זיווג</u> בגרף הוא אוסף צלעות ללא קודקודים משותפים.

למה אם הגרף G מכיל זיווג בגודל t אזי כל כיסוי ע"י קודקודים בגרף הזה מכיל לפחות t קודקודים (אם יש t צלעות זרות נצטרך לפחות t קודקודים כדי לכסות אותן, בלי להתסכל אפילו על שאר הצלעות).

הוכחה: יהי s כיסוי ע"י קודקודים של הגרף G. s חייב להכיל קודקוד אחד לפחות על כל אחת מצלעות הזיווג. כל הקודקודים האלה שונים זה מזה כי לצלעות בזיווג אין קודקודים משותפים. לכן, מכיוון שיש t צלעות בזיווג, s מכיל לפחות t קודקודים. ■

הוכחה: (של המשפט) נסמן ב-s^* פתרון אופטימלי וב-t את מספר האיטרציות של האלג' עד העצירה. יהי s הפתרון המוחזר ע"י האלג'. לכן $|s| = 2t$ כי בכל אירטציה האלג' מוסיף 2 קודקודים ל-s. תהיינה $l_1, \ldots, l_t$ הצלעות שהאלג' בוחר לאורך האיטרציות שלו לפי אופן פעולתו. לצלעות $l_1, \ldots, l_t$ אין קודקודים משותפים ולכן הן מהיות זיווג בגרף בגודל t ולכן הלמה $|s^*| \geq t$ ולכן $|s| = 2t \leq 2|s^*|$. ■

בעיית כיסוי קודקודים ממושקל

קלט גרף לא מכוון $G = (V, E)$ ופ' משקל $w : V \to \mathbb{R}_+$.

פלט כיסוי קודקודים $S \subseteq V$ כך ש-$w(S)$ מינימלי.

סכמה לניסוח אלגוריתם מקרב בעזרת תכנון לינארי

1. נתרגם את הבעיה לבעיית מינימיזציה של פ' לינארית במשתנים שמקבלים ערכים שלמים ומקיימים אילוצים לינאריים (ניסוח אנליטי של הבעיה המקורית) - זו בעיית ILP.
2. נסיר את האילוץ של הערכים השלמים ונחליף אותו באי שוויונים לינאריים (צריכה להיות רלקסציה - הבעיה החדשה צריכה להיות קלה יותר).
3. נפתור את בעיית ה-LP שקיבלנו בשלב הקודם בעזרת אלג' הפותר בעיות תכנון לינארי (Simplex לדוגמה) ונקבל פתרון אופטימלי לבעיה זו.
4. נעגל את הפתרון האופטימלי לבעיית ה-LP לפתרון טוב לבעיה המקורית ונחזיר את הפתרון המקורב.

חלק ב' של ההרצאה

כלומר בעצם, מה שאנחנו עושים זה: בעיה $\Leftarrow$ ILP $\Leftarrow$ LP $\Leftarrow$ פתרון מקרב.

אלגוריתם

1. בהינתן כיסוי קודקודים S, $\forall v \in V$ נתאים משתנה $X(v) = \begin{cases} 1 & v \in S \\ 0 & v \notin S \end{cases}$. לכן מהיות S כיסוי קודקודים, בהינתן צלע $(a, b) \in E$, מתקיים $a \in S$ או $b \in S$, כלומר $X(a) = 1$ או $X(b) = 1$, או לחלופין $X(a) + X(b) \geq 1$. מתקיים גם כי משקל הכיסוי S הוא "פונקציונאל לינארי" על המשתנים שהגדרנו, כלומר $w(s) = \sum_{v \in S} w(v) = \sum_{v \in V} w(v)$.

קיבלנו בעיית ILP:

$$\begin{cases}\min \sum_{v\in V} w(v)X(v)\\ X(v)\in\{0,1\},\forall v\in V\\ X(a)+X(b)\geq 1,\forall(a,b)\in E\end{cases}$$

2. נחליף $\forall v\in V$ את האילוץ $X(v)\in\{0,1\}$ לאילוץ לינארי $0\leq X(v)\leq 1$ (יכול עכשיו להיות גם ממשי). קיבלנו בעיית LP:

$$\begin{cases}\min \sum_{v\in V} w(v)X(v)\\ 0\leq X(v)\leq 1,\forall v\in V\\ X(a)+X(b)\geq 1,\forall(a,b)\in E\end{cases}$$

3. נפתור את בעיית ה-LP שקיבלנו בשלב הקודם ונקבל פתרון אופטימלי.

נסמן X^* פתרון אופטימלי של הבעיה המקורית ונסמן ב-X^*_{LP} את הפתרון האופטימלי שקיבלנו לבעיית ה-LP בשלב הנוכחי.

4. $\forall v\in V$, נעגל $X(v)=\begin{cases}1 & X^*_{LP}(v)\geq\frac{1}{2}\\ & \text{otherwise}\end{cases}$ ונחזיר את X.

משפט הפתרון X שהאלג' הנ"ל מחזיר הוא פתרון חוקי ו-2-מקרב לבעיה המקורית.

הוכחה: חוקיות: לפי הגדרת X, $\forall v\in V$ מתקיים $X(v)\in\{0,1\}$. תהי $(a,b)\in E$. מהיות X^*_{LP} פתרון חוקי של בעיית ה-LP אזי $X^*_{LP}(a)+X^*_{LP}(b)\geq 1$ ולכן $X^*_{LP}(a)\geq\frac{1}{2}$ או $X^*_{LP}(b)\geq\frac{1}{2}$ ולכן $X(a)=1$ או $X(b)=1$.

קירוב:

למה

$$w(X^*_{LP})\leq\sum_{v\in V} w(v)X^*_{LP}(v)\leq\sum_{v\in V} w(v)X^*(v)=w(X^*)$$

הוכחה: X^* הוא פתרון חוקי לבעית ה-ILP. כל פתרון חוקי של בעיית ה-ILP הוא גם פתרון חוקי של בעית ה-LP. ואילו X^*_{LP} הוא פתרון אופטימלי לבעיית ה-LP, ולכן משקלו קטן מכל פתרון חוקי אחר של בעיית ה-LP. ■

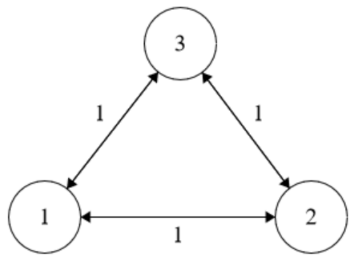
למה $\forall v\in V$ מתקיים $X(v)\leq 2X^*_{LP}(v)$ (מהגדרת העיגול, הסטודנטית המשקיעה תחשב ותגלה שזה אכן המצב).

נסיים את הוכחת הקירוב:

$$\sum_{v\in V} w(v)\,X(v) \overset{\text{למה 2}}{\leq} \sum_{v\in V} w(v)\,2X^*_{LP}(v) = 2\sum_{v\in V} w(v)\,X^*_{LP}(v) \overset{\text{למה 1}}{\leq} 2\sum_{v\in V} w(v)\,X^*(v)$$

■

דוגמה הגרף המלא על שלושה קודקודים עם משקלים אחד, כבציור



הפתרון האופטימלי הוא בחירת שתי צלעות, כלומר $\sum_{v\in V} w(v)\,X^*(w) = 2$. כדי לפתור את הבעיה בתכנון לינארי, נרצה לתת ערכים לכל קודקוד כך שסכום הערכים הוא מינימלי וגם סכום הערכים של כל זוג קודקודים המחוברים בצלע יהיה גדול מאחד. קל לראות כי הפתרון האופטימלי כאן הוא מתן הערך $\frac{1}{2}$ לכל קודקוד, וכך נקבל $\sum_{v\in V} w(v)\,X^*_{LP}(w) = \frac{2}{3}$. הפתרון שיחזיר האלג' שלנו יכלול את כל שולשת הקודקודים והיחס שנקבל הוא $\frac{2}{3}$, אכן קירב ביחס של לכל היותר 2.

תרגול

בעיית 3SAT

קלט נוסחת 3-CNF עם m פסוקיות, $C_1,\ldots,C_m$ (לדוגמה $(x_1 \vee x_2 \vee \neg x_3)$ ב-n משתנים. נניח כי $n \leq 3m$ וגם שאין בפסוקית משתנה שחוזר על עצמו (לדוגמה $x_1 \vee x_2 \vee \neg x_1$).

פלט האם קיימת השמה של $x_1,\ldots,x_n$ בה הנוסחה מחזירה "אמת".

הערה נשים לב שזו לא בעיית אופטימיזציה, אלא בעיית הכרעה. בנוסף, זו בעיה NP קשה ולכן נפתור בקירוב בעיית אופטימיזציה המתאימה לבעיה זו.

בעיית MAX-3SAT

קלט כמו ב-3SAT.

פלט השמה המספקת מספר מקסימלי של פסוקיות אמת.

פסאודו-קוד

1. נבדוק כמה פסוקיות ההשמה $X_T = (T, \ldots, T)$ מעניקה ונסמן מספר זה ב-f.
2. נבדוק כמה פסקויות $X_F = (F, \ldots, F)$ מעניקה ונסמן מספר זה ב-f.
3. אם $t < f$ נחזיר X_T, אחרת נחזיר X_F.

משפט האלג' הנ"ל הוא 2-מקרב לבעיה.

הוכחה: לכל פסוקית, או ש-X_T מספק אותה או ש-X_F מספק אותה או שניהם. לכן $t + f \geq m$ ולכן $\max\{t, f\} \geq \frac{1}{2}m$ והפתרון האופטימלי, s^*, מקיים $s^* \leq m$ ולכן ולכן $s^* \leq 2\max\{t, f\}$. ■

הערה האלג' היה עובד בדיוק באותו האופן גם עבור כל השמה אחרת שלה, ושלילתה.

הערה האלג' הוא 2-מקרב לכל נוסחאת CNF, לא רק כשיש 3 משתנים בפסוקית.

הגדרה אלג' <u>דטרמיניסטי</u> מחזיר פתרון אופטימלי.

אלג' <u>c-מקרב</u> מחזיר פתרון c מקרב.

אלג' <u>הסתברותי</u> מחזיר פתרון אופטימלי בהסתברות גבוהה כרצוננו, אחרת מחזיר fail.

אלג' <u>c-מקרב הסתברותי</u> מחזיר פתרון c-מקרב בהסת' גבוהה כרצוננו, אחרת מחזיר fail.

הערה לרוב "גדולה כרצוננו" יהיה בהסת' $1 - \frac{1}{a^k}$ כשנוכל להגדיל את k כמה שנרצה (הוא מספר טבעי).

פסאודו-קוד לאלג' הבסיסי:

1. לכל משתנה x_i, נגדיר אותו באופן אקראי ואחיד בין T ל-F.
2. נבדוק אם ההשמה שהתקבלה מספקת לפחות $\frac{7}{8}m$ פסוקיות נחזיר אותה, אחרת נחזיר fail.

לאלג' המלא:

נחזור על האלג' הבסיסי $k(m+1)$ פעמים באופן בלתי תלוי, אם באחת הריצות הייתה הצלחה, נחזיר את ההשמה שהתקבלה, אחרת נחזיר fail.

זמן ריצה בכל איטרציה נבצע את האלג' הבסיסי שדורש $\mathcal{O}(m+n) = \mathcal{O}(m)$ (הגרלה וחישוב הפסוקיות) וסה"כ נבצע אותו $k(m+1)$ פעמים, כלומר סה"כ $\mathcal{O}(km^2)$.

משפט האלג' המלא הוא $\frac{8}{7}$-מקרב.

הוכחה: <u>חוקיות</u>: האלג' מחזיר או השמה חוקית ל-CNF או fail, פלטים חוקיים לאלג' התסברותי.

<u>קירוב</u>: אם האלג' הצליח, ההשמה המוחזרת מספקת לפחות $\frac{7}{8}m$ פסוקיות ולכן נקבל $\frac{7}{8}m \geq \frac{7}{8}s^*$ עבור s^* אופטימלי.

<u>הסתברות</u>:

טענה האלג' הבסיסי מצליח בהסת' $\frac{1}{m+1}$.

הוכחה: נגדיר $\Omega=\{T,F\}^n$ מרחב הסתברות אחיד. נגדיר Y מ"מ שערכו מס' הפסוקיות שלא סופקו על השמה כלשהי. נגדיר לכל פסוקית C_i מ"מ $Y_i=\begin{cases}1 & \text{לא סופקה ע"י ההשמה } \omega \ C_i\\ 0 & \text{אחרת}\end{cases}$. מתקיים $Y=\sum\limits_{i=1}^{m}Y_i$.

$$E[Y_i]\overset{\text{אינדיקטור}}{=}P(Y_i=1)\overset{(*)}{=}\left(\frac{1}{2}\right)^3=\frac{1}{8}$$

$(*)$ זו בדיוק ההסת' ששלושה משתנים אקראיים אחידים יתאימו לנוסחה.

$$E[Y]=E\left[\sum_{i=1}^{m}Y_i\right]=\sum_{i=1}^{m}E[Y_i]=\sum_{i=1}^{m}\frac{1}{8}=\frac{1}{8}m$$

$$\begin{aligned}
P(\text{האלג' הבסיסי נכשל}) &= P\left(\text{ההשמה סיפקה לכל היותר } \frac{7}{8}m \text{ פסוקיות}\right)\\
&= P\left(\text{ההשמה לא סיפקה לפחות } \frac{1}{8}m \text{ פסוקיות}\right)\\
&= P\left(Y>\frac{1}{8}m\right)\\
&= P\left(Y\geq\frac{1}{8}m+\frac{1}{8}\right)\\
&= P\left(Y\geq\underbrace{\frac{1}{8}m}_{E[Y]}\underbrace{\left(1+\frac{1}{m}\right)}_{c}\right)\\
&\overset{\text{א"ש מרקוב}}{\leq}\frac{1}{1+\frac{1}{m}}=\frac{m}{m+1}
\end{aligned}$$

ולכן

$$P(\text{האלג' הבסיסי הצליח})=1-P(\text{האלג' הבסיסי נכשל})\geq\frac{1}{m+1}$$

■

נראה כי האלג' מצליח בהסת' $1-\frac{1}{e^{k}}$.

$$\begin{aligned}P\left(\text{האלג' הכללי נכשל}\right) & =P\left(\text{האלג' הבסיסי נכשל }k\left(m+1\right)\text{ פעמים}\right)\\ & \overset{\text{ב"ת}}{=}\left(P\left(\text{האלג' הבסיסי נכשל}\right)\right)^{k(m+1)}\\ & =\left(1-P\left(\text{האלג' הבסיסי הצליח}\right)\right)^{k(m+1)}\\ & \le\left(1-\frac{1}{m+1}\right)^{k(m+1)}\\ & \le\frac{1}{e^{k}}\end{aligned}$$

■

הערה נוכל להכליל ולטעון כי עבור l משתנים בפסוקית, נקבל על אותו האלג' קירוב של $\frac{2^{l}-1}{2^{l}}$.

שבוע $\mathbb{VIII}$ | עוד אלגוריתמי קירוב

הרצאה

חלק א' של ההרצאה

בעיית פתרון אופטימלי של מערכת משוואות לינאריות מודולו 2 (MAX-LIN-2)

סיפור מסגרת (העשרה) צפנים לתיקון שגיאות. נרצה להעביר מ-A ל-B מידע (ביטים) דרך ערוץ רועש שיכול לשנות ביטים לפני הגעת ההודעה. לשם כך בעת שליחת m, נעביר אותו דרך A מטריצה מוסכמת מראש שתוסיף מידע לתיקון שגיאות ונשלח את $M=Am$. ההודעה שתגיע ל-B היא $\widetilde{M}$ אבל גם B יודע מה A ולכן בהנחה שאין יותר מדי שגיאות, הוא יוכל באמצעות המידע שנוסף ע"י A ל-m להסיק אילו ביטים נפגמו ולתקנם.

קלט $A\in M_{k\times n}\left(\mathbb{F}_{2}\right)$ ו-$b\in\mathbb{F}_{2}^{k}$ המייצגים מערכת משוואות לינאריות $Ax=b$ כאשר $X=\begin{pmatrix}x_{1}\\ \vdots\\ x_{n}\end{pmatrix}$ הוא וקטור משתנים.

פלט השמה $s=\begin{pmatrix}s_{1}\\ \vdots\\ s_{n}\end{pmatrix}$ של ערכים למשתנים המספקת כמה שיותר משוואות.

דוגמה $k=n=3$. $\begin{pmatrix}1&0&0\\0&1&1\\1&1&1\end{pmatrix}\begin{pmatrix}x_{1}\\x_{2}\\x_{3}\end{pmatrix}=\begin{pmatrix}1\\1\\1\end{pmatrix}$. למערכת אין פתרון (אין הצעה המספקת כל כל המשוואות). פתרון אופטימלי כאן הוא $\begin{pmatrix}1\\1\\0\end{pmatrix}$.

הערה זו בעיה NP-קשה, נחפש אלג' קירוב.

הערה זו בעיה דומה ל-MAX-3-SAT. שם, האלג' היה בחירת השמה מקרית המספקת פסוקית אחת בסיכוי $\frac{7}{8}$ וכך קיבלנו אלג' הסת' $\frac{8}{7}$-מקרב המחזיר $\frac{8}{7}$-קירוב בסיכוי גבוה כרצוננו.

נציג אלג' קירוב דטרמיניסטי שהוא 2-מקרב. האלג' משתמש בשיקולים הסת'. ניתן להוכיח כי זה הקירוב הכי טוב שניתן להשיג.

נסמן ב-$\mathcal{S}$ את מרחב הפתרונות החוקיים לבעיה - $\mathcal{S}=\mathbb{F}_2^n$, $|\mathcal{S}|=2^n$ ולכן הפתרון הנאיבי הוא לא יעיל. נסמן $r_1,\ldots,r_k\in\mathbb{F}_2^n$ את שורות המטריצה A ו-$c_1,\ldots,c_n\in\mathbb{F}_2^k$ את עמודות המטריצה.

נגדיר פ' $Y:\mathcal{S}\to\mathbb{R}^+$ המוגדר ע"י $Y(s)$ מספר המשוואות ש-s מספקת, פורמלית,

$$Y(s)=|\{i\in[k]:\langle r_i\mid s\rangle=b_i\}|=\{i\in[k]:(As)_i=b_i\}=\left|\left\{i\in[k]:\left(\sum_{i=1}^{n}s_jc_j\right)_i=b_i\right\}\right|$$

נרצה למצוא מקסימום של Y על $\mathcal{S}$.

ננסה לבנות אלג' דטרמיניסטי איטרטיבי שמכניס בכל איטרציה ערך למשתנה נוסף בצורה חמדנית.

דוגמה בדוגמה הנ"ל $\begin{pmatrix}1&0&0\\0&1&1\\1&1&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}1\\1\\1\end{pmatrix}$ אם"ם $x_1\begin{pmatrix}1\\0\\1\end{pmatrix}+x_2\begin{pmatrix}0\\1\\1\end{pmatrix}+x_3\begin{pmatrix}0\\1\\1\end{pmatrix}=\begin{pmatrix}1\\1\\1\end{pmatrix}$.

נציב איזשהו ערך ב-x_1.

- $x_1=0$: נקבל תת בעיה $\begin{pmatrix}0&0\\1&1\\1&1\end{pmatrix}\begin{pmatrix}x_2\\x_3\end{pmatrix}=\begin{pmatrix}1\\1\\1\end{pmatrix}$ (הצבנו את x_1 והחלנו את ערכו על b, במקרה הזה הוא 0 ולכן לא משפיע).
- $x_1=1$: נקבל תת בעיה $\begin{pmatrix}0&0\\1&1\\1&1\end{pmatrix}\begin{pmatrix}x_2\\x_3\end{pmatrix}=\begin{pmatrix}0\\1\\0\end{pmatrix}$ (כאן x_1 השפיע על הסכום הראשון והאחרון אבל לא על האמצעי כי לא סוכמים אותו במשוואה הלינארית המתאימה לו).

נבנה גישה הסתברותית שתאפשר לנו לבנות כלל החלטה חמדני לבחירה בין שתי מערכות משוואות. נגדיר מרחב הסתברות אחיד על $\mathcal{S}$, כל השמה בהסת' $\frac{1}{2^n}$. Y הוא מ"מ ונרצה לחשב את תוחלתו.

עבור $k=1$, מדובר במערכת עם משוואה אחת, כלומר $a_1x_1+\cdots+a_nx_n=b$, כאשר $r=(a_1,\ldots,a_n)$ הוא השורה היחידה של A. $Y(s)=\begin{cases}1&\langle r\mid s\rangle=b\\0&\langle r\mid s\rangle\neq b\end{cases}$.

למה עבור $k=1$, $E[Y]=\begin{cases}1&r=0,b=0\\0&r=0,b=1\\\frac{1}{2}&r\neq 0\end{cases}$.

הוכחה:

1. $r=0,b=0$ ולכן $\forall s\in\mathcal{S}$, $\langle r\mid s\rangle=0=b$ ולכן $Y(s)\equiv 1$.
2. $r=0,b=1$ ולכן $\forall s\in\mathcal{S}$, $\langle r\mid s\rangle=0\neq b$ ולכן $Y(s)\equiv 0$.
3. $r\neq 0$. נגדיר $V=\{s:\langle r\mid s\rangle=0\}$. V הוא ת"מ של $\mathbb{F}_2^n$ שמימדו הוא $n-1$ (הסטודנטית המשקיעה תיזכר בלינארית 2) ולכן $|V|=2^{n-1}$.

- אם $b=0$ אז $Y(s)=1$ אם"ם $\langle r\mid s\rangle=0$ אם"ם $s\in V$ ולכן $P(Y=1)=\frac{|v|}{2^n}=\frac{1}{2}$ ולכן $E[Y]=1\cdot P(Y=1)=\frac{1}{2}$.

• אם $b=1$ אז באופן דומה $Y\left(s\right)=1$ אם"ם $s\in V^{C}$ ולכן $P\left(Y=1\right)=\frac{\left|V^{C}\right|}{2^{n}}=\frac{1}{2}$ ושוב $E\left[Y\right]=\frac{1}{2}$.

■

הגדרה תהי $Ax=b$ מערכת משוואות לינאריות עם k משוואות. תהיינה $r_{1},\ldots,r_{k}$ השורות של A. נאמר כי המשוואה ה-i היא משוואה <u>ריקה טובה</u> אם $r_{i}=0$ ו-$b_{i}=0$ <u>וריקה רעה</u> אם $r_{i}=0$ ו-$b_{i}=1$.

למה מרכזית תהי $Ax=b$ מערכת משוואות לינאריות עם k משוואות מעל $\mathbb{F}_{2}$. יהי β מספר המשוואות הריקות הרעות במערכת ו-γ מספר המשוואות הריקות הטובות. יהי Y מ"מ על $\mathcal{S}$ שערכו מספר המשוואות ש-s מספקת. אזי $E\left[Y\right]=\frac{k+\gamma-\beta}{2}$.

הוכחה: $\forall i\in\left[k\right]$, יהי Y_{i} המשתנה מהמקרי האחראי על המשוואה ה-i, $Y_{i}\left(s\right)=\begin{cases}1 & \left\langle r_{i}\mid s\right\rangle =b_{i}\\0 & \left\langle r_{i}\mid s\right\rangle \neq b_{i}\end{cases}$. מתקיים $Y=\sum\limits_{i=1}^{k}Y_{i}$ ולכן מלינאריות התוחלת

$$E\left[Y\right]=\sum_{i=1}^{k}E\left[Y_{i}\right]=1\cdot\gamma+0\cdot\beta+\frac{1}{2}\left(k-\beta-\gamma\right)=\gamma+\frac{k-\gamma-\beta}{2}=\frac{k+\gamma-\beta}{2}$$

■

מסקנה תהי $Ax=b$ מערכת משוואות לינאריות עם k משוואות ו-n נעלמים. אזי ניתן לחשב את התוחלת של Y בזמן $\mathcal{O}\left(kn\right)$.

הוכחה: מהלמה המרכזית כדי לחשב את התוחלת מספיק למצוא את מספר המשוואות הריקות הטובות והרעות במערכת ולשם כך כל שנצטרך לעשות הוא לעבור על שורות המטריצה, אם מצאנו שורת אפסים, נבדוק האם b_{i} המתאים לה הוא 0 או 1 וכך נסווג אותה בהתאם (רעה/טובה).

■

מסקנה יהי s^{*} פתרון אופטימלי לבעיה. אזי $E\left[Y\right]\geq\frac{1}{2}Y\left(s^{*}\right)$.

הוכחה: יהיו β,γ כמקודם. $Y\left(s^{*}\right)\leq k-\beta$ כי משוואה ריקה רעה אף פעם לא תוכל להסתפק. לכן

$$E\left[Y\right]=\frac{k+\gamma-\beta}{2}\geq\frac{k-\beta}{2}\geq\frac{1}{2}Y\left(s^{*}\right)$$

■

האלג' שלנו יחשב את התוחלת של Y בתתי הבעיות בהן $s_{1}=0$ ו-$s_{1}=1$ וייבחר את ההשמה שמעניקה תוחלת גבוהה יותר למספר המשוואות המסופקות בתחום המצומצם (תחת הדרישה $s_{i}=0$ או $s_{i}=1$). פעולה זאת נעשית בזמן לינארי מהמסקנה הראשונה. לאחר מכן, באופן איטרטיבי, נבחר באופן חמדני לכל $i\in\left[n\right]$ את s_{i} (0 או 1) שמעניקה תוחלת גבוהה יותר כנ"ל במרחב הפתרונות המצומצם (תחת הדרישה הספציפית על s_{i}, ובעצם גם כל הבחירות הקודמות שבוצעו עד כה, $s_{1},\ldots,s_{i-1}$).

עבור (A',b') מערכת משוואות ב-m נעלמים ו-Y פ' האיכות המתאימה למערכת זו (המחשב כמה משוואות מסופקות על כל השמה), נסמן $\mathrm{Av}\left(A',b'\right)=E\left[Y\right]\overset{\text{הגדרה}}{=}\frac{1}{2^{m}}\sum\limits_{s\in\mathbb{F}_{2}^{m}}Y\left(s\right)$ ממוצע המשוואות המסופקות.

יהיו $y_1,\ldots,y_m$ המשתנים במערכת (A',b'). הצבת ערך 0 או 1 ל-y_1 מפצלת את המערכת לשתי מערכות אפשריות $\left(\widetilde{A},b_1\right),\left(\widetilde{A},b_0\right)$.
העקרון החמדני עתה הוא - אם $\mathrm{Av}\left(\widetilde{A},b_1\right)\geq\mathrm{Av}\left(\widetilde{A},b_0\right)$ נציב $y_1=0$ ואחרת $y_1=1$.

זה לא היה מספיק פורמלי ולכן נפרמל (באופן ריגורוזי, רפטטיבי וסזיפי).

למה (עדכון לאחר הצבה) תהי (A',b') מערכת של משוואות לינאריות עם k משוואות ו-m נעלמים. יהי $y=\begin{pmatrix}y_1\\ \vdots\\ y_m\end{pmatrix}$ וקטור משתנים של המערכת. הצבת ערך $s\in\{0,1\}$ במשתנה y_1 הופך את המערכת ל-$\left(\widetilde{A},\widetilde{b}\right)$ כאשר $\widetilde{A}$ מתקבלת מ-A' ע"י הסדרת העמודה הראשונה (c_1) של A' ועדכון $\widetilde{b}=b-s\cdot c_1$.

הוכחה: עבור $c_1,\ldots,c_m$ העמודות של A', המערכת (A',b') היא $y_1c_1+\cdots+y_mc_m=b'$ (c_i הם וקטורים, y_j סקלרים). הצבת ערך s ב-y_1 הופכת את המערכת ל-$s\cdot c_1+y_2\cdot c_2+\cdots+y_mc_m=b'$ או באופן שקול $y_2\cdot c_2+\cdots+y_mc_m=b'-sc_1$, מה שמתאים ל-$\left(\widetilde{A},\widetilde{b}\right)$ בה $\widetilde{A}$ מתקבלת מ-A' ע"י מחיקת העמודה הראשונה ו-$\widetilde{b}=b'-sc_1$. ■

מסקנה (פיצול לאחר הצבה) כאשר מציבים ערכים 0 ו-1 ב-y_1 מקבלים שתי מערכות $\left(\widetilde{A},b_0=b'\right)$ ו-$\left(\widetilde{A},b_1=b'-c_1\right)$ כאשר $\widetilde{A}$ מתקבלת ע"י הסרת העמודה הראשונה מ-A'.

למה (ממוצע המערכות שהתפצלו) $\frac{\mathrm{Av}\left(\widetilde{A},b_0\right)+\mathrm{Av}\left(\widetilde{A},b_1\right)}{2}=\mathrm{Av}\left(A',b'\right)$.

הוכחה: $\mathrm{Av}\left(\widetilde{A},b_0\right)=\frac{1}{2^{m-1}}\sum\limits_{s\in\mathbb{F}^m:s_1=0}Y\left(s\right)$ ובאותו האופן עבור $\mathrm{Av}\left(\widetilde{A},b_1\right)$.

$$\begin{aligned}\frac{\mathrm{Av}\left(\widetilde{A},b_0\right)+\mathrm{Av}\left(\widetilde{A},b_1\right)}{2}&=\frac{\frac{1}{2^{m-1}}\sum\limits_{s\in\mathbb{F}^m:s_1=0}Y\left(s\right)+\frac{1}{2^{m-1}}\sum\limits_{s\in\mathbb{F}^m:s_1=1}Y\left(s\right)}{2}\\&=\frac{1}{2^m}\left(\sum_{s\in\mathbb{F}^m:s_1=0}Y\left(s\right)+\sum_{s\in\mathbb{F}^m:s_1=1}Y\left(s\right)\right)\\&=\frac{1}{2^m}\sum_{s\in\mathbb{F}^m}Y\left(s\right)=\mathrm{Av}\left(A',b'\right)\end{aligned}$$

■

אלגוריתם מקרב נציג פירמול של האלג' החמדן שבו דנו עד כה, שיהווה אלג' 2-מקרב לבעיה.

1. אתחול: נגדיר $A'=A,b'=b$.
2. איטרציה: לכל $0\leq t\leq n-1$, באיטרציה ה-$(t+1)$ של האלג' נציב ערך במשתנה x_{t+1} אחרי שהצבנו כבר את הערכים $x_1,\ldots,x_t$. ההצבה של הערכים 0 ו-1 למשתנה x_{t+1} מפצלת את המערכת (A',b') לשתי מערכות $\left(\widetilde{A},b_1\right),\left(\widetilde{A},b_0\right)$ כפי

שתיארנו במסקנה הנ"ל. אם $\mathrm{Av}\left(\widetilde{A}, b_0\right) \geq \mathrm{Av}\left(\widetilde{A}, b_1\right)$ נגדיר $s_{t+1} = 0$ ואחרת $s_{t+1} = 1$. נציב $x_{t+1} = s_{t+1}$ ונעדכן $A' = \widetilde{A}$ ו-$b' = b_{s_{t+1}}$, $t = t + 1$.

3. עצירה: כאשר $t = n$ נעצור ונחזיר את s.

משפט $Y(s) \geq \frac{1}{2} Y(s^*)$, כלומר האלג' הנ"ל הוא 2-מקרב.

הוכחה: עבור $0 \leq t \leq m$ נסמן ב-(A_t, b_t) את מערכת המשוואות אחרי t איטרציות שלהאלג'. נסמן $E_t = \mathrm{Av}(A_t, b_t)$.

יהי $0 \leq t \leq m - 1$. נוכיח כי מתקיים $E_{t+1} \geq E_t$, כלומר שממוצע המשוואות המסופקת רק עולה לאחר כל בחירה.

משם נסיים כי הרי $E_t = \frac{1}{2^{m-t}} \sum_{r \in \mathbb{F}_2^m : r_1 = s_1, \ldots, r_t = s_t} Y(r)$ ובפרט $E_0 = \mathrm{Av}(A, b)$ וכן $E_m = Y(s)$ ואז נקבל

$$Y(s) = E_m \geq \cdots \geq E_0 = \mathrm{Av}(A, b) \geq \frac{1}{2} Y(s^*)$$

נסמן ב-$\left(\widetilde{A}, b_1\right), \left(\widetilde{A}, b_0\right)$ את שתי המערכות המתקבלות אחרי ההצבה של ערכים 0 ו-1 למשתנה x_{t+1} באיטרציה ה-$t + 1$ של האלג' אזי מהעיקרון החמדני של האלג',

$$\begin{aligned}
E_{t+1} &= \mathrm{Av}(A_{t+1}, b_{t+1}) \\
&= \max\left\{\mathrm{Av}\left(\widetilde{A}, b_0\right), \mathrm{Av}\left(\widetilde{A}, b_1\right)\right\} \\
&\geq \frac{\mathrm{Av}\left(\widetilde{A}, b_0\right) + \mathrm{Av}\left(\widetilde{A}, b_1\right)}{2} \\
&= E_t
\end{aligned}$$

■

תרגול

הגדרה יהי $G = (V, E)$ לא מכוון. <u>חתך</u> בגרף הוא חלוקה A, B של V, (כלומר $A \cap B = \varnothing$, $A \cup B = V$).

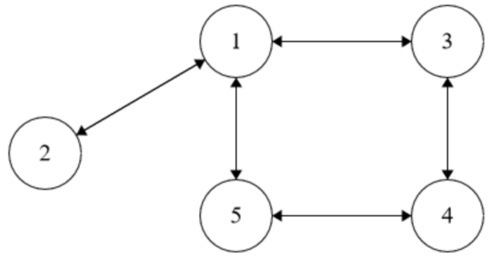
הגדרה נסמן E_C קבוצת הצלעות הנמצאת בחתך, $E_C = \{\{u, v\} \in E : u \in A \wedge v \in B\}$.

בעיית MAX-CUT

קלט גרף לא מכוון $G = (V, E)$.

פלט חתך מקסימלי בגרף (שמס' הצלעות בחתך מקסימלי).

דוגמה בגרף הבא, נבדוק כמה חתכים.



A	B	E_C
$\{1,2,3,4,5\}$	$\varnothing$	$\varnothing$
$\{1,2\}$	$\{3,4,5\}$	$\{\{2,3\},\{2,5\}\}$
$\{1,3,5\}$	$\{2,4\}$	כל הצלעות

אלגוריתם מקרב

1. נאתחל $A=V, B=\varnothing$. $\mathcal{O}(1)$
2. נעבור על כל הקודקודים לפי הסדר (v_1 עד v_n) ועבור כל קודקוד, אם מס' השכנים של הקודקוד בקבוצה שלו גדול ממס' השכנים שלו בקבוצה השנייה, נעביר את הקודקוד לקבוצה השנייה. $\mathcal{O}(|V|\cdot|E|)$
3. נחזור על שלב 2 עד שלא יותרו קודקודים שצריך להעביר.
4. נחזיר את A, B. $\mathcal{O}(1)$

משפט האלג' עוצר, חוקי ו-2-מקרב.

הוכחה: עצירה: בכל העברה אנו מעלים את מס' הצלעות בחתך (אנחנו ממקסמים את הצלעות בחתך בהן מעורב הקודקוד הנוכחי עליו אנחנו מסתכלים ולכן סה"כ מעלים את מספר הצלעות בחתך). גודל החתך חסום ע"י $|E|$ ולכן לאחר לכל היותר $|E|$ איטרציות נעצור.

חוקיות: אנו מאתחלים חתך חוקי ובכל שלב מעבירים קודקודים בין הקבוצות (ללא שכפול או מחיקה) ולכן החתך נשאר חוקי.

קירוב: נראה כי $|E_C|\geq 2s^*$. נסמן $E_{v,C}$ את קבוצת הצלעות בחתך שנובעות מהקודקוד v.

$$|E_C|=\frac{1}{2}\sum_{v\in V}|E_{v,C}|\overset{(*)}{\geq}\frac{1}{2}\sum_{v\in V}\frac{1}{2}\deg v=\frac{1}{4}\sum_{v\in V}\deg v\overset{\text{לחיצות ידיים}}{=}\frac{1}{4}(2\cdot|E|)=\frac{1}{2}|E|\geq\frac{1}{2}s^*$$

$(*)$ במהלך האלג' אנחנו מספקים לפחות חצי מהצלעות כי אם הוא בקבוצה שמספקת פחות מחצי הוא יועבר לקבוצה השנייה שתספק את כל שאר הצלעות שלו. ∎

זמן ריצה מזמני הריצה המופיעים על האלג', יחד עם העובדה ששלב 2 מתקיים לכל היותר $|E|$ פעמים, נקבל זמן ריצה של $\mathcal{O}\left(|V|\cdot|E|^2\right)$.

בעיית הסוכן הנוסע המטרי (M-TSP)

סיפור מסגרת סוכן צריך לעבור דרך כל התחנות הנתונות לו עם צריכת דלק מינימלית (פרופורציונית לקילומטר).

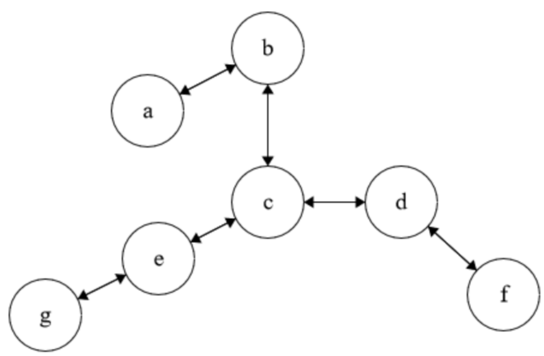
קלט גרף מלא $G=(V,E)$ ופ' משקל $w:E\to\mathbb{R}^+$ המקיימת את אי שוויון המשולש ($w(x,z)\le w(x,y)+w(y,z)$).

פלט מעגל פשוט העבור בכל קודקודי הגרף (מעגל המילטוני) בעל משקל מינימלי.

אלגוריתם מקרב

1. נמצא עפ"מ בגרף, T. $\mathcal{O}\left(|V|^2\right)$
2. נאתחל $H=\varnothing$ המעגל ההמילטוני שנחזיר. $\mathcal{O}(1)$
3. נבחר קודקוד באקראי v_i ונריץ ממנו DFS על T. אם ביקרנו בקודקוד בפעם הראשונה, נוסיף אותו ל-H. בסיום הרצת ה-DFS נוסיף את v_i ל-H. $\mathcal{O}(|V|+|E|)$
4. נחזיר את H. $\mathcal{O}(1)$

דוגמה הצלעות המסומנות הן עפ"מ.



אם נבחר את הקודוקד a אז $H=(a,b,c,e,g,d,f,a)$.

משפט האלג' חוקי ו-2-מקרב.

הוכחה: חוקיות: עוברים על כל הקודקודים בגרף ומוסיפים ל-H את הקודקוד רק אם זוה פעם הראשונה שאהנו מבקרים בו. לכן כל קודקוד מופיע בדיוק פעם אחת ב-H, פרט לראשון, שאותו אנו מוסיפים בסוף הריצה. כיוון שהגרף מלא אז קיימת צלע בין כל 2 קודקודים רצופים ב-H.

קירוב: נסמן H^* פתרון אופטימלי. נראה כי $w(H)\le 2w(H^*)$.

נגדיר F המסלול הנוצר בעזרת הליכה על T ב-DFS והוספה של כל קודקוד בו מבקרים ל-F (יש ב-F כפילויות).

טענה יהי C מסלול בגרף מלא ו-C' מסלול הנוצר מ-C ע"י הסרת קודקודים, אז $w(C')\le w(C)$ (הסרת צלעות למסלול בו המרחקים מקיימים את אי שוויון המשולש מקצרת את המסלול).

הוכחה: יהיו $x,y,z\in V$ קודקודים ב-C בסדר זה. נגדיר C' נוצר מ-C ע"י הסרת y. אז מתקיים $C'=C\setminus\left\{ \left(x,y\right),\left(y,z\right)\right\} \cup\left\{ \left(x,z\right)\right\}$ לכן

$$w\left(C'\right)=w\left(C\right)-w\left(x,y\right)-w\left(y,z\right)+w\left(x,z\right)\overset{\triangle}{\leq}w\left(C\right)$$

■

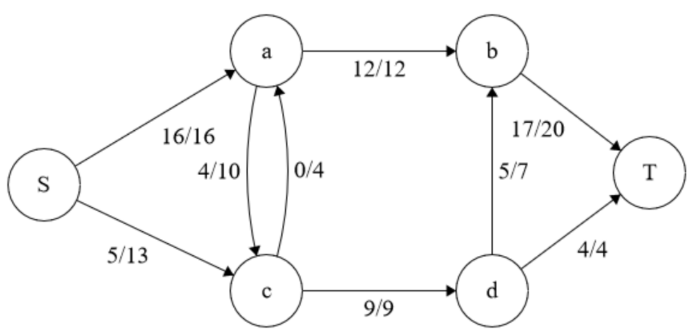
מסקנה $w\left(H\right)\leq w\left(F\right)=2w\left(T\right)$ (חזרה אחורה לקודקוד נחשבת כמעבר דרכו, לכן כל קודקוד נוסף בדיוק פעמיים ל-F).

נראה כי $w\left(T\right)\leq w\left(H^{*}\right)$. נוריד צלע מ-$H^{*}$ ונקבל עץ פורש, נסמנו T^{*} (לא בהכרח עפ"מ). אז $w\left(T^{*}\right)=w\left(H^{*}\right)-w\left(e\right)$ ולכן $w\left(T^{*}\right)\leq w\left(H^{*}\right)$ אבל T עפ"מ ו-T^{*} עץ פורש ולכן $w\left(T\right)\leq w\left(T^{*}\right)\leq w\left(H^{*}\right)$.

לסיכום, $w\left(H\right)\leq w\left(F\right)=2w\left(T\right)\leq2w\left(H^{*}\right)$. ■

זמן ריצה מזמני הריצה הרשומים באלג', סה"כ נקבל $\mathcal{O}\left(\left|V\right|^{2}\right)$.

דוגמה הנושא הבא שלנו הוא בעיית הזרימה. בהינתן מערכת צינורות עם קיבולת מקסימלית לכל צינור, נרצה לדעת מה החלוקה האופטימלית על קיבולת הצינורות שתאפשר מעבר מקסימלי של נוזל מהמקור למטרה. בגרף הבא, הגדרנו לכל צינור כמה נוזל יעבור דרכו, זה פתרון אופטימלי לבעיית הזרימה הספציפית הזו כי היא מעבירה שטף מקסימלי של נוזל.



כאן נגמר החומר לבוחן והחלק הראשון של הקורס

שבוע $\mathbb{IX}$ | בעיית הזרימה

הרצאה

חלק א׳ של ההרצאה

זרימת ברשתות (Network Flow)

הגדרה רשת זרימה היא חמישייה $N=(V,E,c,s,t)$ כאשר $G=(V,E)$ גרף מכוון, $c:E\to\mathbb{R}_{\geq 0}$ פ׳ קיבול, s,t קודקודים מיוחדים: מקור ובור בהתאמה.

הגדרה זרימה ברשת זרימה היא פ׳ $f:E\to\mathbb{R}_{\geq 0}$ המקיימת את האילוצים הבאים:

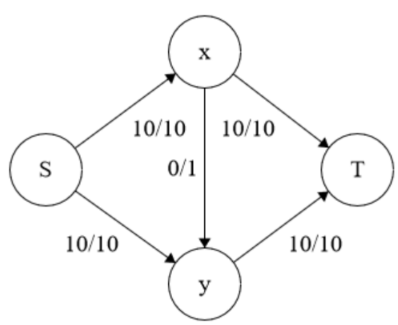
- אילוץ הקיבול: לכל צלע בגרף, הזרימה בצלע אינה גדולה מהקיבול של הצלע, כלומר $\forall e\in E$, מתקיים $f(e)\leq c(e)$.
- חוק שימור החומר: לכל קודקוד בגרף חוץ מקודקודי המקור והבור, הזרימה הנכנסת לקודקוד השווה לזרימה היוצאת מהקודקוד. כלומר, $\forall x\in V\setminus\{s,t\}$ מתקיים

$$\underbrace{\sum_{(u,x)\in E} f(u,x)}_{\text{הזרימה הנכנסת}}=\underbrace{\sum_{(x,w)\in E} f(x,w)}_{\text{הזרימה היוצאת}}$$

שטף הזרימה מוגדר כזרימה הכוללת היוצאת מקודקוד המקור, כלומר $|f|=\sum_{(s,v)\in E} f(s,v)$.

בהינתן רשת זרימה N, נרצה למצוא זרימה מקסימלית (עם $|f|$ מירבי) ברשת.

דוגמה בדוגמה הבאה, משמאל זרימה עם שטף מקסימלי ומימין הקיבול לכל צלע.



הבעיה ניתנת לפתרון באמצעות תכנון לינארי ונפתרה ב-56׳ ע״י Ford & Fulkerson ולאחר מכן גם ע״י Edmonds & Karp.

הערות על מרחב הפתרונות

1. מרחב הפתרונות היחודיים לבעיה זו הוא מרחב הזרימות החוקיות ברשת N.

2. הוא לעולם לא ריק כי $f \equiv 0$ היא תמיד זרימה חוקית.

3. המרחב הזה יכול להיות אינסופי: נשים לב כי אם f, g זרימות חוקיות אז גם $\frac{f+g}{2}$ זרימה חוקית (ברור שאילוץ הקיבול מתקיים והסטודנטית המשקיעה תיווכח שגם שימור החומר מתקיים). באופן כללי, $h = \lambda f + (1-\lambda) g$ הוא זרימה חוקית ולכן מרחב הזרימות החוקיות הוא קבוצה קמורה (ובפרט היא אינסופית).

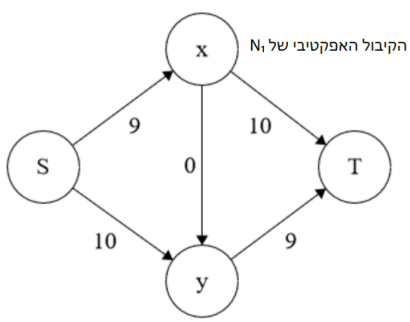
4. למרות שמרחב הפתרונות אינסופי, יש פתרון אופטימלי כי מרחב הפתרונות הוא קבוצה סגורה וחסומה ולכן קומפקטית ופ' השטף היא פ' רציפה על זרימות ולכן מאינפי 3 קיים לקבוצה מקסימום.

רעיון כדי לבנות זרימה לא טריוויאלית (שונה מאפס) ברשת, נמצא באמצעות BFS מסלול פשוט בין s ל-t ונשלח בו זרימה קבועה השווה לקיבול המינימלי של צלעות המסלול, ואפס על השאר. אילוץ הקיבול מתקיים כי הזרימה שווה לקיבול המינימלי במסלול ואילו שימור החומר מתקיים כי אנחנו מעבירים שטף קבוע על מסלול כלשהו.

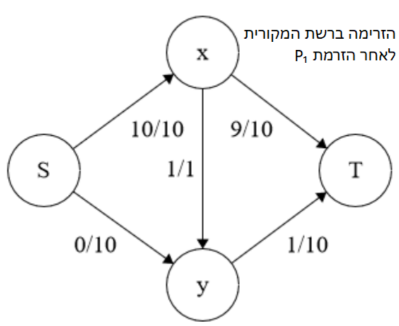
אלגוריתם אלג' איטרטיבי שבכל שלב מוצא מסלול פשוט בין s ל-t, שולח בו זרימה קבועה ומוסיף את הזרימה הדשה הזו לזרימה שקיימת כרגע ברשת. תוך כדי שהוא דואג לשמור על אילוץ הקיבול.

דוגמה נריץ את האלג' על הגרף בדוגמה הנ"ל.

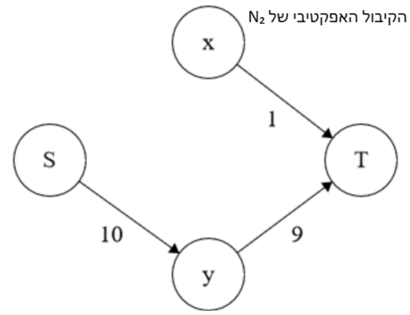
המסלול הראשון הוא $P_0 : s, x, y, t$ ונזרים דרכו 1 (הקיבול המינימלי) לכן הקיבול האפקטיבי של הרשת לאחר ההזרמה, N_1, נהייה



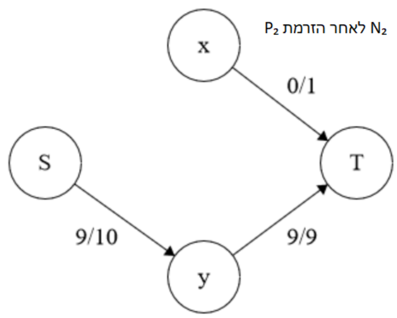
עתה נבחר $P_1 : s, x, t$ ונזרים דרכו 9 ונקבל שכרגע אנחנו מזרימים (ברשת המקורית)



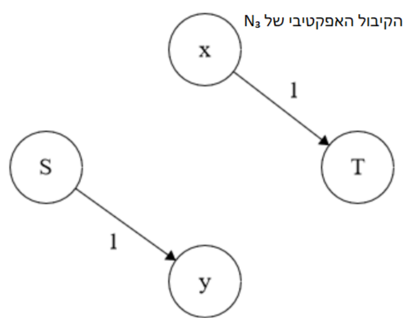
הרשת האפקטיבית החדשה, N_2, היא



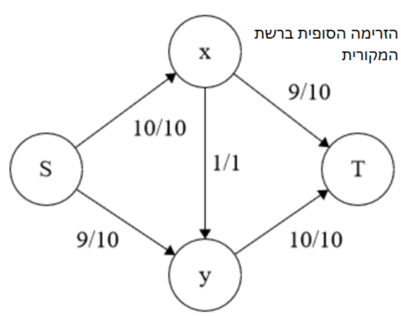
נבחר את המסילה $P_2 : s,y,t$ ונזרים דרכה 9 ולכן הזרימה שלנו כרגע היא



והרשת האפקטיבית עם הקיבול היא עכשיו



ואין כאן מסילה בין s ל-t ולכן סיימנו את ריצת האלגוריתם, עם זרימה סופית



נשים לב שזה אינו פתרון אופטימלי.

למה לכל רשת זרימה קיימת זרימה אופטימלית f כך שלא קיימים שני קודקודים $V \in x,y$ כך ש-$E \in (y,x),(x,y)$ וגם $f(x,y), f(y,x) > 0$, כלומר לא קיימים שני קודקודים עם תנועה בין שניהם כי ניתן לקזז אותה.

הערה נניח הנחות מקלות על הרשת:

- אין בגרף (V,E) של הרשת מעגלים.
- אין צלעות הנכנסות למקור.
- אין צלעות היוצאות מהבור.

הוכחה: תהי g זרימה אופטימלית ברשת. נגדיר $B=\{(x,y)\in E:(y,x)\in E, g(x,y)>0, g(y,x)>0\}$. נשים לב כי $(x,y)\in B$ אם"ם $(y,x)\in B$. נגדיר פ' חדשה f על E באופן הבא:

$$f(e)=\begin{cases} g(e) & e\notin B \\ g(x,y)-\min\{g(x,y),g(y,x)\} & (x,y)=e\in B \end{cases}$$

נוכיח כי f היא זרימה חוקית ברשת כי $|f|=|g|$ ולכן גם f אופטימלי וכי ל-f יש את התכונה הנדרשת בלמה.

1. f זרימה חוקית: f אי-שלילית מהגדרתה. f מקיימת את אילוץ הקיבול כי $\forall e\in E, f(e)\leq g(e)$. f מקיימת את חוק שימור החומר כי $\forall x\in V\setminus\{s,t\}$ במעבר מ-g ל-f הזרימה הנכנסת ל-x והיוצאת מ-x קטנות באותה מידה (אם הקטנו את הזרימה של צלע נכנסת זה אומר שהקטנו את הזרימה של צלע יוצאת באותה כמות ולהפך).
2. $|f|=|g|$: הנחנו כי אין צלעות נכנסות לקודקוד המקור s. לכן לא קיים $x\in V$ כך ש-$(x,s)\in E$, ולכן לכל צלע $(s,x)\in E$ מתקיים $f(s,x)=g(s,x)$. לכן

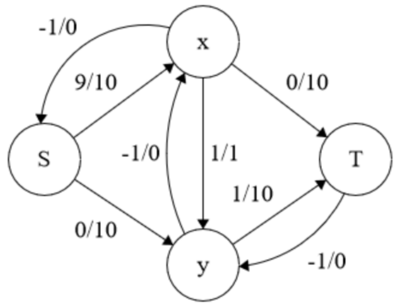
$$|f|=\sum_{(s,x)\in E}f(s,x)=\sum_{(s,x)\in E}g(s,x)=|g|$$

3. קיום תכונת הלמה: אם $(x,y)\in B$ אז מהגדרת f, $\min\{f(x,y),f(y,x)\}=0$ ואם $(x,y)\notin B$ לא ייתכן ש-$\min\{f(x,y),f(y,x)\}>0$ כי $\min\{g(x,y),g(y,x)\}=0$.

■

חלק ב' של ההרצאה

כדי לשפר את האלג' שלנו, שלא מחזיר פתרון אופטימלי כרגע, בכל פעם שנעביר זרם דרך צלע כלשהי, נגיד צלע חדשה בין אותם קודקודים רק בכיוון ההפוך עם קיבול אפקטיבי זהה. בגלל שאנחנו לא רוצים לאפשר קיבול אמיתי נוסף, ניתן לצלעות הפיקטיביות הללו קיבול 0 ונזרים דרכם את מינוס הזרם בכיוון ההפוך (ראו איור לדוגמה). כך אנו פותחים דרכים חדשות בהן נוכל להגדיר זרימות וזה יעזור לנו באלג'.



הגדרות

1. רשת זרימה מורחבת זו הבעיה $N' = (V, c', s, t)$ כאשר V היא קבוצת הקודקודים של הרשת, $s \in V$ הוא קודקוד המקור ו-$t \in V$ קודקוד הבור. $c' : V \times V \to \mathbb{R}_{\geq 0}$ היא פ' הקיבול המורחב המקיימת $\forall v \in V$

$$c'(x, x) = 0, c'(x, s) = 0, c'(t, x) = 0$$

2. תהי $N = (V, E, c, s, t)$ רשת זרימה רגילה. נגדיר ההרחבה של רשת הזרימה הזו היא $N' = (V, c', s, t)$ כאשר

$$c'(x, y) = \begin{cases} c(x, y) & (x, y) \in E \\ 0 & (x, y) \notin E \end{cases}$$

3. תהי $N' = (V, c', s, t)$ רשת זרימה מורחבת. זרימה מורחבת ברשת זו היא פ' $f' : V \times V \to \mathbb{R}$ המקיימת את שלושת התכונות הבאות:

   (א) אנטי סימטריה: $f'(y, x) = -f'(x, y)$, $\forall x, y \in V$.

   (ב) אילוץ הקיבול: $f'(x, y) \leq c'(x, y)$, $\forall x, y \in V$.

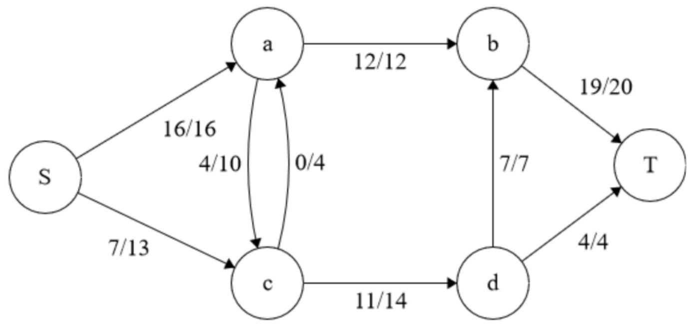
   (ג) חוק שימור החומר: $\sum_{v \in V} f'(x, v) = 0$, $\forall x \in V \setminus \{s, t\}$.

4. תהיינה $N = (V, E, c, s, t)$ זרימה רגילה, $N' = (V, c', s, t)$ הרחבה של N ו-f זרימה רגילה ברשת N. ההרחבה של f לרשת המורחבת N' מוגדרת ע"י

$$f'(x, y) = \begin{cases} f(x, y) & (x, y) \in E \wedge f(x, y) > 0 \\ -f(y, x) & (y, x) \in E \wedge f(y, x) > 0 \\ 0 & \text{אחרת} \end{cases}$$

תרגול

דוגמה נסתכל על זרימה בעלת שטף מקסימלי (כל הדרכים שמובילות אליה הגיעו לרוויה ולכן לא נוכל להעביר עוד חומר אל הבור.



נלמד בהרצאה שני אלגוריתמים:

- פורד-פולקרסון (FF), שרץ ב-$\mathcal{O}\left(|E|\cdot|f^*|\right)$ כאשר f^* הזרימה האופטימלית (כלומר לא בהכרח פולינומיאלי בקלט בכלל).
- אדמונד-קארפ (EK), שרץ ב-$\mathcal{O}\left(|E|^2|V|\right)$ ובנוסף בהינתן שהקיבול שלם גם הפתרון שהוא יחזיר יהיה שלם.

רידוקציה לבעיית הזרימה

1. נבנה רשת זרימה המתאימה לבעיה.
2. נריץ את אלג' EK על הרשת שבנינו ונקבל זרימה מקסימלית f.
3. נחזיר פתרון לבעיה המקורית על סמך f.

כדי להוכיח נכונות נוכיח את הדברים הבאים:

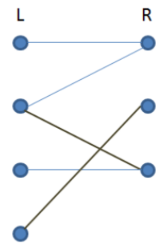
1. כל פתרון חוקי של הבעיה ניתן לתרגום לזרימה חוקית ברשת (מוודא שלא נפספס פתרונות חוקיים של הבעיה, ובפרט האופטימלי).
2. כל זרימה חוקית ברשת ניתנת לתרגום לפתרון חוקי של הבעיה (מוודא שהפתרון לבעיה חוקי).
3. המרה בין פתרונות משמרת את ערך הבעיה (משמש להוכחת אופטימליות).

בעיית מציאת זיווג מקסימלי

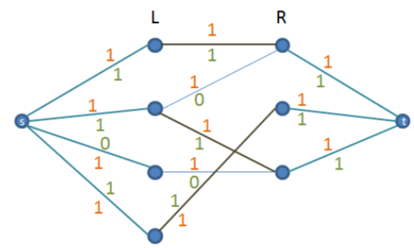
קלט גרף דו"צ לא מכוון $G=\left(L\dot{\cup}R,E\right)$.

פלט זיווג בגודל מקסימלי בגרף.

דוגמה נדגים את הרידוקציה על מקרה פרטי. הגרף שלנו הוא



ונגדיר עליו רשת זרימה באופן הבא



כאשר המספרים בכתום הם הקיבולים (כרגע נתעלם מהירוקים), s,t הם קודקודים חדשים וכל הצלעות הן מכוונות ימינה, כלומר, הצלעות הן או מ-s לקודקודים ב-L, או מקודקודים ב-L לקודקודים ב-R או מקודקודים ב-R ל-t.

הירוקים הם ערכי זרימה מקסימלית בגרף. אם נבחר רק את הצלעות שזורם דרכן 1 (ולא אפס) נקבל שידוך (מקסימלי).

אלגוריתם

1. נבנה רשת זרימה $N=(V,E',c,s,t)$ כאשר:
   - $V'=L\cup R\cup\{s,t\}$.
   - $E'=E_L\cup\vec{E}\cup E_R$ כאשר

$$E_L=\{(s,u):u\in L\}\,,\vec{E}=\{(u,v):\{u,v\}\in E\}\,,E_R=\{(u,t):u\in R\}$$

(פשוט חלוקה לתחומים).

   - $\forall e\in E'$,$c(e)=1$.
2. נריץ על הרשת שבנינו את אלג' EK ונקבל זרימה בעלת שטף מקסימלי, f.
3. נחזיר את הזיווג $M=\left\{e\in\vec{E}:f(e)=1\right\}$.

משפט האלג' הנ"ל הוא חוקי ואופטימלי.

הוכחה: חוקיות: נראה כי M זיווג חוקי. נניח בשלילה ש-M לא זיווג חוקי, לכן קיים $v\in L\cup R$ כך ש-2 צלעות ב-M נוגעות (מכילות) ב-v.

אם $v\in R$ ($v\in L$) אז יש 2 צלעות שנכנסות (יוצאות) ל-v ומהגדרת M יש שתי צלעות ברשת הזרימה שנכנסות (יוצאות מ-) ל-v והזרמנו בהם זרימה בגודל 1 ולכן סך הזרימה הנכנסת (היוצאת) ל-v הוא 2 אבל הזרימה היוצאת (נכנסת) היא לכל היותר 1 בסתירה לחוק שימור החומר.

אופטימליות: נניח בשלילה שקיים M' כך ש-$|M'| > |M|$.

1. נוכיח כי $|f| = |M|$.

$$|f| = \sum_{u \in L} f(s,u) \overset{(*)}{=} \sum_{(u,v) \in \vec{E}} f(u,v) = \sum_{(u,v) \in \vec{E} : f(u,v)=1} 1 = |M|$$

$(*)$ מחוק שימור החומר, כל הזרימה שעוברת ב-E_L חייבת לעבור דרך הצלעות ב-$\vec{E}$.

2. נוכיח כי קיימת זרימה חוקית ברשת f' כך ש-$|f'| = |M'|$.

נסמן $L' \subseteq L$ קבוצת הקודקודים ב-L שקיימת צלע ב-M' שנוגעת בהם ובאותו האופן $R' \subseteq R$. נגדיר $\forall e \in E'$,

$$f'(e) = \begin{cases} 1 & e \in M \cup \{(s,u) : u \in L'\} \cup \{(u,t) : v \in R'\} \\ 0 & \text{otherwise} \end{cases}$$

(א) נראה כי f' זרימה חוקית ברשת.

i. אילוץ הקיבול: f' מזרימה לכל היותר יחידת חומר אחת, לכן $\forall e \in E'$,$f'(e) \le 1 = c(e)$.

ii. חוק שימור החומר: לכל קודקוד $v \in L' \cup R'$, f' מזרימה יחידת חומר אחת ל-v ויחידת חומר אחת מ-v ולכן חוק שימור החומר נשמר. M' זיווג חוקי ולכן לא קיימות 2 צלעות שהזרימה הן 1 ושתיהן נכנסות/יוצאות מ-v. כל קודקוד אחר f' מזרימה 0 ממנו ואליו.

(ב) נראה כי $|M'| = |f'|$.

$$|f'| = \sum_{u \in L} f(s,u) = \sum_{u \in L'} f(s,u) + \sum_{u \in L \setminus L'} f(s,u) = \sum_{u \in L'} 1 = |L'| = |M'|$$

■

לכן $|f| = |M| < |M'| = |f'|$ בסתירה למקסימליות f.

שבוע $\mathbb{X}$ | האלגוריתם של פורד פלקורסון

הרצאה

חלק א' של ההרצאה

האלג' הנאיבי שהצגנו לא עבד כי הוא סגר מסלולים שאולי היינו רוצים להשתמש בהם בהמשך.

הרעיון של Ford & Fulkerson הוא שברגע שמזרימים זרימה בין x ל-y, פותחים צלע עם קיבול מתאים בכיוון ההפוך ברשת כדי שבמקרה הצורך נוכל להחזיר בה זרימה. על הצלעות בכיוון ההפוך אפשר לחשוב כמסמנות "כמה זרימה כבר העברנו באיטרציות קודמות".

הגדרה <u>שטף של זרימה מורחבת</u> f' מוגדרת ע"י $|f'| = \sum\limits_{v \in V} f'(s,v)$.

למה 1 תהי N רשת זרימה רגילה ותהי f זרימה רגילה ב-N. תהי N' ההרחבה של N ו-f' ההרחבה של f ל-N'. אזי N' היא רשת זרימה מורחבת ו-f' היא זרימה מורחבת ברשת זו ומתקיים $|f'| = |f|$.

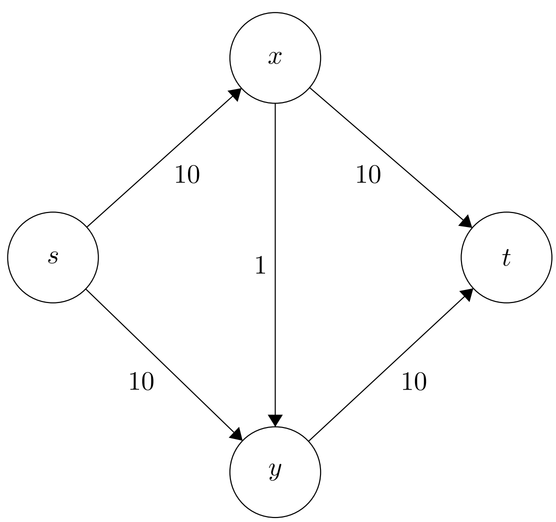
הגדרה תהי $N' = (V, c', s, t)$ רשת זרימה מורחבת. נגדיר $E = \{(x,y) : c'(x,y) > 0\}$, $c : E \to \mathbb{R}_+$ ע"י $c(x,y) = c'(x,y)$, $\forall (x,y) \in E$. נגדיר $N = (V, E, c, s, t)$ להיות <u>הצמצום</u> של הרשת N'.

הגדרה תהי N' רשת זרימה מורחבת ו-$N = (V, E, c, s, t)$ הצמצום של N'. תהי f' זרימה מורחבת ב-N'. נגדיר $f : E \to \mathbb{R}_{\geq 0}$, <u>הצמצום</u> של f', ע"י $f(x,y) = \max\{f'(x,y), 0\}$, $\forall (x,y) \in E$.

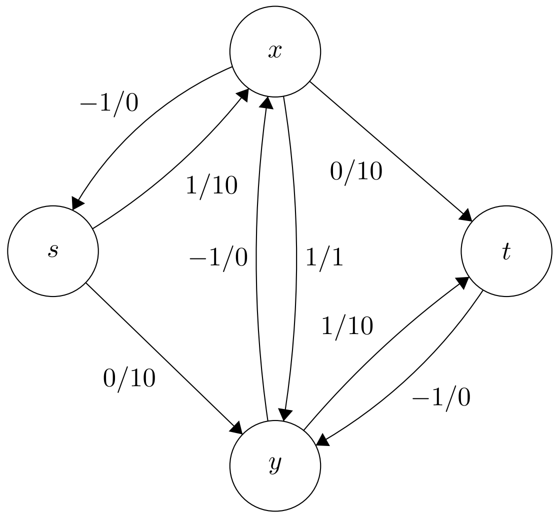
למה 2 תהי N' רשת זרימה מורחבת ותהי f' רשת זרימה ב-N'. תהי N הצמצום של N' ו-f הצמצום של f'. אזי:

1. N היא רשת זרימה רגילה ו-f היא זרימה רגילה ברשת זו.
2. $|f| = |f'|$.
3. אם N' הרחבה של N אז הצמצום של N' הוא N.

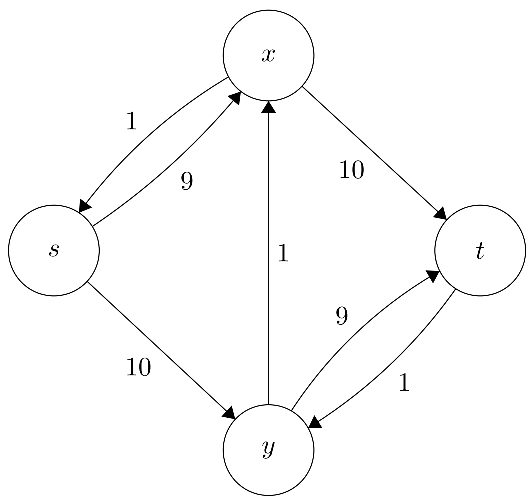
דוגמה נריץ את הרעיון על הדוגמה המקורית, עם המסלול שבזמנו הרס לנו את הריצה הנאיבית.



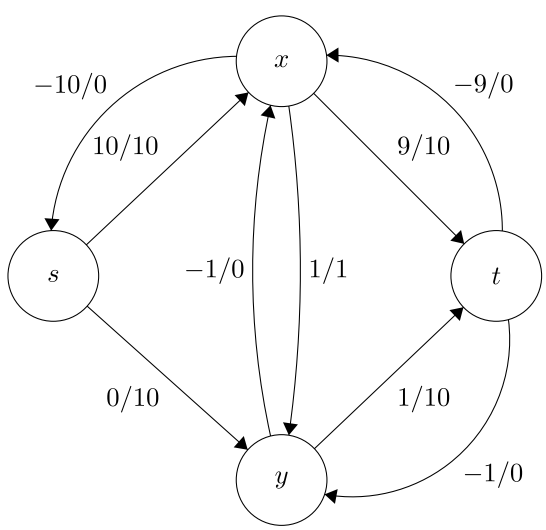
לאחר הזרמת 1 דרך המסלול s, x, y, t נקבל עם הצלעות החדשות את הגרף



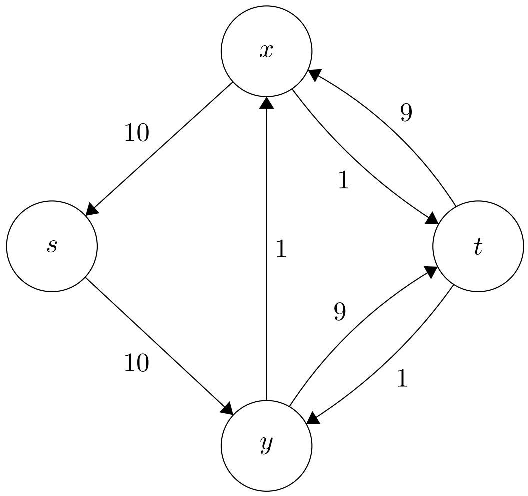
לכן עכשיו נריץ את האלג' על הרשת החדשה, N_1 שהיא



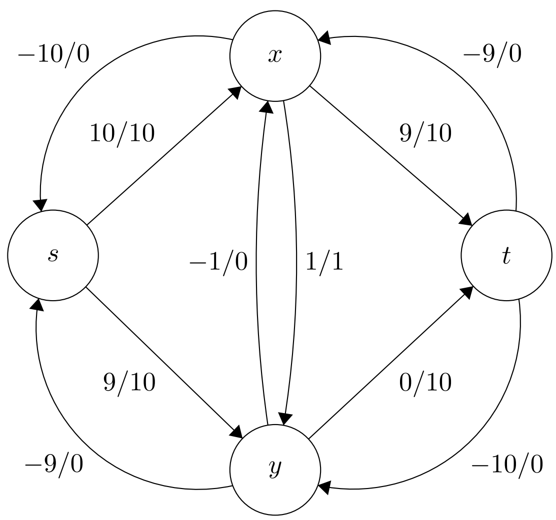
נבחר את המסלול s, x, t ונזרים דרכו 9, עכשיו הרשת המורחבת המתקבלת היא



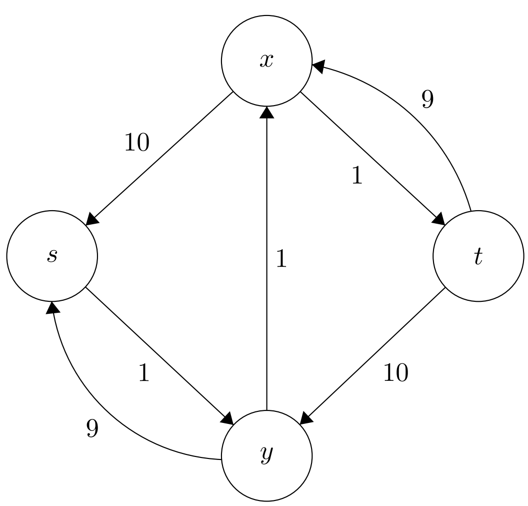
לכן עכשיו הרשת הרגילה החדשה היא N_2, שהיא



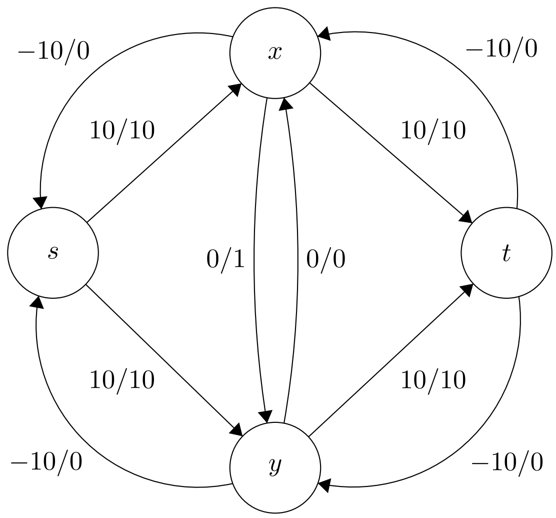
נזרים דרך s, y, t זרימה של 9 ונקבל רשת מורחבת



עתה הרשת הפאקטיבית החדשה N_3 היא



באלג' הנאיבי כאן היינו עוצרים כי לא פתחנו צלעות חדשות. אבל אנחנו נוכל לבחור את המסלול s,y,x,t ולהזרים בו 1 בעזרת הצלעות החדשות.



ועכשיו אין יותר מסלולים ולכן נפסיק את האיטרציה ואכן אם נמצמם את הזרימה המורחבת ברשת המורחבת נקבל את הזרימה המקסימלית המוכרת לנו מהדוגמה מההרצאה הקודמת.

הערה מעתה נעבוד רק עם זרימות מורחבות ולכן נסיר את ה-' מכל הסימונים.

הגדרות לכתיבה הפורמלית של F&F

תהי $N=(V,c,s,t)$ רשת זרימה מורחבת, f זרימה מורחבת על רשת מורחבת זו.

1. <u>הקיבול השיורי</u> הוא פ' $c_f:V\times V\to\mathbb{R}_{\geq 0}$ המגודרת ע"י $c_f(x,y)=c(x,y)-f(x,y)$, $\forall(x,y)\in E$.
2. <u>הרשת השיורית</u> היא הרביעיה $N_f=(V,c_f,s,t)$.
3. <u>אוסף הצלעות השיורי</u> הוא $E_f=\{(x,y):c_f(x,y)>0\}$.
4. <u>הגרף השיורי</u> הוא $G_f=(V,E_f)$.
5. <u>מסילת הרחבה</u> זו מסילה פשוטה מ-s ל-t בגרף השיורי G_f.
6. <u>הקיבול השיורי של מסילת הרחבה</u> p הוא $c_f(p)=\min_{e\in p}\{c_f(e)\}$.
7. <u>הזרימה השיורית במסילת הרחבה</u> p היא פ' $\Delta_{f,p}:V\times V\to\mathbb{R}$ המוגדר ע"י $\Delta_{f,p}(x,y)=\begin{cases}c_f(p) & (x,y)\in p\\ -c_f(p) & (y,x)\in p\\ 0 & \text{otherwise}\end{cases}$.

הערה הזרימה השיורית היא למעשה הקיבול האפקטיבי לאחר האיטרציה.

האלגוריתם של פורד ופולקרסון למציאת זרימה אופטימלית ברשת זרימה

קלט רשת זרימה רגילה $\widetilde{N}$.

פלט זרימה אופטימלית g ב-$\widetilde{N}$.

פסאודו-קוד

1. הרחבה: נרחיב את רשת $\widetilde{N}$ לרשת מורחבת N.
2. אתחול: נאתחל את f להיות זרימת האפס המורחבת ב-N.
3. איטרציה: בכל שלב נמצא מסילת הרחבה p בגרף השיורי G_f ונעדכן $f = f + \Delta_{f,p}$.
4. עצירה: כאשר אין מסילות הרחבה בגרף השיורי, נעצור.
5. צמצום: נצמצם את הרשת N חזרה לרשת הרגילה $\widetilde{N}$ ואת הזרימה המורחבת f לזרימה רגילה g ברשת $\widetilde{N}$.
6. סיום: נחזיר את g.

חלק ב' של ההרצאה

משפט (לא פורמלי) האלגוריתם של F&F מחזיר זרימה חוקית ואופטימלית.

הערה אי-הפורמליות של המשפט היא שהאלג' לא בהכרח עוצר, אבל אם הוא כן אז הוא מחזיר זרימה חוקית ואופטימלית.

הוכחה: חוקיות:

למה תהי N רשת זרימה (מורחבת) ותהי f זרימה ברשת זו. אזי:

1. הרשת השיורית N_f היא רשת זרימה (מורחבת) חוקית (כלומר שהיא אי-שלילית ומאפסת לולאות, צלעות הנכנסות למקור והיוצאות מהבור).
2. הזרימה השיורית $\Delta_{f,p}$ היא זרימה חוקית ברשת השיורית ו-$|\Delta_{f,p}| = c_f(p)$ ($\Delta_{f,p}$ מתעלמת מכל הצלעות שלא במסלול).
3. הזרימה $f_1 = f + \Delta_{f,p}$ היא זרימה חוקית ברשת N ו-$|f_1| = |f| + c_f(p)$.

הוכחה:

1. מספיק לוודא כי עבור $x, y \in V$ מתקיים $c_f(x,y) \geq 0$. מכיוון ש-f- זרימה מורחבת חוקית הרי ש-$f(x,y) \leq c(x,y)$ ולכן $c_f(x,y) = c(x,y) - f(x,y) \geq 0$. שאר התכונות טריוויאליות.

2.

- אנטי סימטריה נובעת מההגדרה.
- אילוץ הקיבול: יהיו $x, y \in V$.

– אם הצלע $p \ni \widetilde{e} = (x,y)$. אזי מהגדרת הזרימה השיורית, $\Delta_{f,p}(\widetilde{e}) = c_f(p) = \min_{e\in p}\{c_f(e)\} \le c_f(\widetilde{e})$.

– אם $(x,y) \notin p$ אזי $\Delta_{f,p}(x,y) \le 0 \le c_f(x,y)$.

- חוק שימור החומר: יהי $x \in V\setminus\{s,t\}$.

– אם x על המסילה אז בגלל ש-$x \neq s,t$, קיים קודקוד y הקודם ל-x על המסילה וקודקוד z הבא אחרי x על המסילה. במקרה זה יוצאות מ-x בדיוק שתי צלעות עם זרימה שיורית שונה מ-0 (הזרימה השיורית שונה מאפס רק על המסילה): $(x,y),(x,z)$. לכן

$$\sum_{v\in V}\Delta_{f,p}(x,v) = \Delta_{f,p}(x,y) + \Delta_{f,p}(x,z) = \underbrace{-c_f(p)}_{\text{נגד כיוון המסילה}} + \underbrace{c_f(p)}_{\text{עם כיוון המסילה}}$$

– אם x לא על המסילה אז כל הצלעות היוצאות מ-x הן בעלות זרימה שיורית 0 ולכן סכומן הוא גם אפס.

השטף של $\Delta_{f,p}$: יהי x הקודקוד על המסילה הבא אחרי קודקוד המקור s. אזי הצלע (s,x) היא היחידה שיוצאת מ-s ובעלת זרימה שיורית שונה מ-0 ולכן

$$|\Delta_{f,p}| = \sum_{v\in V}\Delta_{f,p}(s,v) = \Delta_{f,p}(s,x) = c_f(p)$$

.3

- אנטי סימטריה של $f_1 = f + \Delta_{f,p}$. יהיו $x,y \in V$. אזי

$$f_1(y,x) = f(y,x) + \Delta_{f,p}(y,x) = \underbrace{-f(x,y)}_{\text{זרימה חוקית ב-}N\ f} + \left(\underbrace{-\Delta_{f,p}(x,y)}_{\text{זרימה חוקית ב-}N_f\ \Delta_{f,p}}\right) = -f_1(x,y)$$

- שימור החומר של f_1: יהי $x \in V$.

$$\sum_{v\in V} f_1(x,v) = \sum_{v\in V} f(x,v) + \sum_{v\in V}\Delta_{f,p}(x,v) = 0 + 0 = 0$$

- אילוץ הקיבול: יהיו $x,y \in V$. אזי

$$\begin{aligned} f_1(x,y) &= f(x,y) + \Delta_{f,p}(x,y) \\ &\le f(x,y) + c_f(x,y) \\ &= f(x,y) + (c(x,y) - f(x,y)) = c(x,y) \end{aligned}$$

$$
\begin{aligned}
|f_1| &= \sum_{v\in V} f_1(s,v) \\
&= \sum_{v\in V} \left(f(s,v) + \Delta_{f,p}(s,v)\right) \\
&= \sum_{v\in V} f(s,v) + \sum_{v\in V} \Delta_{f,p}(s,v) \\
&= |f| + |\Delta_{f,p}| = |f| + c_f(p)
\end{aligned}
$$

■

■

מסקנה בכל שלב (לאחר כל איטרציה של האלג') זורמת ברשת זרימה חוקית. בנוסף, כל איטרציה מעלה את שטף הזרימה ברשת.

הערה זו מעיין הוכחת חוקיות באינווריאנטה.

תרגול

הגדרה חתך ברשת זרימה הוא חלוקה של הקודקודים ל-2 קבוצות זרות $V = S\dot{\cup}T$ כך ש-$s \in S$, $t \in T$.

הגדרה קיבול של חתך (S,T) מוגדר ע"י $c(S,T) = \sum_{(u,v)\in E} c(u,v)$.

טענה כל חתך (S,T) ברשת זרימה וכל זרימה חוקית f ברשת מקיימים $|f| \le c(S,T)$.

הערה אינטואטיבית, הזרימה מתחילה מ-s ולכן מ-S ועד הסוף היא צריכה לעבור ל-t, כלומר לקודקודים ב-T. לכן לא יכול להיות שעברה יותר זרימה מאשר הקיבול של הצלעות.

משפט (משפט השטף והחתך) קיים חתך (S,T) וקיימת זרימה f שמקיימים $|f| = c(S,T)$.

מסקנה אם מצאנו חתך (S,T) וזרימה חוקית f כך שמתקיים $|f| = c(S,T)$ אז בהכרח ש-f זרימה מקסימלית ו-(S,T) חתך בעל קיבול מינימלי.

הערה קיים אלג' שמוצא את החתך המינימלי ב-$\mathcal{O}\left(|E|^2 |V|\right)$.

בעיית המשקיעות והשחקנים

סיפור מסגרת רוצים להפיק סרט ומשקיעות מוכנות להשקיע את הונן אם"ם כל השחקנים שהן אוהבות נמצאים בסרט.

קלט $A=\{a_1,\dots,a_n\}$ קבוצת שחקנים כך שלכל $a_i\in A$ נתונה משכורת s_i שאותה הוא דורש. $B=\{b_1,\dots,b_k\}$ קבוצת המשקיעות כך שלכל $b_i\in B$ נתון סכום d_i אותו היא מוכנה להשקיע רק אם כל השחקנים שהיא אוהבת $A_i\subseteq A$ משתתפים בסרט.

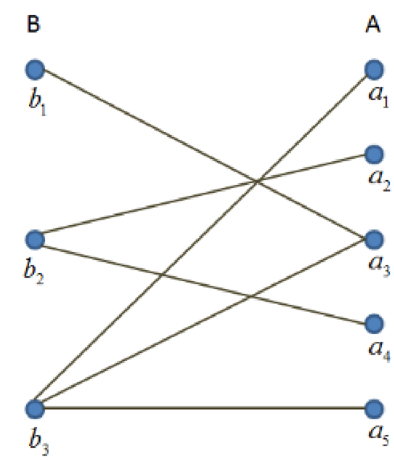
פלט תתי קבוצות $A'\subseteq A, B'\subseteq B$ המקיימים:

- חוקיות: לכל $b_i\in B$ מתקיים $A_i\subseteq A'$.
- אופטימליות: הרווח על A',B' מקסימלי כאשר הרווח מוגדר ע"י $P\left(A',B'\right)=\sum_{b_i\in B'} d_i-\sum_{a_i\in A'} s_i$.

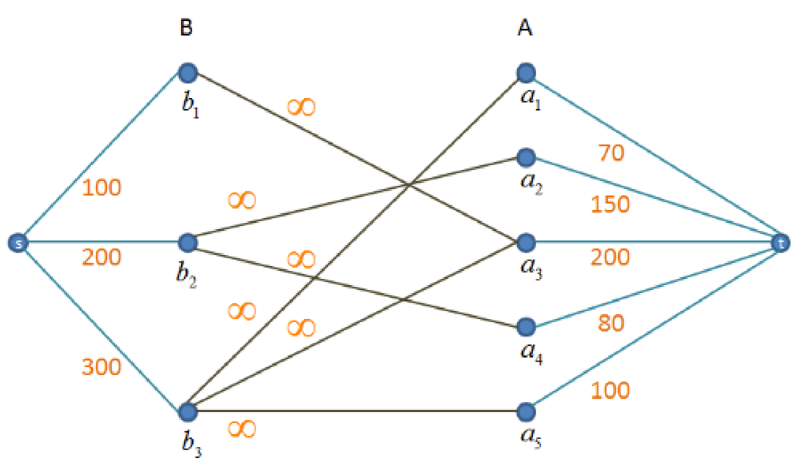
דוגמה

משקיעה	שחקנים אהובים	השקעה
b_1	$A_1=\{a_3\}$	100
b_2	$A_2=\{a_2,a_4\}$	200
b_3	$A_3=\{a_1,a_3,a_5\}$	300

עם שכר השחקנים $(s_1,\dots,s_5)=(50,150,200,80,100)$. נוכל לייצג זאת כגרף דו"צ:



פעם קודמת שראינו גרף דו"צ הוספנו לו צלעות כיווניות והפכנו אותו לרשת זרימה, נעשה את זה גם כאן.



נעשה רידוקציה למציאת חתך מינימלי - כאן החתך הוא ההפסד שלנו. כדי להבטיח חוקיות מבחינת השחקנים של כל משקיעה, נקבע את כל הצלעות המחברות משקיעות לשחקנים להיות אינסוף ולכן אם המשקיעה ב-S אז ממינימליות החתך גם כל השחקנים שלה יהיו ב-S. S היא הבחירה הסופית שלנו, זה אומר שאנחנו מחייבים את בחירת כל השחקנים של משקיעה אם היא נבחרה.

אם $b_i \in S$ אז $c\left(s, b_i\right)$ לא בחישוב החתך ולכן הפסדנו כסף. כלומר בצד שמאל נרצה כמה שיותר משקיעות ב-S, שזה שקול לכמה שפחות צלעות מהאגף השמאלי בחתך, ואילו מימין נרצה כמה שפחות שחקנים ב-S (או יותר מדויק הזולים ביותר), שזה שקול לכמה שפחות צלעות בחתך מהאגף הימני (כי אם לא בחרנו הרבה שחקנים אז הרבה שחקנים יהיו ב-T ואז לא שכרם לא יהיה בחתך).

פסאודו-קוד

1. נגדיר רשת זרימה $N=(V,E,c,s,t)$ ע"י $V=A\cup B\cup\{s,t\}$, $E=E_s\cup\widetilde{E}\cup E_t$ כאשר

$$E_s=\left\{(s,b_i):b_i\in B\right\},\ \widetilde{E}=\left\{(b_i,a_j):b_i\in B,a_j\in A_i\right\}E_t=\left\{(a_i,t):a_i\in A\right\}$$

וגם
$$c\left(u,v\right)=\begin{cases}d_i & u=s,v=b_i\\ s_i & u=a_i,v=t\\ \infty & (u,v)\in\widetilde{E}\end{cases}.$$

2. נריץ את האלג' למציאת חתך מינימלי ונסמנו (S,T).
3. נגדיר $A'=S\cap A, B'=S\cap B$ ונחזיר את A',B'.

משפט האלג' מחזיר פתרון חוקי ואופטימלי.

הוכחה: <u>הכנה</u>: נגדיר התאמה חח"ע ועל מהחתכים ברשת לתתי קבוצות $A'\subseteq A, B'\subseteq B$ (לא בהכרח חוקיות) באופן הבא:

- בהינתן חתך (S,T) נגדיר $A'=S\cap A, B'=S\cap B$.
- בהינתן (A',B') נגדיר $S=A'\cup B'\cup\{s\}$, $T=V\backslash S$.

<u>חוקיות</u>: נראה כי $\forall b_i\in B'$ מתקיים $A_i\subseteq A'$.

טענה עבור חתך (S,T) ו-A',B' המתאימות לחתך מההתאמה הנ"ל מתקיים ש-A',B' חוקיות אם"ם $c\left(S,T\right)$ סופי.

הוכחה: A',B' חוקיות אם"ם $\forall b_i\in B'$ מתקיים $A_i\subseteq A'$ אם"ם $\forall b_i\in B$ מתקיים ש-A_i נמצאת באותו צד של החתך אם"ם אין צלע עם קיבול ∞ בחתך אם"ם $c\left(S,T\right)$ סופי. ■

מסקנה האלג' מחזיר A',B' שמתאימות לחתך מינימלי ברשת. החתך המינימלי הוא מקיבול סופי (כיוון שקיים חתך כלשהו עם קיבול סופי, למשל $S=\{s\}, T=V\backslash S$) ולכן מהטענה הנ"ל A',B' חוקיות.

<u>אופטימליות</u>:

טענה יהי חתך מקיבול סופי (S,T) ותתי קבוצות A',B' שמתאימות לו אזי קיים קבוע D כלשהו כך שמתקיים

$$c(S,T)=D-P(A',B')$$

הוכחה:

$$\begin{aligned}c(S,T)&=\sum_{b_i\notin B'}c(s,b_i)+\sum_{a_i\in A'}(a_i,t)\\&=\sum_{b_i\notin B'}d_i+\sum_{a_i\in A'}s_i\\&=\sum_{b_i\in B}d_i-\sum_{b_i\in B'}d_i+\sum_{a_i\in A}s_i\\&=\underbrace{\sum_{b_i\in B}d_i}_{D}-\underbrace{\left(\sum_{a_i\in A'}s_i-\sum_{b_i\in B'}d_i\right)}_{P(A',B')}\end{aligned}$$

■

מסקנה $P(A',B')$ מקסימלי.

הוכחה: נניח בשלילה ש-$P(A',B')$ לא מקסימלי. לכן קיימים $A''\subseteq A,B''\subseteq B$ חוקיות כך ש-$P(A'',B'')>P(A',B')$. נגדיר (S'',T'') החתך המתאים ל-A'',B''. מהמשפט הנ"ל

$$c(S'',T'')=D-P(A'',B'')<D-P(A',B')=c(S,T)$$

סתירה. ■

■

זמן ריצה

1. בניית הרשת: $\mathcal{O}(|E|+|V|)=\mathcal{O}(nk+(n+k))$.
2. מציאת חתך מינימלי: $\mathcal{O}\left(|E|^2|V|\right)=\mathcal{O}\left(n^2k^2(n+k)\right)$.
3. הגדרת A',B' והחזרתם $\mathcal{O}(|V|)=\mathcal{O}(n+k)$.

לכן סה"כ קיבלנו $\mathcal{O}\left(n^2k^2(n+k)\right)$.

שבוע $\mathbb{XI}$ | הוכחת נכונות של F & F ושל E & K

הרצאה

חלק א' של ההרצאה

הגדרה חתך (S,T) ברשת זרימה מורחבת $N=(V,c,s,t)$ הוא חלוקה זרה לשתי קבוצות כך ש-$s\in S$, $t\in T$. הקיבול של החתך (S,T) הוא $\sum_{x\in S,y\in T} c(x,y)$. עבור f זרימה ברשת, הזרימה בחתך היא $f(S,T)=\sum_{x\in S,y\in T} f(x,y)$.

טענה תהי N רשת זרימה מורחבת, תהי f זרימה ברשת ויהי (S,T) חתך ברשת אזי $f(S,T)=|f|$.

הוכחה:

$$
\begin{aligned}
f(S,T) &= \sum_{x\in S,y\in T} f(x,y)\\
&= \sum_{x\in S}\sum_{y\in T} f(x,y)\\
&= \sum_{x\in S}\left(\sum_{v\in V} f(x,v)-\sum_{u\in S} f(x,u)\right)\\
&= \sum_{x\in S}\sum_{v\in V} f(x,v)-\sum_{x\in S}\sum_{u\in S} f(x,u)
\end{aligned}
$$

$$
\sum_{x\in S}\sum_{v\in V} f(x,v)=\sum_{v\in V} f(s,v)+\underbrace{\sum_{x\in S,x\neq s}\sum_{v\in V} f(x,v)}_{\text{0 משימור חומר}}=|f|+\sum_{x\in S,x\neq s} 0=|f|
$$

$$
\begin{aligned}
\sum_{x\in S}\sum_{u\in S} f(x,u) &= \sum_{x,u\in S} f(x,u)\\
&= \sum_{\{x,u\}\subseteq S} \left(f(x,u)+f(u,x)\right)\\
&= \sum_{\{x,u\}\subseteq S} 0=0
\end{aligned}
$$

כלומר $f(S,T)=|f|+0=|f|$. ■

דוגמה עבור הגרף

מסקנה יהי (S,T) חתך ברשת ותהי g זרימה ברשת. אזי $|g|\le c(S,T)$.

הוכחה: מהטענה הנ"ל,

$$|g| = g(S,T) = \sum_{x\in S, y\in T} g(x,y) \le \sum_{x\in S, y\in T} c(x,y) = c(S,T)$$

■

משפט (Min Cut - Max Flow, השטף והחתך) תהי N רשת זרימה ותהי f זרימה ברשת N. אזי שלושת התנאים הבאים שקולים זה לזה:

1. f היא זרימה אופטימלית ב-N.
2. אין מסילת הרחבה בגרף השיורי G_f.
3. קיים חתך (S,T) כך ש-$|f| = c(S,T)$.

הוכחה:

$(1) \Leftarrow (2)$ נניח בשלילה שקיימת מסילת הרחבה p בגרף השיורי G_f. נגדיר $g = f + \Delta_{f,p}$. ממה שהוכחנו, g זרימה חוקית ברשת N ולכן $|g| = |f| + c_f(p) > |f|$ בסתירה לאופטימליות f.

$(3) \Leftarrow (1)$ תהי g זרימה ברשת N. מהמסקנה הקודמת $|g| \le c(S,T) = |f|$ ולכן f אופטימלית.

$(2) \Leftarrow (3)$ נגדיר $S = \{s\} \cup \{x \in V$ ניתן להגיע מ-s ל-x בגרף השיורי $G_f\}$, $T = V \backslash S$. מכיוון שאין מסילת הרחבה בגרף השיורי, $t \notin S$ ולכן $t \in T$ ולכן זהו חתך חוקי.

יהי $x \in S$, $y \in T$. נוכיח כי $f(x,y) = c(x,y)$. נניח בשלילה כי $f(x,y) \neq c(x,y)$. לכן מאילוץ הקיבול $f(x,y) < c(x,y)$ ולכן $c_f(x,y) > 0$. לכן $(x,y) \in E_f$, כלומר (x,y) היא צלע בגרף השיורי G_f. נבנה מסלול מ-s ל-y בגרף G_f באופן הבא. נעבור על צלעות הגרף G_f מ-s ל-x (ניתן לעשות זאת כי $x \in S$) ונמשיך לצלע (x,y). לכן מהגדרת S, $y \in S$ בסתירה לכך ש-$y \in T$. לכן $f(x,y) = c(x,y)$ ולכן

$$|f| = f(S,T) = \sum_{x\in S, y\in T} f(x,y) = \sum_{x\in S, y\in T} c(x,y) = c(S,T)$$

■

מסקנה אם האלג' של פורד ופולוקרסון עוצר, הוא מחזיר זרימה אופטימלית וזה כי התנאי לעצירתו הוא שאין מסילות הרחבה בגרף השיורי.

מסקנה נניח כי האלג' של פורד ופולקרסון עצר עם זרימה f ונריץ BFS על G_f מקודקוד המקור ברשת וכך נגלה את כל הקודקודים שניתן להגיע אליהם מ-S. נסמן קבוצה זו ב-S. אזי $(S, V\backslash S)$ הוא חתך מינימלי ברשת.

האלגוריתם של Edmonds & Karp למציאת זרימה אופטימלית ברשת

קלט רשת זרימה רגילה $\widetilde{N} = (V, E, \widetilde{c}, s, t)$.

פלט זרימה אופטימלית g ברשת $\widetilde{N}$.

פסאודו-קוד

1. הרחבה: נרחיב את $\widetilde{N}$ לרשת מורחבת N.
2. אתחול: נאתחל $f \equiv 0$.
3. איטרציה: כל עוד קיימת מסילת הרחבה בגרף השיורי G_f, נבחר (באמצעות BFS) p להיות מסילת הרחבה בעלת אורך מינימלי (מספר צלעות מינימלי) ונעדכן $f = f + \Delta_{f,p}$.
4. אם אין מסילות הרחבה בגרף השיורי נעצור.
5. נצמצם את N חזרה ל-$\widetilde{N}$ ואת f לזרימה רגילה g ברשת $\widetilde{N}$.
6. נחזיר את g.

הערה אדמונדס וקארפ זה פשוט פורד ופולקורסון עם בחירת מסלול בעל אורך מינימלי.

משפט האלגוריתם של אדמונדס וקארפ עוצר לאחר $\mathcal{O}\left(|V| \cdot |E|\right)$ איטרציות.

מסקנה זמן הריצה שלו הוא לכל היותר $\mathcal{O}\left(|V| \cdot |E|\right) \mathcal{O}\left(|V| + |E| + |V|\right) = \mathcal{O}\left(|V| \cdot |E|^2\right)$.

נסמן לאחר כל איטרציה את הזרימה הנוכחית f_i. לכן יש לנו $f_0, f_1, \ldots, f_N \equiv 0$. נסמן לאחר כל איטרציה $G_{f_0}, G_{f_1}, \ldots, G_{f_N}$ הגרף השיורי לאחריו.

הגדרות להוכחה פורמלית של אדמונדס וקארפ

1. נסמן ב-f_i את הזרימה ברשת לאחר i איטרציות של האלג' של EK. בפרט $f_0 \equiv 0$.
2. נסמן ב-p_i את מסילת ההרחבה שהאלג' בחר באיטרציה ה-i. לכן $f_i = f_{i-1} + \Delta_{f_{i-1},p_i}$. עבור $x \in V$ נסמן $\delta_i\left(x\right)$ המרחק של x מקודקוד המקור s בגרף השיורי G_{f_i}, כלומר $\delta_i\left(x\right)$ הוא האורך המינימלי של המסלול בין s ל-x ב-G_{f_i}, אם אין מסלול כזה נגדיר $\delta_i\left(x\right) = \infty$.

חלק ב' של ההרצאה

למה 1 לכל $x \in V$ ולכל $i \geq 0$ מתקיים $\delta_{i+1}\left(x\right) \geq \delta_i\left(x\right)$. כלומר:

1. אם $\delta_i\left(x\right) < \infty$ אזי $\delta_{i+1}\left(x\right)$ הוא מספר סופי הגדול או שווה ל-$\delta_i\left(x\right)$ או $\delta_{i+1}\left(x\right) = \infty$.
2. אם $\delta_i\left(x\right) = \infty$ אז גם $\delta_{i+1}\left(x\right) = \infty$.

הערה רעיון ההוכחה למקרה הראשון הוא כזה: נניח בשלילה שקיימים x ו-i כך ש-$\infty > \delta_i\left(x\right) > \delta_{i+1}\left(x\right)$. נבחר קודקוד x עם תכונה זו במרחק מינימלי. יהי y לפניו במסילה באורך מינימלי, לכן y מקיים את התכונה ולכן ניתן להגיע מ-s ל-y ב-G_{f_i} כי הוא ישיג ב-$G_{f_{i+1}}$. אם בשלילה $\left(y,x\right)$ קיימת ב-$G_{f_{i+1}}$ אז $\left(x,y\right)$ היא במסילת ההרחבה (כי $\left(y,x\right)$ לא קיימת ב-G_{f_i} אבל כן ב-$G_{f_{i+1}}$ ולכן זה אומר

שהזרמנו דרך מסילת ההרחבה זרם בכיוון (x,y)). לכן מסילת ההרחבה p_i עד x נותנת מסלול ב-G_{f_i} קצר יותר מזה שיש ב-$G_{f_{i+1}}$ בסתירה להנחה.

המקרה השני זהה לחלוטין רק שבמקום ש-p_i היא מסילה ב-G_{f_i} שקצרה מזו ב-$G_{f_{i+1}}$, הסתירה היא ש-x ישיג ב-G_{f_i}.

הוכחה:

1. נניח בשלילה שקיימים x ו-i כך ש-$\delta_{i+1}(x) < \delta_i(x) < \infty$.יהי x עם מרחק מינימלי מ-s ב-$G_{f_{i+1}}$. נשים לב כי $s \neq x$.

   נתבונן במסילה בעלת אורך מינימלי בין s ל-x ב-$G_{f_{i+1}}$ ונסמן ב-y את הקודקוד הקודם ל-x במסילה זו. מתקיים $\delta_{i+1}(y) = \delta_{i+1}(x) - 1 < \delta_{i+1}(x)$ ולכן y מקיים את טענת הלמה (כי אחרת זו סתירה למינימליות x) כלומר $\delta_i(y) \leq \delta_{i+1}(y)$.

   נוכיח כי $(y,x) \notin G_{f_i}$. נניח בשלילה כי $(y,x) \in G_{f_i}$. לכן

$$\delta_i(x) \leq \delta_i(y) + 1 \leq \delta_{i+1}(y) + 1 = \delta_{i+1}(x)$$

   כלומר $\delta_i(x) \leq \delta_{i+1}(x)$ ולכן x מקיים את טענת הלמה בסתירה לבחירת x.

   לכן $(y,x) \notin G_{f_i}$ ולכן $c_{f_i}(y,x) = 0$ ולכן $f_i(y,x) = c(y,x)$.

   מנגד, $(y,x) \in G_{f_{i+1}}$ (כך y הוגדר) ולכן $c_{f_{i+1}}(y,x) > 0$ ולכן $f_{i+1}(y,x) < c(y,x)$ ולכן $f_{i+1}(y,x) < f_i(y,x)$.

   מדרך פעולות של האלג', זה שקול לכך ש-$\Delta_{f_i,p_i}(y,x) < 0$, כלומר ש-$\Delta_{f_i,p_i}(x,y) > 0$, ולכן $(x,y) \in p_i$.

   מהגדרת p_i, הוא מסלול באורך מינימלי מ-s ל-t ב-G_{f_i} ולכן מתקיים כי $\delta_i(y) = \delta_i(x) + 1$ (y בא אחרי x במסלול קצר ביותר) ולכן

$$\delta_i(x) = \delta_i(y) - 1 \leq \delta_{i+1}(y) - 1 = \delta_{i+1}(x) - 2 < \delta_i(x)$$

   סתירה.

2. בשל הדמיון למקרה הראשון, נסתפק בהוכחה שלו ולרעיון ההוכחה של המקרה השני לעיל.

■

להמשך הוכחת הנכונות

תרגול

הגדרה בהינתן בעיית תכנות לינארי בצורה הסטנדרטית $\begin{matrix}\max c^T \cdot x \\ s.t\ Ax \leq b, x \geq 0\end{matrix}$, <u>הבעיה הדואלית</u> לה מוגדרת ע"י $\begin{matrix}\min b^T \cdot y \\ s.t\ A^T y \geq c, y \geq 0\end{matrix}$

הערה הבעיה הדואלית של הבעיה הדואלית נקראת הבעיה הפרימלית.

טענה (דואליות חלשה) לכל פתרון חוקי $x \in \mathbb{R}^n$ (שמקיים את אילוצי הבעיה הפרימלית) ולכל פתרון חוקי $y \in \mathbb{R}^m$ (שמקיים את אילוצי הבעיה הדואלית) מתקיים

$$c^T x \leq b^T y$$

הערה כלומר המינימום בבעיה הדואלית $\leq$ מקסימום בבעיה הפרימלית.

הוכחה:

$$c^T x = \sum_{i=1}^{n} c_i x_i \leq \sum_{i=1}^{n} \left(A^T y\right)_i x_i \overset{\text{לינארית}}{=} y^T A x \leq y^T b = b^T y$$

■

טענה (דואליות חזקה בבעיות תכנון לינארי) הפתרון שממקסם את הבעיה הפרימלית x^* שווה לפתרון שממזער את הבעיה הדואלית y^*, כלומר $b^T y^* = c^T x^*$.

נשים לב כי אם הבעיה היא בעיית הזרימה, אז הדואליות החזקה היא בדיוק Min Cut Max Flow.

בעיית הזרימה כבעיית תכנון לינארי

תהי $N = (V, E, c', s, t)$ רשת זרימה. נסדר את הצלעות באופן שרירותי $E = \left\{e_1, \ldots, e_{|E|}\right\}$.

- נגדיר $x = f = \begin{pmatrix} f(e_1) \\ \vdots \\ f\left(e_{|E|}\right) \end{pmatrix}$.
- נגדיר c באופן הבא: $c\left(e\right) = \begin{cases} 1 & e = (s, u) \\ 0 & \text{otherwise} \end{cases}$ (זה וקטור עמודה בגודל $|E|$, סימונים מבלבלים). נקבל

$$c^T f = \sum_{(s,u) \in E} 1 \cdot f\left(s, u\right) = |f|$$

- האילוץ הראשון הוא $f \geq 0$ והוא ברירת מחדל לכל בעיית תכנון לינארי.

את אילוץ הקיבול נשיג באמצעות $f\left(e\right) \leq c'\left(e\right)$ $\left(0, \ldots, \underset{e}{1}, 0, \ldots, 0\right)$.

בחוק שימור החומר נדרוש $\sum\limits_{(v,u)\in E} f\left(v,u\right)=0 \sum\limits_{(u,v)\in E} f\left(u,v\right)-$, $\forall v\in V\setminus\{s,t\}$ אין לנו שוויון באילוצי תכנון לינארי ולכן נפצל זאת לשני אי שוויונות לינאריים שאנחנו יודעים לכתוב:

$$\sum_{(u,v)\in E} f\left(u,v\right)-\sum_{(v,u)\in E} f\left(v,u\right)\leq 0,\quad -\sum_{(u,v)\in E} f\left(u,v\right)+\sum_{(v,u)\in E} f\left(v,u\right)\leq 0$$

נשים לב כי אם נסיר את השיקול השני, כלומר נישאר רק עם $\sum\limits_{(v,u)\in E} f\left(v,u\right) \leq \sum\limits_{(u,v)\in E} f\left(u,v\right)$, נקבל בעיה דומה לבעיית הזרימה שהשינוי היחיד בה הוא שכל קודקוד משמש גם כמקור ומלבד הזרימה שהוא מקבל שהוא כן מחויב להוציא, הוא יכול להוסיף עוד זרימה נוספת אם ירצה. נשים לב שהגדרת השטף בבעיה המקורית וגם בוריאציה הזו זהה, כי מספיקה לנו ההבטחה שכל זרם שנוציא מ-s יגיע ל-t, גם אם יגיע יותר מדי.

בבעייה שכזו, $A=\begin{pmatrix} I \\ \hline M\end{pmatrix}$, כאשר I מטריצת היחידה בגודל $|E|$ ואילו $M\in M_{(|V|-2)\times|E|}\left(\mathbb{R}\right)$ מוגדרת ע"י

$$M_{v,e}=\begin{cases}1 & \exists u\in V:e=(u,v)\\ -1 & \exists u\in V:e=(v,u)\\ 0 & \text{otherwise}\end{cases}$$

נמצא את הבעיה הדואלית לזו. $y=\begin{pmatrix} y_{e_1}\\ \vdots \\ y_{e_{|E|}}\\ z_{v_1}\\ \vdots\\ z_{v_{|V|}}\end{pmatrix}$, נמזער את $b^Ty=\sum\limits_{e\in E} c'\left(e\right)y_e$ (שזה ממש במקרה קיבול החתך), כלומר אנחנו מתעלמים מ-$z_{v_1},\ldots,z_{v_{|V|}}$ בחישוב הזה והם ישמשו רק לשם סיפוק האילוצים. $A^T=\left(\begin{array}{c|c} I & M^T\end{array}\right)$.

נסתכל על M. כל שורה בה מייצגת קודקוד v שאינו מקור או בור, בה קוורדינטת צלע מקבלת 1 אם היא מהצורה (u,v), -1 אם היא מהצורה (v,u) ו-0 אחרת.

כל עמודה מייצגת צלע $e=(u,v)$, בה קוורדינטת קודקוד מקבלת ערך 1 אם הוא u, -1 אם הוא v, ו-0 אחרת. כל זאת עם הסייג שאם הצלע מתחילה מהמקור או מסתיימת בבור אז למעשה בעמודה יהיה לנו רק ערך אחד (1 או -1 בהתאמה).

כלומר הדרישות הן, לכל $u\neq s,v\neq t$ כך ש-$(u,v)\in E$, שיתקיים $y_{(u,v)}+z_v-z_u\geq 0$. בנוסף, $\forall\left(u,t\right)\in E$ שיתקיים $y_{(u,t)}-z_u\geq 0$ וכן $\forall\left(s,v\right)\in E$, $y_{(s,v)}+z_v\geq 1$.

לאינטואיציה אפשר לחשוב על z_v כאינדיקטור האם v ב-S או ב-T. בנוסף התנאי הראשון דואג שאם $y_{(u,v)}=1$ כלומר (u,v) בחתך שנובע מ-y אז $u\in S$ ו-$v\in T$ והשאר עושים את אותו הדבר במקרים המיוחדים של המקור והבור.

שבוע $\mathbb{XIII}$ | התמרת פורייה דיסרקטית ומהירה

הרצאה

חלק א' של ההרצאה

הגדרה נסמן ב-V_{n-1} את המרחב הוקטורי של הפולינומים ממעלה קטנה או שווה ל-$n-1$ מעל הממשיים.

ייצוג של פולינום לפי מקדמים

הגדרה ייצוג המקדמים של הפולינום $p(x)=\sum\limits_{k=0}^{n-1}a_kx^k$ הוא וקטור $\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}\in\mathbb{R}^n$. נשים לב כי ייצוג המקדמים מגדיר איזומורפיזם בין $\mathbb{R}^n$ ו-V_{n-1}.

הגדרה נגדיר פעולות על V_{n-1} באופן ריגורוזי מדי:

1. חיבור:

   קלט שני פולינומים $P,Q\in V_{n-1}$ בייצוג המקדמים.

   פלט הפולינום $R=P+Q\in V_{n-1}$ בייצוג המקדמים.

   אלגוריתם יהיו $\begin{pmatrix}b_0\\ \vdots\\ b_{n-1}\end{pmatrix},\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}$ וקטורי ייצוג המקדמים של P ו-Q בהתאמה. אז וקטור המקדמים של R הוא $\begin{pmatrix}a_0+b_0\\ \vdots\\ a_{n-1}+b_{n-1}\end{pmatrix}$.

   זמן ריצה $\Theta(n)$.

2. הצבת ערך:

   קלט פולינום $P\in V_{n-1}$ ו-$x_0\in\mathbb{R}$.

   פלט $P(x_0)$.

   אלגוריתם יהי $\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}$ וקטור המקדמים של P. צריך לחשב את $P(x_0)=\sum\limits_{k=0}^{n-1}a_kx_0^k$. נחשב את סדרת החזקות של x_0: $1,x_0,x_0^2,\dots,x_0^{n-1}$ באמצעות $x_0^{k+1}=x_0\cdot x_0^k$ ומשם נחשב את $P(x_0)$.

   זמן ריצה $\Theta(n)$.

3. כפל פולינומים:

   קלט שני פולינומים $P,Q\in V_{n-1}$ בייצוג המקדמים.

   פלט הפולינום $R=PQ\in V_{2n-2}$ בייצוג המקדמים.

   אלגוריתם יהיו $\begin{pmatrix}b_0\\ \vdots\\ b_{n-1}\end{pmatrix},\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}$ וקטורי המקדמים של P ו-Q בהתאמה. יהי $\begin{pmatrix}c_0\\ \vdots\\ c_{n-1}\end{pmatrix}$ וקטור המקדמים של R. מתקיים לכל $0\le m\le 2n-2$, $c_m=a_0b_m+a_1b_{m-1}+\cdots+a_mb_0$ (צריך שהחזקה תהיה m ואלו כל הקומבניציות של

מקדמים שמאפשרים זאת). אם a_k, b_k לא קיימים, נחשבם כאפסים. יש לנו $2n-2$ חישובים, כל אחד לכל היותר n פעולות בסיסיות כלומר $\Theta\left(n^2\right)$.

הערה לפי ייצוג מקדמים, פעולת הכפל היא $\Theta\left(n^2\right)$ שזה יקר מדי. נחפש שייצוג אחר יותר זול.

ייצוג של פולינומים לפי ערכים

טענה לכל פולינום ממעלה $\leq n-1$ מעל $\mathbb{F}$ השונה מאפס יש לכל היותר $n-1$ שורשים שונים.

מסקנה פולינום ממעלה $\leq n-1$ נקבע באופן יחיד ע"י הערכים שלו ב-n נקודות שונות.

הוכחה: תהיינה $x_0, \ldots, x_{n-1}$ n נקודות שונות ב-$\mathbb{F}$. נניח בשלילה שקיימים $P \neq Q$ כך ש-$P\left(x_i\right) = Q\left(x_i\right)$. אזי הפולינום $R = P - Q$ הוא פולינום ממעלה לכל היותר $n-1$ המתאפס ב-n נקודות ולכן מהטענה הנ"ל, $R = 0$ ולכן $P = Q$ סתירה. ■

נקבע n נקודות ממשיות שונות $x_0, \ldots, x_{n-1}$.

הגדרה ייצוג הערכים של פולינום $P \in V_{n-1}$ הוא וקטור הערכים $\begin{pmatrix} P(x_0) \\ \vdots \\ P(x_{n-1}) \end{pmatrix}$. מהמסקנה הנ"ל זה ייצוג נאמן (איזומורפיזם) של V_{n-1}.

פעולות בייצוג ערכים

1. חיבור:

קלט שני הפולינומים $P, Q \in V_{n-1}$ בייצוג הערכים.

פלט הפולינום $R = P + Q$ בייצוג הערכים.

אלגוריתם נחשב

$$\begin{pmatrix} R(x_0) \\ \vdots \\ R(x_{n-1}) \end{pmatrix} = \begin{pmatrix} (P+Q)(x_0) \\ \vdots \\ (P+Q)(x_{n-1}) \end{pmatrix} = \begin{pmatrix} P(x_0) \\ \vdots \\ P(x_{n-1}) \end{pmatrix} + \begin{pmatrix} Q(x_0) \\ \vdots \\ Q(x_{n-1}) \end{pmatrix}$$

זמן ריצה $\Theta\left(n\right)$.

2. כפל:

קלט שני פולינומים $P, Q \in V_{n-1}$ בייצוג הערכים.

פלט הפולינום $V_{2n-2} \ni R = PQ$ בייצוג הערכים.

אלגוריתם וקטור הערכים של R הוא $\begin{pmatrix} R(x_0) \\ \vdots \\ R(x_{n-1}) \end{pmatrix} = \begin{pmatrix} (PQ)(x_0) \\ \vdots \\ (PQ)(x_{n-1}) \end{pmatrix} = \begin{pmatrix} P(x_0)Q(x_0) \\ \vdots \\ P(x_{n-1})Q(x_0) \end{pmatrix}$. נשים לב כי זה לא מספיק כי R במעלה (אולי) גבוהה מ-$n-1$ ולכן נדרוש את ערכי P ו-Q ב-$2n-1$ נקודות שונות $x_0, \ldots, x_{2n-2}$ ואז נריץ את האלג' כמו שנתון לעיל.

זמן ריצה $\Theta\left(n\right)$ בהנחה שמציאת הנקודות הנוספות נעשה בזמן לינארי (הנחה נכונה כפי שנראה).

3. הצבת ערך (בעיית האינטרפולציה הפולינומית):

קלט פולינום $P\in V_{n-1}$ בייצוג הערכים (בנקודות $x_0,\ldots,x_{n-1}$) ונקודה נוספת x_n.

פלט $P\left(x_n\right)$.

האלגוריתם נציג פתרון (של לגרנז') שרץ ב-$\Theta\left(n^2\right)$, אפע"פ שידוע פתרון שרץ ב-$\Theta\left(n\log n\right)$. נגדיר n פולינומים $L_0,\ldots,L_{n-1}\in V_{n-1}$ עם התכונה $L_m\left(x_k\right)=\delta_{km}=\begin{cases}1 & k=m\\ 0 & k\neq m\end{cases}$. באופן מפורש,

$$L_m\left(x\right)=\frac{\left(x-x_0\right)\ldots\left(x-x_{m-1}\right)\left(x-x_{m+1}\right)\ldots\left(x-x_{n-1}\right)}{\left(x_m-x_0\right)\ldots\left(x_m-x_{m-1}\right)\left(x_m-x_{m+1}\right)\ldots\left(x_m-x_{n-1}\right)}$$

($x_m\neq x_i$, $i\neq m$). הרעיון הוא להשתמש בפוקנציונאל הלינארי φ_i ובעזרתו להגיד איזומורפיזם של בין וקטור מספרים לפולינום (ליאנרית בקיצור).

למה לכל $P\in V_{n-1}$, $P=\sum\limits_{m=0}^{n-1}P\left(x_m\right)L_m$.

הוכחה: (העשרה) נסמן $\tilde{P}=\sum\limits_{m=0}^{n-1}P\left(x_m\right)L_m$. $\tilde{P}$ הוא פולינום ממעלה $\leq n-1$. נוכיח כי $P=\tilde{P}$. לכל $0\leq k\leq n-1$ מתקיים $P\left(x_k\right)=\tilde{P}\left(x_k\right)$ ולכן $P=\tilde{P}$ מהמסקנה בתחילת ההרצאה. אכן ניתן להציב את $x_0,\ldots,x_{n-1}$ ולראות שמתקבל שוויון. ■

מסקנה $P\left(x_n\right)=\sum\limits_{m=0}^{n-1}P\left(x_m\right)L_m\left(x_n\right)$.

זמן ריצה החישוב של $L_m\left(x_n\right)$ עולה $\Theta\left(n\right)$ לכל $0\leq m\leq n-1$ וסה"כ נקבל $\Theta\left(n^2\right)$.

מסקנה בייצוג הערכים פעולות חיבור וכפל פולינומים הן יעילות (זמן לינארי) ופעולת הצבת הערך יקרה ב-$\Theta\left(n^2\right)$. פעולת כפל קשה בייצוג מקדמים ופעולת הצבת הערך קשה בייצוג הערכים.

רעיון נוכל לעבור בין הייצוגים ולבצע כל פעולה בייצוג בו הפעולה יעילה. כמה עולה המעבר בין הייצוגים?

שני הייצוגים של הפולינומים הם המקדמים של פיתוח של פולינום לפי שני בסיסים שונים - בסיס החזקות $\left\{1,\ldots,x^{n-1}\right\}$ ובייצוג המקדמים ובסיסי פולינומי לגראנז' $\left\{l_0,\ldots l_{m-1}\right\}$ בייצוג הערכים. כדי לעבור בין הייצוגים צריך לכפול במטורצית המעבר בין הבסיסים.

למה יהי $P\in V_{n-1}$. יהי $\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}$ ייצוג המקדמים של P ויהי $\begin{pmatrix}P(x_0)\\ \vdots\\ P(x_{n-1})\end{pmatrix}$ ייצוג הערכים של P ב-n נקודות שונות $x_0,\ldots,x_{n-1}$ אזי $M\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}=\begin{pmatrix}P(x_0)\\ \vdots\\ P(x_{n-1})\end{pmatrix}$ כאשר $M=\begin{pmatrix}1 & x_0 & \cdots & x_0^{n-1}\\ \vdots & & & \\ 1 & x_{n-1} & \cdots & x_{n-1}^{n-1}\end{pmatrix}$ (מטריצת ון דר מונדה).

הוכחה: נבדוק כי שני הוקטורים בטענת הלמה שווים בכל הקוואורדינטות.

$$\left(M\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}\right)_k=\sum_{m=0}^{n-1}M_{km}a_m=\sum_{m=0}^{n-1}x_k^m a_m=P\left(x_k\right)$$

■

מסקנה כדי לעבור בין הייצוגים, צריך לכפול את וקטור המקדמים במטריצת ון דר מונדה כדי להגיע לייצוג הערכים וזה $\Theta\left(n^2\right)$ ובכיוון ההפוך צריך לחשב את ההופכית ולכפול בה וזה גם $\Theta\left(n^2\right)$.

רעיון שימוש במספרים מרוכבים מכיוון ש-$\mathbb{R}$ הוא תת שדה של שדה המספרים המרוכבים $\mathbb{C}$. ניתן לחשוב את $P\in V_{n-1}$ מעל $\mathbb{C}$ ובפרט להציב בו מספרים מרוכבים. גם מעל $\mathbb{C}$ יש לכל היותר $n-1$ שורשים לכל פולינום ממעלה $n-1\geq$ ולכן ייצוג הערכים מעל $\mathbb{C}$ מוגדר היטב.

מספר מרוכב $z\in\mathbb{C}$ ניתן לכתיבה כ-$z=a+bi$ כאשר $a,b\in\mathbb{R}$. ייצוג אחר של z הוא הייצוג הפולרי, $z=R\left(\cos\theta+i\sin\theta\right)$. כאשר $R=\sqrt{a^2+b^2}=|z|$ ו-θ הזווית של הוקטור המחבר את z עם ראשית הצירים.

$$R_1\left(\cos\theta+i\sin\theta\right)\cdot R_2\left(\cos\varphi+i\sin\varphi\right)=R_1R_2\left(\cos\left(\theta+\varphi\right)+i\sin\left(\theta+\varphi\right)\right)$$

הגדרה <u>מעגל היחידה</u> הוא $\{z\in\mathbb{C}:|z|=1\}$ כלומר כל מרוכבים עם $R=1$, $\{\cos\theta+i\sin\theta:i\in[0,2\pi)\}$.

דוגמה $1=\cos0+i\sin0$, $i=\sin\frac{\pi}{2}+i\sin\frac{\pi}{2}$, $-i=\cos\frac{3\pi}{2}+i\sin\frac{3\pi}{2}$.

אנחנו נדבר רק על שורשי היחידה מסדר n, כלומר מספרים z כך ש-$z^n=1$.

חלק א' של ההרצאה

מסקנה (מהגדרת מעגל היחידה) יהי $z=\cos\theta+i\sin\theta$. אז $z^k=\cos k\theta+i\sin k\theta$.

הגדרה <u>שורשי היחידה</u> מסדר n הם $\{z:z^n=1\}$.

נסמן $\omega_n=\cos\frac{2\pi}{n}+i\sin\frac{2\pi}{n}$. מתקיים

$$\omega_n^n=\cos n\frac{2\pi}{n}+i\sin n\frac{2\pi}{n}=\cos2\pi+i\sin2\pi=1$$

עבור $0\leq k\leq n-1$, $\omega_n^k=\cos\frac{2\pi k}{n}+i\sin\frac{2\pi k}{n}$. אלה n נקודות שונות על מעגל היחידה והן כולן שורשים מסדר n כי

$$\left(\omega_n^k\right)^n=\omega_n^{kn}=1^k=1$$

הגדרה ω_n נקרא <u>שורש יחידה פרימיטיבי</u>.

דוגמה עבור $n=2$, שורשי היחידה הם $\{-1,1\}$, עבור $n=4$, $\omega=i$ ולכן שורשי היחידה הם $\{1,i,-1,-i\}$.

הגדרה יהי $\begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \end{pmatrix} \in \mathbb{C}^n$. יהי $P\in V_{n-1}=\sum_{k=0}^{n-1} a_k z^k$. אזי <u>התמרת פורייה בדידה</u> מסדר n היא $\begin{pmatrix} P(\omega_n^0) \\ \vdots \\ P(\omega_n^{n-1}) \end{pmatrix} = \begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \end{pmatrix} \mathrm{DFT}_n$.

הערה התמרת פורייה DFT_n הינה מעבר מייצוג המקדמים של פולינום לייצוג הערכים שלו בשורשי היחידה מסדר n. ראינו כי בהינתן $P=\sum_{k=0}^{n-1} a_k z^k \in V_{n-1}$ מתקיים שייצוג הערכים של P בנקודות $z_0,\dots,z_{n-1}$ ניתן לכתיבה כ-

$$\begin{pmatrix} P(z_0) \\ \vdots \\ P(z_{n-1}) \end{pmatrix} = M \begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \end{pmatrix}, \quad M=\begin{pmatrix} 1 & z_0 & \cdots & z_0^{n-1} \\ \vdots \\ 1 & z_{n-1} & \cdots & z_{n-1}^{n-1} \end{pmatrix}$$

אבל אצלנו מתקיים $z_k=\omega_n^k$ ולכן $M=\begin{pmatrix} \omega_n^{0\cdot 0} & \omega_n^{1\cdot 0} & \cdots & \omega_n^{(n-1)0} \\ \vdots \\ \omega_n^{0\cdot(n-1)} & \omega_n^{1\cdot(n-1)} & \cdots & \omega_n^{(n-1)(n-1)} \end{pmatrix}$ כלומר $[M]_{ij}=\omega_n^{ij}$

דוגמה עבור $n=4$, $M=\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & -1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$ ($\omega_4=i$).

הגדרה יהי $\begin{pmatrix} p_0 \\ \vdots \\ p_{n-1} \end{pmatrix} \in \mathbb{C}^n$. יהי $P\in V_{n-1}=\sum_{k=0}^{n-1} a_k z^k$ כך ש-$\begin{pmatrix} P(\omega_n^0) \\ \vdots \\ P(\omega_n^{n-1}) \end{pmatrix} = \begin{pmatrix} p_0 \\ \vdots \\ p_{n-1} \end{pmatrix}$. <u>התמרת פורייה בדידה הפוכה</u> מסדר n היא $\begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} p_0 \\ \vdots \\ p_{n-1} \end{pmatrix}\mathrm{DFT}_n^{-1}$, כלומר במפורש

$$\mathrm{DFT}_n^{-1}\begin{pmatrix} p_0 \\ \vdots \\ p_{n-1} \end{pmatrix} = M^{-1}\begin{pmatrix} p_0 \\ \vdots \\ p_{n-1} \end{pmatrix}$$

הערה התמרת פורייה הפוכה היא מעבר בין ייצוג הערכים לייצוג המקדמים.

למה $M^{-1}=\frac{1}{n}\begin{pmatrix} & \vdots & \\ \cdots & \omega_n^{-km} & \cdots \\ & \vdots & \end{pmatrix}$.

משפט (משפט ה-FFT) יהי n חזקה של 2.

1. לכל $\begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \end{pmatrix} \in \mathbb{C}^{n-1}$ ניתן לחשב את $\mathrm{DFT}_n\begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \end{pmatrix}$ בזמן $\mathcal{O}(n\log n)$.
2. לכל $\begin{pmatrix} p_0 \\ \vdots \\ p_{n-1} \end{pmatrix} \in \mathbb{C}^{n-1}$ ניתן לחשב את $\mathrm{DFT}_n^{-1}\begin{pmatrix} p_0 \\ \vdots \\ p_{n-1} \end{pmatrix}$ בזמן $\mathcal{O}(n\log n)$.

תרגול

משפט (משפט האינטרפולציה של לגראנז') לכל n נקודות $(x_0,y_0),\dots,(x_{n-1},y_{n-1})$ כך ש-$i\neq j \Rightarrow x_i\neq x_j$ קיים פולינום יחיד מדרגה $\leq n-1$ כך ש-$P(x_i)=y_i$.

כפל פולינומים מהיר

קלט $a=(a_0,\dots,a_{n-1})^T, b=(b_0,\dots,b_{n-1})^T$ ייצוג המקדמים של P,Q.

פלט $c=(c_0,\dots,c_{n-1})^T$ ייצוג המקדמים של PQ.

אלגוריתם

1. נסמן את $m=\min\{k:2^k\geq 2n-2\}$ ונרפד באפסים $a=(a_0,\dots,a_{n-1},0,\dots,0)^T\in\mathbb{R}^m$ וכך גם עבור b. $\Theta(n)$
2. נחשב $a'=\mathrm{FFT}_n(a), b'=\mathrm{FFT}_n(b)$ (כאשר FFT_n הוא אלג' לחישוב DFT_n שנראה בהרצאה). $\Theta(n\log n)$
3. נחשב $r'=a'\cdot b'$ (איבר איבר). $\Theta(n)$
4. נחשב את $R=\mathrm{FFT}_n^{-1}(r')$ ונחזיר את R. $\Theta(n\log n)$.

זמן ריצה $\Theta(n\log n)$.

הערה האלגוריתם היה נכון גם אילו הפולינומים היו בדרגות שונות, כי היינו מרפדים עד לקבלת דרגה שווה.

הגדרה יהיו $b=\begin{pmatrix}b_0\\ \vdots\\ b_{n-1}\end{pmatrix}, a=\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}$. הקונבולוציה שלהם היא $c=a*b=\begin{pmatrix}c_0\\ \vdots\\ c_{n-1}\end{pmatrix}$ כאשר $c_i=\sum_{j=0}^{i}a_jb_{i-j}$.

דוגמה $b=(1,1,6,2)^T, a=(5,7,3,2,1,5,0,7)^T$. $c=a*b$. $c_0=a_0b_0$, $c_6=a_00+a_10+a_20+a_3b_3+a_4b_2+a_5b_1+a_6b_0$.

במובן הזה, אפשר לחשוב על הקונבולוציה כעל ריצה מלמעלה בוקטור אחד ומלמטה בוקטור האחר, אבל יש עוד דרך.

אם נחשוב על שני הוקטורים כפסים רצים מרופדים באפסים, אז האיבר ה-i בקונבולוציה היא הזזה של i צעדים את אחד הוקטורים אל עבר האחר וחישוב המכ"פ שלהם.

הערה חישוב קונבולוציה שקול לחישוב המקדמים של פולינום המכפלה הנוצר מהסתכלות על וקטורי הקלט כפולינום בייצוג מקדמים.

בעיית התאמת המחרוזות

קלט $S=(s_0,\dots,s_{n-1})\in\{\pm1\}^n$, $P=(p_0,\dots,p_{m-1})\in\{\pm1\}^m$. עם $m\leq n$.

פלט d אוסף כל האינדקסים כך ש-$k\in D$ אמ"ם קיימת הופעה רציפה של p החלק מהאינדקס k ב-S, כלומר

$$D=\{k\in\{0,\dots,n-m-1\}:p_i=s_{k+i},\forall i\in\{0,m-1\}\}$$

הערה האלגוריתם הנאיבי של מעבר על כל האינדקסים דורש $\mathcal{O}(nm)$ שזה מיותר, כי אפשר לעשות יותר טוב עם קונבולוציה.

נשים לב כי $p_is_{k+i}=1\iff p_i=s_{k+i}$ ולכן $\sum p_is_{k+i}=m$ אמ"ם יש מופע של p ב-s החל מ-k.

אלגוריתם

1. נבנה p^R ע"י היפוך התווים של P, כלומר $p_i^R = p_{m-i-1}$. $\mathcal{O}(m)$
2. נחשב $c = p^R * s$. $\mathcal{O}(n \log n)$
3. נחזיר $D = \{k-(m-1) : c_k = m\}$. $\mathcal{O}(m)$

זמן ריצה $\mathcal{O}(n \log n)$ שזה עדיף על האלג' הנאיבי אלא אם $m = \mathcal{O}(\log n)$.

הערה הסיבה לחיסור האינדקס היא משום שבחישוב הקונבולוציה אנחנו מתחילים m צעדים קדימה.

למה $c_k = m$ אם"ם קיים מופע של P ב-S שמסתיים באינדקס k אם"ם קיים מופע רציף של P ב-S החל מ-$k-(m-1)$.

הוכחה: נזכור כי $p_i = s_{k+i}$ אם"ם $p_i s_{k+i} = 1$ ולכן קיים מופע רציף של P ב-S שמתחיל ב-k' אם"ם $\sum_{i=0}^{m-1} p_i s_{k'+i} = m$.

$$\begin{aligned} c_k &= \sum_{j=0}^{k} p_j^R s_{k-j} \\ &= \sum_{j=0}^{k} p_{m-1-j} s_{k-j} \\ &\overset{i=m-1-j}{=} \sum_{i=m-1-k}^{m-1} p_i s_{k-(m-1-i)} \end{aligned}$$

אם $k < m-1$ אז $i > 0$ תמיד ולכן סכום לעולם לא יהיה שווה ל-m (פחות מ-m נסכמים שהם ± 1). אם $m-1 \leq k$ אז $i \leq 0$ ואז

$$\begin{aligned} c_k &= \underbrace{\sum_{i=m-1-k}^{0} p_i s_{k-(m-1)+i}}_{\text{0 מריפוד באפסים}} + \sum_{i=0}^{m-1} p_i s_{k-(m-1)+i} \\ &= \sum_{i=0}^{m-1} p_i s_{k-(m-1)+i} \end{aligned}$$

ומההבחנה בהתחלה $c_k = m$ אם"ם מתקיים התנאי בלמה. ■

שבוע $\mathbb{XIII}$ | משפט ה-FFT, קריפטוגרפיה ואלגוריתמים על מספרים

הרצאה

חלק א' של ההרצאה

הערה הסדטונטית המשקיעה תוכיח כי אם $\mathrm{DFT}_n\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}=\begin{pmatrix}p_0\\ \vdots\\ p_{n-1}\end{pmatrix}$ ונסמן $Q=\sum\limits_{k=0}^{n-1}p_kz^k$ אז

$$\mathrm{DFT}_n^{-1}\begin{pmatrix}p_0\\ \vdots\\ p_{n-1}\end{pmatrix}=\frac{1}{n}\begin{pmatrix}Q\left(\omega_n^{-0}\right)\\ \vdots\\ Q\left(\omega_n^{-(n-1)}\right)\end{pmatrix}$$

משפט (FFT) יהי n חזקה של 2. לכל $\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}\in\mathbb{C}^n$ ניתן לחשב את $\mathrm{DFT}_n\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}$ בזמן $\mathcal{O}\left(n\log n\right)$.

למה 1 יהי n מספר זוגי ו-$P\left(z\right)=\sum\limits_{k=0}^{n-1}a_kz^k$. נגדיר $P_0\left(y\right)=\sum\limits_{j=0}^{\frac{n}{2}-1}a_{2j}y^j$, $P_1\left(y\right)=\sum\limits_{l=0}^{\frac{n}{2}-1}a_{2l+1}y^l$ אזי $P\left(z\right)=P_0\left(z^2\right)+zP_1\left(z^2\right)$.

דוגמה $P\left(z\right)=z^3-7z^2-3z+1$.

$$P\left(z\right)=\left(-7z^2+1\right)+\left(z^3-3z\right)=\left(-7z^2\right)+z\left(z^2-3\right)=P_0\left(z^2\right)+zP_1\left(z^2\right)$$

הוכחה:

$$P_0\left(z^2\right)+zP_1\left(z^2\right)=\sum_{j=0}^{\frac{n}{2}-1}a_{2j}z^{2j}+z\sum_{l=0}^{\frac{n}{2}-1}a_{2l+1}z^{2l}=\sum_{k=0}^{n-1}a_kz^k=P\left(z\right)$$

■

למה 2 יהי n מספר זוגי. לכל $0\le j\le\frac{n}{2}-1$ מתקיים $\left(\omega_n^{\frac{n}{2}+j}\right)^2=\left(\omega_n^j\right)^2=\omega_{\frac{n}{2}}^j$, כלומר כשמעלים בריבוע את שורשי היחידה מסדר n מקבלים (פעמיים) את שורשי היחידה מסדר $\frac{n}{2}$.

דוגמה $n=4$, $\omega_n=i$.

$$\begin{aligned}&1,i,-1,-i\\&1^2,i^2,\left(-1\right)^2,\left(-i\right)^2\\=&1,-1,1,-1\end{aligned}$$

כלומר פעמיים שורשי היחידה מסדר 2.

הוכחה: $\omega_n=\cos\frac{2\pi}{n}+i\sin\frac{2\pi}{n}$ וגם $\omega_{\frac{n}{2}}=\cos\frac{4\pi}{n}+i\sin\frac{4\pi}{n}$. נשים לב כי $\omega_n^2=\omega_{\frac{n}{2}}$ (אפשר להציב פשוט) ולכן

$$\left(\omega_n^j\right)^2=\omega_n^{2j}=\left(\omega_n^2\right)^j=\omega_{\frac{n}{2}}^j$$

ובאותו האופן

$$\left(\omega_n^{\frac{n}{2}+j}\right)^2 = \omega_n^{n+2j} = \omega_n^n\omega_n^{2j} = \left(\omega_n^2\right)^j = \omega_{\frac{n}{2}}^j$$

■

הוכחה: (של משפט ה-FFT) מלמה 1, ניתן לרשום את $P\left(z\right)=P_0\left(z^2\right)+zP_1\left(z^2\right)$ ולכן

$$\begin{aligned}
\mathrm{DFT}_n\begin{pmatrix} a_0 \\ \vdots \\ a_{n-1}\end{pmatrix} &= \begin{pmatrix} P\left(\omega_n^0\right) \\ \vdots \\ P\left(\omega_n^{n-1}\right)\end{pmatrix} \\
&= \left(\begin{array}{c} P_0\left(\left(\omega_n^0\right)^2\right)+\omega_n^0 P_1\left(\left(\omega_n^0\right)^2\right) \\ \vdots \\ P_0\left(\left(\omega_n^{\frac{n}{2}-1}\right)^2\right)+\omega_n^{\frac{n}{2}-1} P_1\left(\left(\omega_n^{\frac{n}{2}-1}\right)^2\right) \\ \hline P_0\left(\left(\omega_n^{\frac{n}{2}}\right)^2\right)+\omega_n^{\frac{n}{2}} P_1\left(\left(\omega_n^{\frac{n}{2}}\right)^2\right) \\ \vdots \\ P_0\left(\left(\omega_n^{n-1}\right)^2\right)+\omega_n^{n-1} P_1\left(\left(\omega_n^{n-1}\right)^2\right)\end{array}\right) \\
&\overset{\text{למה 2}}{\underset{(*)}{=}} \left(\begin{array}{c} P_0\left(\omega_{\frac{n}{2}}^0\right) \\ \vdots \\ P_0\left(\omega_{\frac{n}{2}}^{\frac{n}{2}-1}\right) \\ \hline P_0\left(\omega_{\frac{n}{2}}^0\right) \\ \vdots \\ P_0\left(\omega_{\frac{n}{2}}^{\frac{n}{2}-1}\right)\end{array}\right) + \begin{pmatrix} \omega_n^0 \\ \vdots \\ \omega_n^{\frac{n}{2}-1} \\ \omega_n^{\frac{n}{2}} \\ \vdots \\ \omega_n^{n-1}\end{pmatrix}\left(\begin{array}{c} P_1\left(\omega_{\frac{n}{2}}^0\right) \\ \vdots \\ P_1\left(\omega_{\frac{n}{2}}^{\frac{n}{2}-1}\right) \\ \hline P_1\left(\omega_{\frac{n}{2}}^0\right) \\ \vdots \\ P_1\left(\omega_{\frac{n}{2}}^{\frac{n}{2}-1}\right)\end{array}\right)
\end{aligned}$$

$(*)$ מכפלת הוקטורים היא איבר איבר.

נזכור כי $P_0\left(y\right)=\sum_{j=0}^{\frac{n}{2}-1} a_{2j}y^j$ הוא פולינום ב-$V_{\frac{n}{2}-1}$ ולכן מהגדרת התמרת פורייה בדידה מתקיים

$$\begin{pmatrix} P_0\left(\omega_{\frac{n}{2}}^0\right) \\ \vdots \\ P_0\left(\omega_{\frac{n}{2}}^{\frac{n}{2}-1}\right)\end{pmatrix} = \mathrm{DFT}_{\frac{n}{2}}\begin{pmatrix} a_0 \\ a_2 \\ \vdots \\ a_{n-2}\end{pmatrix}$$

באופן דומה $\mathrm{DFT}_{\frac{n}{2}}\begin{pmatrix} a_1 \\ a_3 \\ \vdots \\ a_{n-1}\end{pmatrix} = \begin{pmatrix} P_1\left(\omega_{\frac{n}{2}}^0\right) \\ \vdots \\ P_1\left(\omega_{\frac{n}{2}}^{\frac{n}{2}-1}\right)\end{pmatrix}$. לכן

$$\mathrm{DFT}_n\begin{pmatrix} a_0 \\ \vdots \\ a_{n-1}\end{pmatrix} = \left(\begin{array}{c} \mathrm{DFT}_{\frac{n}{2}}\begin{pmatrix} a_0 \\ a_2 \\ \vdots \\ a_{n-2}\end{pmatrix} \\ \hline \mathrm{DFT}_{\frac{n}{2}}\begin{pmatrix} a_0 \\ a_2 \\ \vdots \\ a_{n-2}\end{pmatrix}\end{array}\right) + \begin{pmatrix} \omega_n^0 \\ \vdots \\ \omega_n^{n-1}\end{pmatrix}\left(\begin{array}{c} \mathrm{DFT}_{\frac{n}{2}}\begin{pmatrix} a_1 \\ a_3 \\ \vdots \\ a_{n-1}\end{pmatrix} \\ \hline \mathrm{DFT}_{\frac{n}{2}}\begin{pmatrix} a_1 \\ a_3 \\ \vdots \\ a_{n-1}\end{pmatrix}\end{array}\right)$$

כאשר המכפלה היא שוב איבר איבר. המשוואה הנ"ל נותנת דרך ריקורסיבית לחישוב $\mathrm{DFT}_n\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}$: נחשב את $\mathrm{DFT}_{\frac{n}{2}}\begin{pmatrix}a_0\\ a_2\\ \vdots\\ a_{n-2}\end{pmatrix}$ ואת $\mathrm{DFT}_{\frac{n}{2}}\begin{pmatrix}a_1\\ a_3\\ \vdots\\ a_{n-1}\end{pmatrix}$ ונשתמש בזהות הנ"ל.

זמן הריצה: נסמן ב-$T(n)$ את זמן הריצה לחישוב DFT_n ונקבל $T(n)\le 2T\left(\frac{n}{2}\right)+3n$ (n פעולות העתקה, n פעולות חיבור ו-n פעולות כפל) ולכן ממשפט האב $T(n)=\mathcal{O}(n\log n)$. ■

אלגוריתם יעיל לכפל פולינומים בייצוג המקדמים

קלט שני פולינומים $P,Q\in V_{n-1}$ בייצוג המקדמים.

פלט הפולינום $V_{2n-2}\ni R=PQ$ בייצוג המקדמים.

נסמן ב-$\begin{pmatrix}c_0\\ \vdots\\ c_{2n-2}\end{pmatrix},\begin{pmatrix}b_0\\ \vdots\\ b_{n-1}\end{pmatrix},\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\end{pmatrix}$ ייצוגי המקדמים של P,Q,R ו-m החזקה של 2 הקטנה ביותר שגדולה מ-$2n-2$ (כדי שיהיו מספיק נתונים, $n-1$ לא מספיק).

אלגוריתם

1. נגדיר $\mathrm{DFT}_m\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\\ 0\\ \vdots\\ 0\end{pmatrix}=\begin{pmatrix}p_0\\ \vdots\\ p_{m-1}\end{pmatrix}$ ובאופן דומה $\mathrm{DFT}_m\begin{pmatrix}b_0\\ \vdots\\ b_{n-1}\\ 0\\ \vdots\\ 0\end{pmatrix}=\begin{pmatrix}q_0\\ \vdots\\ q_{m-1}\end{pmatrix}$.
2. נחשב $\mathrm{DFT}_m^{-1}\begin{pmatrix}p_0q_0\\ \vdots\\ p_{m-1}q_{m-1}\end{pmatrix}=\begin{pmatrix}\tilde{c_0}\\ \vdots\\ \tilde{c}_{m-1}\end{pmatrix}$.
3. נחזיר את $\begin{pmatrix}\tilde{c_0}\\ \vdots\\ \tilde{c}_{2n-2}\end{pmatrix}$.

טענה

1. האלגוריתם מחזיר את הוקטור $\begin{pmatrix}c_0\\ \vdots\\ c_{2n-2}\end{pmatrix}$ כלומר את וקטור המקדמים של R.
2. זמן הריצה של האלג' הוא $\mathcal{O}(n\log n)$.

הוכחה:

1. יהי ω_n שורש היחידה הפרימיטיבי מסדר m. נסמן $z_k=\omega_m^k$ לכל k. לכן מההגדרה של DFT_m מתקיים

$$\begin{pmatrix}p_0\\ \vdots\\ p_{m-1}\end{pmatrix}=\begin{pmatrix}P(z_0)\\ \vdots\\ P(z_{m-1})\end{pmatrix},\quad\begin{pmatrix}q_0\\ \vdots\\ q_{m-1}\end{pmatrix}=\begin{pmatrix}Q(z_0)\\ \vdots\\ Q(z_{m-1})\end{pmatrix}$$

ולכן $\begin{pmatrix} R(z_0) \\ \vdots \\ R(z_{m-1}) \end{pmatrix} = \begin{pmatrix} p_0q_0 \\ \vdots \\ p_{m-1}q_{m-1} \end{pmatrix}$ מהטענה שראינו (ומכך ש-$m \geq 2n-2$) מתקיים ש-R הוא הפולינום היחיד ממעלה $\geq m-1$ המקבל את הערכים $\begin{pmatrix} p_0q_0 \\ \vdots \\ p_{m-1}q_{m-1} \end{pmatrix}$ ולכן $\begin{pmatrix} c_0 \\ \vdots \\ c_{2n-2} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} p_0q_0 \\ \vdots \\ p_{m-1}q_{m-1} \end{pmatrix}$ ולכן בצעד השלישי של האלג' נחזיר את הוקטור DFT_m^{-1} $\begin{pmatrix} \tilde{c_0} \\ \vdots \\ \tilde{c}_{2n-2} \end{pmatrix} = \begin{pmatrix} c_0 \\ \vdots \\ c_{2n-2} \end{pmatrix}$ שהוא וקטור המקדמים של R.

2. לפי משפט ה-FFT זמן הריצה של האלג' הוא $\mathcal{O}(m\log m)$. נשים לב כי מתקיים $\frac{m}{2} \leq 2n-2$ ממינימליות m ולכן $m \leq 4n-4$ ולכן זמן הריצה הוא $\mathcal{O}(n\log n)$.

■

הצפנה

נתונים שני אנשים A(lice), B(ob) שרוצים לתקשר אחד עם השני בלי שאחרים יוכלו לפענח אותה את ההודעות שלהם בדרך. פורמלית, A שולחת ל-B הודעה m, מצפינה אותה להודעה m' ו-B מפענח אותה בחזרה ל-m.

ההצפנה הקלאסית היא להצפין באמצעות מפתחות הצפנה ופענוח סודיים המגדירים פ' הצפנה ופענוח סודיות $f_E : M \to M'$ ממרחב ההודעות להצפנות ו-$f_D : M' \to M$ כך שלכל $m \in M$ מתקיים $f_D(f_E(m)) = m$. זה נקרא הצפנה סימטרית.

דוגמה לכך היא OTP (One Time Pad), כלומר הגרלת סדרת ביטים אקראיית שמבצעים לה XOR עם ההודעה להצפנה (וגם לפענוח).

רעיון ההצפנה הפומבית הוצע לראשונה ב-1977 ע"י Diffie & Hellman, כאשר אלו הציבו דרישות להצפנה פומבית: לכל משתמש יש מפתח הצפנה פומבי (ולכן פ' הצפנה פומבית) ומפתח הצפנה סודי (המתאים לפ' פענוח סודית). כולם יכולים לשלוח ל-B הודעות מוצפנות שרק B יודע לפענח. נרצה גם ש-$M = M'$ ולכן הפ' $f_E : M \to M$ היא פ' חח"ע ושיתקיים $f_E(f_D(m)) = f_D(f_E(m)) = m$.

ב-1978 RSA הציעו את שיטת ההצפנה הפומבית הפופולרית ביותר כיום. יש מעט מאוד שיטות להצפנה פומבית ואף אחת מהן לא הוכחה באופן פורמלי ומתבססות על היוריסטיקה (הנחה בלתי מוכחת שטרם הופרכה).

שימושים להצפנה פומבית

1. "ספר טלפונים" של מפתחות פומביים. כל משתמש מכניס לרשימה את מפתח ההצפנה הפומבי שלו. כולם יכולים לשלוח לכולם הודעות מוצפנות.

2. "חתימה דיגיטלית". אם A רוצה לחתום על הודעה m היא מפרסמת את הזוג $(m, f_D(m))$. כולם יכולים לוודא ש-$f_E(f_D(m)) = m$ אבל רק A יכולה לחתום.

נניח ש-$M = \left\{0 \leq m \leq 2^{1000}-1\right\}$ (כל המספרים באורך 1000 ביטים). נרצה $f_E : M \to M$ קלה לחישוב אך קשה להיפוך.

דוגמה עבור $N \sim 2^{1000}$, $f_E(m) = m^2 \mod N$ היא פ' קשה להיפוך שהוצעה על ידי Rabbin (שעבד בהמשך גם עם Karp על אלג' Rabbin-Karp - כמה מקורי - למציאת מחרוזות ברצף מחרוזות בדומה לזה שראינו בתרגיל) אבל זה לא תפס.

שיטת ההצפנה הפומבית של RSA

אלגוריתם

1. נגריל מספרים ראשוניים $p, q \sim 2^{500}$ ונגדיר $N = pq$ ו-$M = \{0, \ldots, N-1\}$.
2. נמצא e הזר ל-$(p-1)(q-1)$ (שיהיה יחסית קטן, $e \sim \text{polylog} N$ כלומר פולינומיאלי ב-$\log N$).
3. נמצא $d = e^{-1} \mod (p-1)(q-1)$ כלומר $0 \leq d \leq N-1$ כך ש-$de = 1 \mod (p-1)(q-1)$.
4. נפרסם את (N, e) כמפתח הצפנה פומבי ונשמור את d כמפתח פענוח סודי.

שימוש פ' ההצפנה היא $f_E(e) = x^e \mod N$ ופ' הפענוח היא $f_D(y) = y^d \mod N$.

הערה הקושי בהפיכת f_E טמון בכך שתוקף לא יודע מה הם p, q כי מאוד קשה לפענח מה הם מ-N. מחשב קוואנטי כן יכול לפרוץ זאת כי הוא יודע לכפול וקטור במטריצה אורתוגנולית $n \times n$ ב-$\Theta(\log n)$ (ממש יעיל) ואילו האלג' של Shor לפירוק לגורמים ראשוניים משתמש במטריצה אורתוג' כנ"ל.

חלק ב' של ההרצאה

שאלות מתמטיות

1. למה קיימים p, q כאלה?
2. למה קיים e כזה?
3. למה קיים d כזה (ולמה הוא יחיד)?
4. למה לכל $x \in M$ מתקיים $f_E(f_D(x)) = f_D(f_E(x)) = x$?

שאלות אלגוריתמיות

1. איך ניתן למצוא p, q באופן יעיל ואיך נוודא שהמספרים שמצאנו ראשוניים (אלגוריתם מילר-רבין).
2. איך ניתן למצוא e באופן יעיל?
3. איך ניתן למצוא d באופן יעיל?
4. האם ניתן לחשב את f_E, f_D באופן יעיל?

החלק המתמטי מבוסס על תורת המספרים האלמנטרית (סיכום שלי ברמה נמוכה מאוד) ברובה ובעיקר על אריתמטיקה מודולרית.

טענה (חילוק עם שארית) בהינתן שני מספרים טבעיים a, b כך ש-$b \neq 0$, ניתן לחלק את a ב-b עם שארית, כלומר לרשום $a = qb + r$ כאשר $0 \leq r < b$. הדרך לרשום היא יחידה. q נקראת המנה ו-r השארית.

הגדרה בהינתן $b \geq 2$, נגדיר $\bmod b : \mathbb{N} \to \{0, \ldots, b-1\}$ באופן הבא: עבור $a \in \mathbb{N}$ נחלק את a ב-b עם שארית $a = qb + r$ ונגדיר $(\bmod b)(a) = r$. מסיבות הסיטוריות נרשום $a \mod b = r$.

בהינתן a ו-b ניתן לחשב את $a \mod b$ באופן יעיל (בזמן פולינומיאלי באורכי הייצוגים של a ו-b).

סימון נרשום $a \equiv c \mod b$ כדי להגיד ש-$a \mod b = c \mod b$.

טענה $a \equiv c \mod b$ אם"ם $a - c$ מתחלק ב-b.

תכונות מודולו

יהיו a, b, c כאשר $b \neq 0$.

1. $(a + c) \mod b = (a \mod b + c \mod b) \mod b$.
2. $ac \mod b = (a \mod b \cdot c \mod b) \mod b$.

דוגמה $(2022)^{22} \overset{(*)}{\equiv} 9^{22} = 81^{11} \equiv 4^{11} = 4 \cdot 16^5 \equiv 4 \cdot 5^5 = \ldots$.

$(*)\ 2000 = 20 \cdot 100 \equiv 9 \mod 9$.

הגדרות מתורת המספרים האלמנטרית

הגדרה <u>המחלק המשותף המקסימלי</u> של a, b הוא המספר הגדול ביותר המחלק גם את a וגם את b. נסמן אותו ב-$\gcd(a, b)$.

דוגמה $\gcd(24, 16) = 8$.

ניתן לחשב את $\gcd(a, b)$ באופן יעיל, כלומר בזמן פולינומיאלי באורכי הייצוגים (באמצעות אלג' אוקלידס).

הגדרה a, b נקראים <u>זרים</u> אם $\gcd(a, b) = 1$.

דוגמה $\gcd(40, 21) = 1$ ולכן הם זרים.

הגדרה p הוא <u>ראשוני</u> אם p מתחלק רק בעצמו וב-1.

טענה (המשפט היסודי של האריתמטיקה) לכל $a \in \mathbb{N}$ ניתן להציג את a כמכפלת חזקות של מספרים ראשוניים $a = p_1^{s_1} \cdot \ldots p_k^{s_k}$ כאשר p_i ראשוניים ו-s_i טבעיים.

טענה a, b זרים זה לזה אם אין מספר ראשוני המשותף בפירוק של שניהם לגורמים ראשוניים.

דוגמה $21 = 3 \cdot 7, 40 = 2^3 \cdot 5$.

משפט (התפלגות הראשוניים, צ'בישב) לכל $a \in \mathbb{N}$, נגדיר $\pi(a) = |\{2 \leq p \leq a : \text{ראשוני } p\}| \sim \frac{a}{\log a}$ כלומר $\lim_{a \to \infty} \frac{\pi(a)}{\frac{a}{\log a}} = 1$.

על סמך המשפט נוכל לבצע את השלב הראשון באלג' של RSA. נגריל מספר בן k ביטים (ע"י הגרלת k ביטים באופן ב"ת) ונבדוק באמצעות משפט מילר-רבין האם המספר שהגרלנו ראשוני. לפי המשפט הסיכוי להגריל ראשוני הוא $\frac{1}{k\log 2} = \frac{1}{\log 2^k}$ ולכן אחרי $\mathcal{O}\left(k\right)$ נסיונות נקבל ראשוני בסיכוי גדול.

הערה הסיכוי שהמספר יהיה גדול הוא גבוה כי כדי שהוא לא יהיה, נדרש שכל הביטים הגבוהים שלו 0 וזה קורה בהסת' נמוכה.

נוכל לבצע גם את השלב השני באלג' RSA. נגריל e ונבדוק האם הוא ראשוני. אחרי $\mathcal{O}\left(k\right)$ נסיונות נמצא ראשוני בסיכוי גבוה. בסיכוי גבוה הוא לא יחלק את $\left(p-1\right)\left(q-1\right)$ ולכן יהיה זר לו (מספר המספרים הראשוניים שלא זרים ל-$\left(p-1\right)\left(q-1\right)$ הוא מספר הגורמים הראשוניים השונים בפירוק לגורמים, שזה מעט מאוד יחסית לסדרי הגודל שלנו).

תרגול

במהלך רוב התרגול, נניח כי $0\in\mathbb{N}$.

פעולות על מספרים

1. חיבור - $\mathcal{O}\left(k\right)$
2. כפל וחילוק (ארוך) - $\mathcal{O}\left(k^2\right)$.
3. חישוב מודולו ($a \mod b = a-\lfloor\frac{a}{b}\rfloor b$) - $\mathcal{O}\left(k^2\right)$
4. העלאה בחזקה. נחשב את $a \mod b$, $a^2 \mod b = \left(a \mod b\right)\left(a \mod b\right) \mod b$ ושאר חזקות של חזקות של 2. מכאן נכפיל את a^{2^i} לכל i שעבורו הביט ה-i הוא 1. לדוגמה $2^{11}=2^8 2^2 2^1$. $\mathcal{O}\left(k^3\right)$ (מבצעים k פעולות כפל, ואז שוב עוד לכל היותר k פעולות כפל).

הגדרה בהינתן $a,b\in\mathbb{N}$, נאמר כי b <u>מחלק</u> את a ונסמן $b\mid a$ אם קיים $k\in\mathbb{N}$ כך ש-$a=kb$.

סקרנו הגדרות נוספות שהופיעו בחלק ב' של ההרצאה.

הערה 0 מתחלק בכל מספר שלם אבל אף מספר (אולי פרט ל-0) אינו מתחלק ב-0 ולכן $\gcd\left(a,0\right)=a$, $\forall a\geq 0$ ו-לא מוגדר $\gcd\left(0,0\right)=$.

טענה לכל $a\geq b\in\mathbb{N}$, $\gcd\left(a,b\right)=\gcd\left(b,a \mod b\right)$.

הערה את כל ההוכחות הקרובות אנחנו סתם מוכיחים בריגורוזיות בלתי קריאה, אנחנו בעצם משתמשים כל הזמן בתכונה שאם $d\mid x,y$ אז $d\mid x+y$.

הוכחה: נסמן $d=\gcd\left(a,b\right)$, $d'=\gcd\left(b,a \mod b\right)$.

נוכיח כי $d \mid a \mod b$. קיימים k_1, k_2 כך ש-$k_1 d = a, k_2 d = b$ ולכן

$$a \mod b = a - \lfloor \frac{a}{b} \rfloor b = dk_1 - \lfloor \frac{a}{b} \rfloor dk_2 = d\left(k_1 - \lfloor \frac{a}{b} \rfloor k_2\right)$$

ולכן $d \mid a \mod b$. לכן $d \le d'$ (d' מחלק משותף מקסימלי של $a, a \mod b$ ו-d סתם מחלק אותם).

נוכיח כי $a \mid d'$.

קיימים k_3, k_4 כך ש-$k_3 d' = b$, $k_4 d' = a \mod b$ ולכן

$$\begin{aligned} a &= a \mod b + \lfloor \frac{a}{b} \rfloor b = k_4 d' + \lfloor \frac{a}{b} \rfloor k_3 d' \\ &= d'\left(k_4 + \lfloor \frac{a}{b} \rfloor k_3\right) \end{aligned}$$

ולכן $a \mid d'$. לכן $d' \le d$ (d' מחלק משותף מקסימלי של a, b ו-d סתם מחלק אותם).

לכן לסיכום $d = d'$. ■

טענה יהיו $a, b \in \mathbb{N}$. $S = \{ax + by : x, y \in \mathbb{Z}, ax + by \ge 1\}$ אזי $\gcd(a, b) = \min S$.

הוכחה: נסמן $t = \gcd(a, b)$ וכן $z = \min S$.

נוכיח כי $t \le z$. לשם כך נראה כי $t \mid z$. קיימים k_5, k_6 כך ש-$a = k_5 t, b = k_6 t$. נשים לב כי z מהצורה $ax + by$ עבור $x, y \in \mathbb{Z}$ כלשהם ולכן

$$z = ax + by = k_5 tx + k_6 ty = t\left(k_5 x + k_6 y\right)$$

נוכיח כי $z \le t$. לשם כך נראה כי z מחלק משותף של a, b, וממקסימליות t ינבע כי $z \le t$.

ברור כי $z \le a$ ($a \cdot 1 + b \cdot 0 \in S$ ו-z מינימלי). נחלק את a ב-z עם שארית, $a = kz + r$, $r < z$. נוכיח כי $r = 0$ ונסיים.

נניח בשלילה כי $r > 0$, כלומר $r \ge 1$,

$$r = a - kz = a - (ax + by)k = a(1 - kx) + b(-yk) \in S$$

ולכן $r \in S$ כי $r \ge 1$ ומהצורה $ax' + by'$ אבל $r < z$ בסתירה למינימליות ולכן $z \mid a$.

באותו האופן ניתן להוכיח כי $z \mid b$ ולכן $z \le t$.

לסיכום $z = t$. ■

אלגוריתם אוקלידס המורחב (EE)

קלט $a \le b \in \mathbb{N}$.

פלט (g,x,y) כך ש-$g = \gcd(a,b) = ax + by$.

הגדרה יהיו $a,n \in \mathbb{N}$ מספרים זרים. אז ההופכי הכפלי של a מודולו n הוא מספר $p \in \mathbb{N}$ כך ש-$ap \mod n = 1$.

דוגמה $a=4, n=7$ אז $p=2$. אם $a=0, n=20$ אז $p=9$. אם $a=6, n=9$ אז אין p כזה.

הערה אם a,n זרים אזי $1 = \gcd(a,n) = ax + ny$. x הוא ההופכי הכפלי של a מודולו n כי $ax \mod n = ax + ny \mod n = 1$.

פסאודו-קוד

1. אם $b=0$ נחזיר $(a,1,0)$.
2. אחרת נחשב $(g',x',y') = \mathrm{EE}(b, a \mod b)$.
3. נחזיר $\left(g', y', x' - \lfloor \frac{a}{b} \rfloor y'\right)$.

דוגמה

$$
\begin{array}{ccccc}
EE(117,91) & = & \left(13,-3,1-(-3)\lfloor\frac{117}{91}\rfloor\right) & = (13,-3,4) \\
\downarrow & & & \\
EE(91,26) & & \left(13,1,0-1\cdot\lfloor\frac{91}{26}\rfloor\right) & = (13,1,-3) \\
\downarrow & & & \\
EE(26,13) & = & \left(13,0,1-0\cdot\lfloor\frac{26}{13}\rfloor\right) & = (13,0,1) \\
\downarrow & & & \\
EE(13,0) & = & (13,1,0) &
\end{array}
$$

ולכן $\gcd(117,91) = 13 = -3 \cdot 117 + 4 \cdot 91$.

משפט EE מחזיר (g,x,y) כרצוי.

הוכחה: באינדוקציה שלמה על b:

בסיס ($b=0$): $\gcd(a,b) = \gcd(a,0) = a \cdot 1 + b \cdot 0$ ואכן $(a,1,0)$.

צעד ($0,\ldots,b-1 \to b$):

$$
\begin{aligned}
\gcd(a,b) = \gcd(b, a \mod b) &= g' \\
&\overset{\text{ה"א}}{=} bx' + (a \mod b)\, y' \\
&= bx' + \left(a - \lfloor \frac{a}{b} \rfloor\right) y'
\end{aligned}
$$

ולכן $(g,x,y) = \left(g' y', x' - \lfloor \frac{a}{b} \rfloor y'\right)$ כמוחזר באלג'. $\blacksquare$

טענה האלג' EE מבצע לכל היותר $2\log b$ קריאות ריקורסיביות.

הוכחה: נחלק ל-2 מקרים:

1. אם $b > \frac{a}{2}$ אז $a \mod b = a - \lfloor \frac{a}{b} \rfloor b \le a - b < \frac{a}{2}$ (השארית קטנה מ-$a \mod b$)

2. אם $b \le \frac{a}{2}$ אז $a \mod b < b \le \frac{a}{2}$. כלומר אחרי קריאה אחת a קטן פי 2 ואחרי k קריאות הוא קטן פי 2^k ולכן אחרי לכל היותר $2\log b$ קריאות, b יהיה 0.

$\blacksquare$

מסקנה אם a, b מיוצגים ע"י k ביטים, EE מבצע $\mathcal{O}\left(k\right)$ קריאות ריקורסיביות ובכל קריאה פעולות כפל וחילוק שדורשות $\mathcal{O}\left(k^2\right)$, כלומר הוא רץ ב-$\mathcal{O}\left(k^3\right)$.

שבוע $\mathbb{XIV}$ | הוכחת נכונות RSA ואלגוריתם מילר-רבין

הרצאה

חלק א' של ההרצאה

למה 1

1. b מתחלק ב-a אם"ם כל גורם ראשוני שמופיע בפירוק של a לגורמים ראשוניים מופיע גם בפירוק של b, והוא מופיע בפירוק של b בחזקה גדולה או שווה לזו בה הוא מופיע בפירוק של a.
2. אם a, b זרים ל-c אז גם $a \cdot b$ זר ל-c וגם $ab \mod c$ זר ל-c (לכל אחד מהמספרים בנפרד אין גורמים משותפים ל-c ולכן גם ל-$a \cdot b$ אין).
3. אם a זר ל-c ו-$a \cdot b$ מתחלק ב-c אז b מתחלק ב-c.

מסקנה יהיו a, b מספרים טבעיים כך ש-a זר ל-b. נניח כי $1 \le c, d \le b - 1$ כך ש-$a \cdot c \equiv a \cdot d \mod b$ אזי $c = d$.

הוכחה: מכיוון ש-$ac \equiv ad \mod b$ הרי ש-$b \mid ac - ad$ ולכן $b \mid a\left(c - d\right)$ אבל a זר ל-b ולכן מסעיף 3 של למה 1 מתקבל כי $b \mid c - d$ ולכן מהיות $1 \le c, d \le b - 1$, $|c - d| < b$ אזי $c - d = 0$ ולכן $c = d$. $\blacksquare$

הגדרה יהי n מספר טבעי, נגדיר $\mathbb{Z}_n = \{0, \ldots, n - 1\}$ (זוהי חבורה ביחס לפעולת החיבור).

הגדרה נגדיר $\mathbb{Z}_n^* = \{0 < a < n : \gcd\left(a, n\right) = 1\}$ (זוהי חבורה כפלית ביחס לפעולת הכפל).

הגדרה <u>פונקציית אוילר</u> היא $\varphi(n) = |\mathbb{Z}_n^*|$.

דוגמה $n = 12$, $\mathbb{Z}_{12} = \{0, \ldots, 12\}$ ואילו $\mathbb{Z}_{12}^* = \{1, 5, 7, 11\}$ ולכן $\varphi(12) = 4$.

למה

1. אם p ראשוני אז $\varphi(p) = p - 1$.
2. אם p, q ראשוניים שונים אזי $\varphi(pq) = (p-1)(q-1)$.

הערה φ היא פ' כפלית ככלל, כלומר אם $\gcd(a, b) = 1$ אזי $\varphi(ab) = \varphi(a)\varphi(b)$.

הוכחה:

1. אם p ראשוני אזי $\mathbb{Z}_p^* = \{1, \ldots, p-1\}$ ולכן $\varphi(p) = p - 1$.
2. $\mathbb{Z}_{pq}^* = \{1 \leq a \leq pq - 1 : \gcd(a, pq) = 1\}$. נביט ב

$$B = \{1, \ldots, pq - 1\} \setminus \mathbb{Z}_{pq}^* = \{1 \leq a \leq pq - 1 : \gcd(a, pq) \neq 1\}$$

אזי B מכיל את כל המספרים בין 1 ל-$pq - 1$ המתחלקים ב-p או q (בלעדי - אם הוא מתחלק בשניהם הוא או אפס או מכפלה של pq וזה לא ייתכן בשל הטווח שלו).

$$B = \{p, 2p, \ldots, (q-1)p\} \dot{\bigcup} \{q, 2q, \ldots, (p-1)q\}$$

ולכן $|B| = (q-1) + (p-1)$ ולכן

$$\varphi(pq) = \left|\mathbb{Z}_{pq}^*\right| = (pq - 1) - |B| = pq - p - q + 1 = (p-1)(q-1)$$

■

הגדרה יהי n מספר טבעי ויהי a זר ל-n.

נוודא כי לכל $b \in \mathbb{Z}_n^*$ מתקיים כי גם $c = ab \mod n$ שייך ל-$\mathbb{Z}_n^*$: אכן $0 \leq c \leq n - 1$ ומהיות a, b זרים ל-n הרי שמחלק 2 של למה 1 גם c זר ל-n. לכן $c \in \mathbb{Z}_n^*$.

הגדרה נגדיר פ' $f_a : \mathbb{Z}_n^* \to \mathbb{Z}_n^*$ באופן הבא: לכל $b \in \mathbb{Z}_n^*$, $f_a(b) = ab \mod n$.

דוגמה $n=12$, $a=5$,

$$f_a(1)=5$$
$$f_a(5)=1$$
$$f_a(7)=11$$
$$f_a(11)=7$$

למה 2 יהי a זר ל-n. אזי f_a היא חחע"ל $\mathbb{Z}_n^*$.

הוכחה: מספיק להוכיח כי f_a חח"ע. נניח כי עבור $c,d\in\mathbb{Z}_n^*$ מתקיים $f_a(c)=f_a(d)$ ונוכיח כי $c=d$.

מההגדרה, $ac\equiv ad \mod n$ ואילו a זר ל-n ולכן מהמסקנה הנ"ל, $c=d$. ■

מסקנה יהי $e\in\mathbb{Z}_{(p-1)(q-1)}^*$ אזי קיים $d\in\mathbb{Z}_{(p-1)(q-1)}^*$ כך ש-$de\equiv 1 \mod (p-1)(q-1)$.

הוכחה: נתבונן בפ' f_e שהגדרנו למעלה. מלמה 2, f_e היא על $\mathbb{Z}_{(p-1)(q-1)}^*$. מהיות $1\in\mathbb{Z}_{(p-1)(q-1)}^*$ (כי $1\cdot 1\equiv 1 \mod (p-1)(q-1)$)
אזי קיים $d\in\mathbb{Z}_{(p-1)(q-1)}^*$ כך ש-$f_e(d)=1$, כלומר $de\equiv 1 \mod (p-1)(q-1)$. ■

משפט (אוילר) יהיו a,n מספרים זרים. אזי $a^{\varphi(n)}\equiv 1 \mod n$.

דוגמה $n=12$, $a=5$, $\varphi(12)=4$ ואכן $5^4=25^2\equiv 1^2=1 \mod 12$.

הוכחה: מלמה 2, f_a מוגדרת הנ"ל היא חח"ע ולכן

$$\begin{aligned}\left(\prod_{x\in\mathbb{Z}_n^*} f_a(x)\right) \mod n &= \left(\prod_{x\in\mathbb{Z}_n^*} ax \mod n\right) \mod n\\ &= \left(\prod_{x\in\mathbb{Z}_n^*} ax\right) \mod n\\ &= \left(a^{\varphi(n)}\prod_{x\in\mathbb{Z}_n^*} x\right) \mod n\\ &= a^{\varphi(n)} \mod n \left(\prod_{x\in\mathbb{Z}_n^*} x\right) \mod n\end{aligned}$$

ולכן עבור $A=a^{\varphi(n)} \mod n$, $B=\left(\prod_{x\in\mathbb{Z}_n^*} x\right) \mod n$ נקבל $1\cdot B=B=A\cdot B \mod n$ ונשים לב כי B זר ל-n מחלק 2 של למה 1 וגם $1\le 1, A\le n-1$ ולכן מהמסקנה מהלמה, $A=1$, כלומר $a^{\varphi(n)}\equiv 1 \mod n$. ■

מסקנה (ממשפט אוילר, נקרא משפט פרמה הקטן) יהי p מספר ראשוני ו-a כך ש-a לא מתחלק ב-p. אזי $a^{p-1} \equiv 1 \mod p$.

משפט יהי $x \in M$ אזי $f_E\left(f_D\left(x\right)\right) = f_D\left(f_E\left(x\right)\right) = x$.

הוכחה:

$$\begin{aligned} f_D\left(f_E\left(x\right)\right) &= \left(x^e \mod n\right)^d \mod n \\ &= x^{ed} \mod n \\ &= x^{de} \mod n \\ &\overset{\text{בדומה}}{\equiv} f_E\left(f_D\left(x\right)\right) \end{aligned}$$

נותר להוכיח כי $x^{de} \equiv x \mod N$. מבחירת e, d, מתקיים $de \equiv 1 \mod (p-1)(q-1)$ כלומר $de = a(p-1)(q-1)+1$ כאשר $a \in \mathbb{N}$. נוכיח כי $x^{a(p-1)(q-1)+1} \equiv x \mod N$. נחלק למקרים

1. x זר ל-N - ממשפט אוילר, $x^{\varphi(N)} \equiv 1 \mod N$ ואילו $\varphi(N) = (p-1)(q-1)$ ולכן $x^{(p-1)(q-1)} \equiv 1 \mod N$ ולכן

$$x^{a(p-1)(q-1)+1} = x \cdot \left(x^{(p-1)(q-1)}\right)^a \equiv x \cdot 1^a = x \mod N$$

2. x מתחלק ב-p אבל לא ב-q - מפרמה הקטן, מתקיים $x^{q-1} \equiv 1 \mod q$ ולכן מפרמה הקטן

$$x^{a(p-1)(q-1)+1} = x \cdot x^{a(p-1)(q-1)} = x\left(x^{q-1}\right)^{a(p-1)} \equiv x \cdot 1^{a(p-1)} = x \mod q$$

ולכן $q \mid x^{a(p-1)(q-1)+1} - x$. מצד שני x מתחלק ב-p ולכן $x^{a(p-1)(q-1)+1}$ מתחלק ב-p ולכן x ולכן $p \mid x^{a(p-1)(q-1)+1} - x$ ולכן $N \mid x^{a(p-1)(q-1)+1} - x$ שזה בדיוק $x^{a(p-1)(q-1)+1} \equiv x \mod N$.

3. x מתחלק ב-q אבל לא ב-p - אותו הדבר כמו המקרה הקודם.

4. x מתחלק ב-p וב-q - מהיות $x \in M$, הרי ש-$x \in \{0, \ldots, N-1\}$ אבל $N \mid x$ ולכן $x = 0$ ולכן גם $x^{a(p-1)(q-1)+1} - x = 0$ $N \mid$.

■

האלגוריתם של מילר ורבין לבדיקת ראשוניות

קלט מספר טבעי n.

פלט "n ראשוני" או "n אינו ראשוני".

הערה זו בעיית הערכה ולא אופטימיזציה.

הערה האלג' אמור לרוץ ב-$\mathcal{O}\left(\text{polylog}\left(n\right)\right)$ כלומר יעיל באורך הייצוג של n.

פסאודו-קוד

1. נבחר a מקרי ב-$\{1,\ldots,n-1\}$.
2. נחשב $f=a^{n-1} \mod n$
3. אם $f\neq 1$ נעצור ונחזיר "לא ראשוני", אחרת נמשיך לאיטרציה הבאה.
4. אם בכל ה-t איטרציות קיבלנו $f=1$, נחזיר "ראשוני"

הערות לגבי האלגוריתם של מילר-רבין

1. לפי המשפט הקטן של פרמה, אם n באמת ראשוני אז השוויון תמיד יתקיים.
2. אם n לא ראשוני, אז נוכיח שבהסת' לפחות חצי השוויון לא מתקיים, כלומר $a^{n-1}\not\equiv 1 \mod n$ (זה נכון לכל מספר שאינו מספר Carmichael, אבל יש ממש קצת כאלה והאלג' המלא של מילר-רבין מטפל גם בהם).

   לכן אם נחזור על הצעד הבסיסי t פעמים אז:

   (א) אם n ראשוני נענה (תמיד) "כן ראשוני".

   (ב) אם n אינו ראשוני נטעה (נענה "כן ראשוני") לכל היותר בסיכוי $\frac{1}{2^t}$.

לסיכום השגיאה היחידה שיכולה להיות היא false positive, כלומר שאנחנו טוענים שלא ראשוני הוא ראשוני (שזה רע). לכן זהו אלג' הסת' שמחזיר פתרון נכון בהסת' לפחות $1-\frac{1}{2^t}$ (אלא אם זה Carmichael, כאמור).

להמשך הניתוח

סוף.