① 4/11/08

כלים מתמטיים

(נתי ליניאל)

מתרגל: אילון נחמה

יהיו תרגילים ויהיו בחנים (שלא משפיעים על הציון הסופי) ויהיה מבחן / מבחן בית.

> באתר הקורס יש סכומים. נתי ישמח לקבל הערות/תיקונים.

לקורס יש ארבעה חלקים:

- הסתברות
- אלגברה לינארית (חשובה)
- אנליזה (הרמונית)
- אופטימיזציה

על כל אחד מהנושאים יהיה בוחן שדורש ממש מינימום של החומר.

אזהרה: כשנתי מתלהב הוא מתחיל לרוץ אז לחיצה על הבלמים תתקבל בסבר פנים יפות.

# הסתברות

נתחיל עם מקרה מאוד מצומצם של התיאוריה. היתרון הוא שהוא די פשוט ואינטואיטיבי וקומבינטורי ביסודו. אלא שיש לו כמה מגבלות ושם כבר נצטרך מרחבי הסתברות כלליים.

## מרחבי הסתברות בדידים

<u>הגדרה</u>: <u>מרחב הסתברות בדיד</u> $\Omega$ הוא קבוצה סופית או בת-מניה ביחד עם פונקצית משקל: לכל $x\in\Omega$ מותאם מספר ממשי אי-שלילי $p_x = P_r(x) \ge 0$ ביחד עם התנאי $\sum_{x\in\Omega} p_x = 1$.

<u>דוגמה</u>: מרחב ההסתברות הוא קבוצת הקורסים שלומד תלמיד ולכל קורס מותאם אחוז השעות שהתלמיד משקיע בו...

הגדרה: יהי $\Omega$ מרחב הסתברות בדיד. כל תת קבוצה של $\Omega$ נקראת מאורע. אם $A \subseteq \Omega$ מאורע אז מייחסים לו גם הסתברות

$$Pr(A) = \sum_{x\in A} Pr(x)$$

אם $A=\{x\}$ יחידון אז אומרים ש- $A$ מאורע אלמנטרי.

כאשר מביטים בחלק של מרחב הסתברות ניתן (לאחר נורמליזציה מתאימה) לראות גם אותו כמרחב הסתברות.

הגדרה: אם $B \subseteq \Omega$ מאורע עם הסתברות חיובית (כלומר $Pr(B) \gneq 0$) ואם $A \subseteq \Omega$ מאורע אז נגדיר את ההסתברות של $A$ בהינתן $B$:

$$Pr(A|B) = \frac{Pr(A\cap B)}{P(B)}$$

דוגמה: תלמיד לומד תואר דו-חוגי מדעי המחשב וביולוגיה וההסתברות כמו קודם. יתכן שנרצה לבחון רק את לימודי הביולוגיה שלו, ואלו הם מהווים רק 40% מכלל הקורסים. נניח שהקורס מתא לאורגניזם הוא 10% מכלל הקורסים. אז כשאנחנו מסתכלים רק על הביולוגיה מתא לאורגניזם לוקח

$$25\% = 0.25 = \frac{0.1}{0.4} = \frac{Pr(\text{מתא...})}{Pr(\text{ביולוגיה})}$$

הגדרה: אומרים שמאורעות $A,B \subseteq \Omega$ הם בלתי תלויים (בת"ל) אם מתקיים $Pr(A\cap B) = Pr(A)\,Pr(B)$

אם $Pr(B) \neq 0$, $A,B$ בת"ל אמ"מ

$$Pr(A|B) = Pr(A)$$

הגדרה: משתנה מקרי ממשי הוא פונקציה $f: \Omega \to \mathbb{R}$

דוגמה: במקרה של הסטודנט עם הקורסים $f$ יכולה להיות הציון של הסטודנט בכל קורס ואז מי שזוכר מה זה תוחלת יודע שהתוחלת של $f$ היא הממוצע (המשוקלל) של הסטודנט...

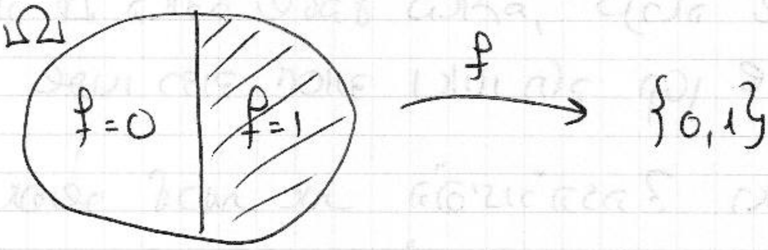
באופן דומה ניתן להגדיר משתנה מקרי לכל קבוצה סופית $R$ (ולאו דווקא $\mathbb{R}$)

<u>הגדרה</u>: יהי $\Omega$ מרחב הסתברות ויהי $f$ משתנה מקרי. <u>התוחלת</u> של $f$ הינו הממוצע הממושקל של $f$, כלומר

$$\mathbb{E}f = \sum_{x\in\Omega} f(x)p_x$$

יש קשר הדוק בין מאורעות לבין משתנים מקריים (מ"מ). אם $f$ מ"מ המקבל את הערכים $\{0,1\}$ (מ"מ כזה נקרא גם משתנה מקרי מציין) מותאם ל-$f$ המאורע

$$A_f = \{x\in\Omega : f(x)=1\}$$




<u>הגדרה</u>: אם $f:\Omega\to R_1$, $g:\Omega\to R_2$ שני מ"מ אומרים ש- $f,g$ <u>משתנים מקריים בלתי תלויים</u> אם

$$\forall y_1\in R_1\ \forall y_2\in R_2\quad Pr(f=y_1, g=y_2) = Pr(f=y_1)\,Pr(g=y_2)$$

<u>הסבר</u>: מה פירוש הכתיב $Pr(f=y_1)$? ובכן, אין הסתברות בלי מרחב הסתברות! $Pr$ עובדת על מאורעות. אז $f=y_1$ צריך לסמל מאורע כלשהו - תת קבוצה של מרחב ההסתברות. ובאופן די אינטואיטיבי $f=y_1$ הוא

$$\{x\in\Omega : f(x)=y_1\} = f^{-1}(\{y_1\})$$

ובאופן דומה "$f=y_1, g=y_2$" זה למעשה המאורע

$$\{x\in\Omega : f(x)=y_1 \wedge g(x)=y_2\} = f^{-1}(\{y_1\})\cap g^{-1}(\{y_2\})$$

אז ההגדרה למעשה אומרת שהמאורעות $f=y_1$ ו- $g=y_2$ הם בלתי תלויים (לפי ההגדרה למאורעות שכבר ראינו קודם) וזה לכל $y_1\in R_1$ ו- $y_2\in R_2$ שנבחר!

יש סיפור על אדם שטבע באגם שעומקו הממוצע 10 מ'. אז הממוצע עצמו לא נותן לנו הרבה מידע. היינו רוצים גם איזשהו מידע על "הפיזור".

בהינתן מ"מ מקרי $f$ על מרחב הסתברות $\Omega$, מעניין להבין עד כמה הערכים של $f$ "מפוזרים". למשל, נסתכל על הציונים של שני סטודנטים:

| | א' | ב' |
|---|---|---|
| אינפי | 90 | 100 |
| לינארית | 90 | 80 |

יש להם אותו ממוצע אבל ה"פיזור" שונה. או למשל באותו סיפור על האדם שטבע באגם, כנראה שהיה לאגם חלק מאוד גדול שהוא רדוד מאוד ואולי חלק קטן שעמוק רב מאוד.

איך אפשר לכמת את ה"פיזור" הזה? ראשית, נזיז את המספרים רק ש"ינועו סביב 0". כלו' במקום להביט ב- $\{f(x)\}_{x\in\Omega}$ נביט במספרים $\{f(x)-\mathbb{E}f\}_{x\in\Omega}$. ובעת, את השאלה האינטואיטיבית "עד כמה מפוזרים ערכי הפונקציה $f$" מחליפה כעת השאלה "עד כמה גדולים/קטנים המספרים $|f(x)-\mathbb{E}f|$ ?"
את הכימות הזה נעשה ע"י הסתכלות בממוצע של ריבועי ההפרשים.

<u>הגדרה:</u> בהינתן משתנה מקרי $f:\Omega\to\mathbb{R}$ נגדיר את <u>השונות</u>
$$\mathrm{Var} f = \mathbb{E}(f^2) - (\mathbb{E}f)^2 \quad (\geq 0)$$

עוד מעט נראה איך ההגדרה מתחברת לדיון האינטואיטיבי.

נשים ♡ שהתוחלת היא לינארית: אם $f,g:\Omega\to\mathbb{R}$ מ"מ ו- $\alpha,\beta\in\mathbb{R}$ אז $\alpha f+\beta g$ מ"מ ומתקיים
$$\mathbb{E}(\alpha f+\beta g) = \alpha\mathbb{E}f + \beta\mathbb{E}g$$
זה נובע באופן טריוויאלי מחוק הדיסטריבוטיבי...

עכשיו אנחנו יכולים לקשר בין הדיון האינטואיטיבי לבין ההגדרה הפורמלית.

(3) בדיון האינטואיטיבי דיברנו על מדידת השונות כ-

$$\mathbb{E}((f-\mathbb{E}f)^2)$$

עכשיו נשתמש בלינאריות:

$$\mathbb{E}((f-\mathbb{E}f)^2) = \mathbb{E}(f^2 - 2f\mathbb{E}f + (\mathbb{E}f)^2) =$$

פתיחת סוגריים

לינאריות $= \mathbb{E}(f^2) - 2\mathbb{E}(f\mathbb{E}f) + \mathbb{E}((\mathbb{E}f)^2)$

$$= \mathbb{E}(f^2) - 2(\mathbb{E}f)(\mathbb{E}f) + (\mathbb{E}f)^2 =$$

$\mathbb{E}f$ גודל קבוע לכן יוצא מהלינאריות והתוחלת של ערך קבוע הוא אותו ערך $= \mathbb{E}(f^2) - (\mathbb{E}f)^2$

עוד כמה עובדות בסיסיות חשובות:

- <u>חסם האיחוד</u>: אם $A_1,\dots,A_k \subseteq \Omega$ מאורעות אז
$$Pr\left(\bigcup_{i=1}^{k} A_i\right) \le \sum_{i=1}^{k} Pr(A_i)$$
- <u>אי שוויון מרקוב</u>: אם $X:\Omega \to \mathbb{R}^+$ מ"מ אי שלילי אז
$$\forall c\ge 1 \quad Pr(X \ge c\mathbb{E}X) \le \frac{1}{c}$$

(אפשר גם לרשום לכל $0<c$ אבל המקרה ש- $0<c\le 1$ הוא טריוויאלי כי כל הסתברות חסומה ע"י 1)

<u>דוגמה לשימוש</u>: שוב אותו אגם שממוצעו 10 ס"מ. ניקח $c=2$ ו- $X$ מתאר את עומק האגם. אז

$$Pr(X \ge \tfrac{1}{2}\cdot 10) = Pr(X\ge 5) \le \frac{1}{2}$$

אז לכל היותר חצי מהאגם הוא בעומק גדול מ- 5 ס"מ...

- <u>אי-שוויון צ'בישב</u>:

דיון אינטואיטיבי: "ריכוז מידה" פירושו שמ"מ מסוים נוטה לקבל ערכים הקרובים לממוצע (= התוחלת) שלו. זו תכונה שמ"מ יכול לקיים או לא לקיים. למשל, נתבונן בהתפלגות משכורות במדינה סוציאליסטית ומדינה קפיטליסטית:

יש ריכוז מידה — אין ריכוז מידה

האינטואיציה אומרת שאם השונות של מ"מ נתון היא קטנה אז יש לו תכונה של ריכוז מידה, כלומר אם מ"מ נתון נוטה להיות רחוק מתוחלתו אז בהכרח יש לו שונות גדולה.

וזה מה שאי שוויון צ'בישב אומר:

יהי $X$ מ"מ ממשי ונסמן $\mu = \mathbb{E}X$ ו- $\operatorname{Var} X = \sigma^2$. אז

לכל $c \geq 1$
$$\Pr(|X-\mu| \geq c\sigma) \leq \frac{1}{c^2}$$

כתוב כאן למעשה שהסיכוי שהמשתנה סוטה בהרבה מהתוחלת שלו הוא קטן. וזה כתלות ב-$\sigma$. באותה מדינה קפיטליסטית ה-$\sigma$ גדול מאוד... ואז המדידה של המרחק מהתוחלת היא ביחידות של $\sigma$.

את אי-שוויון צ'בישב אפשר להוכיח בשתי דקות מאי-שוויון מרקוב.

נגדיר $Y=(X-\mu)^2$. כעת $Y$ כמובן מ"מ אי שלילי. לכן אפשר להשתמש באי-שוויון מרקוב. רק צריך לברר את התוחלת של $Y$.

אבל כבר ראינו קודם ש- $\mathbb{E}Y = \mathbb{E}((X-\mu)^2) = \operatorname{Var} X$

נציב את זה בתוך אי שוויון מרקוב ונקבל:
$$\Pr((X-\mu)^2 \geq a \operatorname{Var} X) \leq \frac{1}{a}$$

לכל $a$, ובפרט אם ניקח $a=c^2$ נקבל
$$\Pr((X-\mu)^2 \geq c^2 \operatorname{Var} X) \leq \frac{1}{c^2}$$

אבל $u^2 \geq v^2$ אמ"מ $|u| \geq |v|$ לכן המאורעות $\{(X-\mu)^2 \geq c^2 \operatorname{Var} X\}$ ו- $\{|X-\mu| \geq c\sigma\}$ הם <u>זהים</u>.

לכן נקבל את הדרוש:
$$\Pr(|X-\mu| \geq c\sigma) \leq \frac{1}{c^2}$$ ☺

שוב, נתי רצה לפטש (hammer) לנו לתוך הראש: אין הסתברות ללא מרחב הסתברות וכשמדברים על המאורע $(X-\mu)^2 \geq c^2 \operatorname{Var} X$ הכוונה היא למעשה למאורע שבאופן פורמלי יש לרשום אותו כ-
$$\{\omega \in \Omega : (X(\omega)-\mu)^2 \geq c^2 \operatorname{Var} X\}$$

עכשיו נעבור לשימושים בכל מה שלמדנו...

רוצים להבין "איך נראית" תמורה מקרית. זאת שאלה לא מוגדרת היטב. אבל ממש אפשר לשאול כמה מחזורים יש בתמורה. זה כבר שאלה קונקרטית יותר. יש עוד שאלות קונקרטיות שאפשר לשאול.

אבל, אין הסתברות בלי מרחב הסתברות!

קודם צריך להגדיר את הבסיס שאנחנו עובדים עליו. מרחב ההסתברות שלנו יהיה $\Omega = S_n$ עם התפלגות אחידה. כלומר

$S_n$ = אוסף כל התמורות של $[n] = \{1, \dots, n\}$

וההסתברות היא $\Pr(\pi) = \frac{1}{|S_n|} = \frac{1}{n!}$

נגדיר מ"מ $X : \Omega \to \mathbb{N}^+$ ע"י $X(\pi) =$ מספר המעגלים בתמורה $\pi$

סוג השאלות שנרצה לשאול:

- לחשב (בדיוק או בערך) את $\mathbb{E}X$
- לחשב (בדיוק או בערך) את $\operatorname{Var} X$
- להעריך בהינתן $a > b$ שלמים חיוביים את $\Pr(a \geq X \geq b)$

אומרים ש-$1 \leq j \leq n$ היא <u>נקודת שבת</u> (fixed point) של התמורה $\pi \in S_n$ אם $\pi(j) = j$.

נסתכל במרחב ההסתברות $\Omega$ מקודם ($S_n$ עם הסתברות אחידה). ונגדיר $Z(\pi) =$ מספר נקודות השבת של $\pi$ מ"מ טבעי.

<u>בעיה:</u> מהו $\mathbb{E}Z$?

<u>תשובה:</u> $\mathbb{E}Z = 1$

<u>הוכחה:</u> אפשר לכתוב $Z = Z_1 + \dots + Z_n$ כאשר

$$Z_i(\pi) = \begin{cases} 1 & \pi(i) = i \\ 0 & \text{אחרת} \end{cases}$$

וקל להשתכנע שאכן $Z(\pi) = \sum_{i=1}^{n} Z_i(\pi)$. כעת, מהלינאריות של התוחלת אפשר לרשום: $\mathbb{E}Z = \sum_{i=1}^{n} \mathbb{E}Z_i$

אבל לאף אחד מה-$Z_i$-ים אין מעמד מיוחד ולכן התוחלות שלהם שווה ולכן $\mathbb{E}Z = n \cdot \mathbb{E}Z_1$. ולכן מספיק להוכיח ש-$\mathbb{E}Z_1 = \frac{1}{n}$.

לשם כך צריך לעשות חישוב קטן. מכיוון ש-$Z_1$ היא מ"מ מציין $\mathbb{E}Z_1$ היא בדיוק ההסתברות של-$\pi$ יש נקודת שבת ב-1.

טוען,

$$\mathbb{E} Z_1 = Pr_{\pi \in S_n}(\pi(1)=1) = \frac{\text{מספר התמורות } \pi \text{ רק ש- } \pi(1)=1}{n!} \overset{(*)}{=} \frac{(n-1)!}{n!} = \frac{1}{n}$$

וזה מה שרצינו! ☺

* רק נסביר למה מספר התמורות $\pi$ רק ש- $\pi(1)=1$ הוא $(n-1)!$. באופן כללי $|S_k| = k!$. ברגע שקבענו נקודה אחת של $\pi$ יש לנו משחק מלא עם $n-1$ הנקודות הנותרות, כלומר למעשה נותרנו עם תמורות ב- $S_{n-1}$ ולכן יש $(n-1)!$ כאלה...

באופן כללי יותר, בהינתן $1 \le k \le n$ נגדיר מ"מ $Y$ רק ש-
$Y(\pi) =$ מספר המעגלים באורך $k$ בתמורה $\pi$

משל $k=1$ $Y$ הוא למעשה $Z$ הנ"ל.

נרצה כעת לחשב את התוחלת של $Y$.

לאור ניסיוננו, ננסה להציג את $Y$ כסכום של מ"מ פשוטים יותר.

לכל $T \subseteq [n]$ רק ש- $|T| = k$ נסמן מ"מ

$$Y_T(\pi) = \begin{cases} 1 & \pi|_T = \text{מעגל} \\ 0 & \text{אחרת} \end{cases}$$

כמקודם $Y = \sum_{\substack{|T|=k \\ T \subseteq [n]}} Y_T$ ! מהלינאריות $\mathbb{E}Y = \sum_{\substack{|T|=k \\ T\subseteq[n]}} \mathbb{E}Y_T$

אבל יש $\binom{n}{k}$ קבוצות $T$ כאלה ומשיקולי סימטריה התוחלות של כל ה- $Y_T$ שוות. ולכן $\mathbb{E}Y = \binom{n}{k}\mathbb{E}Y_{\{1,\dots,k\}}$.

היות ש- $Y_{\{1,\dots,k\}}$ מ"מ מציין, $\mathbb{E}Y_{\{1,\dots,k\}}$ זו ההסתברות שאיברי $\{1,\dots,k\}$ מעגל בתמורה מקרית $\pi$. אז נחשב:

$$\mathbb{E}Y_{\{1,\dots,k\}} = \frac{\text{מספר התמורות שבהן } \{1,\dots,k\} \text{ מעגל}}{n!} = \frac{\text{מספר התמורות שבהן } 1,\dots,k \text{ מעגל} \cdot (n-k)!}{n!} = \frac{(k-1)!\,(n-k)!}{n!}$$

$$\Rightarrow \mathbb{E}Y = \binom{n}{k}\cdot\frac{(k-1)!\,(n-k)!}{n!} = \frac{\cancel{n!}}{k!\,\cancel{(n-k)!}}\cdot\frac{(k-1)!\,\cancel{(n-k)!}}{\cancel{n!}} = \frac{1}{k}$$

אז התוחלת של מספר המעגלים מאורך $k$ לתמורה מקרית $\pi \in S_n$ היא $\frac{1}{k}$.

<u>שאלה</u>: מהי התפלגות של מספר המעגלים בתמורה מקרית?
נסמן ב- $X$ את המ"מ המונה את המעגלים בתמורה.
נסמן ב- $X_k$ את המ"מ המונה את מספר המעגלים מאורך $k$
בתמורה. קל מאוד לראות ש- $X = X_1 + \dots + X_n$.
מלינאריות של התוחלת נובע ש-

$$\mathbb{E}X = \mathbb{E}X_1 + \dots + \mathbb{E}X_n = \sum_{k=1}^{n} \frac{1}{k} = \ln(n) + O(1)$$

<u>תזכורת</u>: למה יש $(k-1)!$ תמורות מעגליות באורך $k$?
נניח שנתחיל את המעגל ב-$1$. הוא יכול לעבור לכל מקום חוץ מעצמו.
אז יש $k-1$ אפשרויות. נניח ש- $1 \to j$. ל-$j$ יש $k-2$
אפשרויות (כולם חוץ מ-$1$ ו-$j$) וכן הלאה. המכפלה היא $(k-1)!$

<u>משפט רמזי בגרפים (Ramsey)</u>: לכל $k, l$ טבעיים,
אם $n \geq \binom{k+l-2}{k-1}$ אז בכל גרף עם $n$ קודקודים יש
$k$-קליקה או $l$-אנטי-קליקה

⑥ 11/11/08
טים

<u>משפט רמזי</u>: (אי סדר מוחלט לא ייתכן - יש כאן מערכת גדולה וחייבים להיות איים של סדר. בגרפים איים של סדר זה קליקה - גרף מלא ואנטי קליקות - גרף שבו אף שני קודקודים לא מחוברים).

לכל טבעיים $k,l$, כל גרף עם $n$ קודקודים כך ש- $n \ge \binom{k+l-2}{k-1}$ מכיל $k$-קליקה או $l$-אנטי-קליקה.

יש למשפט הוכחה אינדוקטיבית נחמדה. לא נתעמק בזה. רק נשים לב שאפשר לנסח את המשפט באופן הבא: אם $n, l, k$ כנ"ל ואם צובעים את הצלעות של $K_n$ בכחול ואדום אז יש $k$-קליקה כחולה או $l$-קליקה אדומה (כאן פשוט כחול מייצג קיום צלע ואדום מייצג אי-קיום צלע).

נתבונן במקרה ש- $k=l$. אם $n \ge \binom{2k-2}{k-1} \sim \frac{4^k}{c\sqrt{k}}$ אז בכל גרף עם $n$ קודקודים יש $k$-קליקה או $k$-אנטי-קליקה.

באופן טבעי נשאלת השאלה האם החסם הזה הדוק. לא יודעים לתת לזה הוכחה מבנית. אבל בעזרת השיטה ההסתברותית אנו נראה שהחסם הנ"ל לא מאוד רחוק מן האמת. באופן יותר מדויק, נראה שיש גרפים עם $\frac{2^{k/2}}{poly(k)}$ קודקודים ללא $k$-קליקה וללא $k$-אנטי-קליקה.

אנחנו מחפשים את הטענה הטובה ביותר מהסוג: בכל גרף עם $f(k)$ קודקודים יש $k$-קליקה או $k$-אנטי-קליקה. ראינו ש

$$\frac{2^{k/2}}{k^c} < f(k) < \frac{4^k}{c\sqrt{k}}$$

נראה ע"י השיטה ההסתברותית

<u>הוכחה</u>: נבחר בגרף מקרי $G(n,\frac{1}{2})$ (תכף נסביר) על $n$ קודקודים ונראה שיש גרף כזה ללא $k$-קליקה וללא $k$-אנטי-קליקה. כאשר $n=\frac{2^{k/2}}{k^c}$. $G(n,\frac{1}{2})$ הוא גרף שמתקבל באופן הבא: מתחילים עם $n$ קודקודים ולכל זוג מהם $(x,y)$ באופן בלתי תלוי ובהסתברות $\frac{1}{2}$ מוסיפים

את הצלע לעצמך בגרף. נסמן ב- $\Omega_n$ את קבוצת כל הגרפים בעלי n קודקודים. אז את $\Omega_n$ אנו הופכים למרחב הסתברות בהתפלגות אחידה. ז"א לכל גרף (מתויג) בעל n קודקודים G מתקיים $Pr(G) = (\frac{1}{2})^{\binom{n}{2}}$ (כי ב- n קודקודים יש $\binom{n}{2}$ זוגות ולכל זוג מחליטים באופן בת"ל אם יש צלע או לא וזה נותן לנו גרף).

על $\Omega_n$ נגדיר שני משתנים מקריים:

$X(G) =$ מספר ה-k-קליקות בגרף G

$Y(G) =$ מספר ה-k-אנטי-קליקות בגרף G

מה הולך לקרות עכשיו? אנחנו נראה ש- $\mathbb{E}X + \mathbb{E}Y < 1$

$\Leftarrow$ קיים G רק ש- $X(G)+Y(G) < 1$. אבל $X(G)+Y(G)$ הם מספר טבעי אי שלילי. לכן $X(G)+Y(G) = 0$. אבל $X(G), Y(G)$ מספרים טבעיים אי שליליים $\Leftarrow X(G)=Y(G)=0$

$\Leftarrow$ ב- G זה אין k-קליקה (כי $X(G)=0$) ואין אנטי קליקה (כי $Y(G)=0$). וזה מה שרצינו. גאוני!

אז צריך רק להוכיח ש- $\mathbb{E}X + \mathbb{E}Y < 1$. נתמקד ב- $\mathbb{E}X$. באופן דומה מטפלים גם ב- $\mathbb{E}Y$. הכלי היחיד שיש לנו בארגז הכלים הוא פירוק של X לסכום משתנים מקריים פשוטים יותר ושימוש בלינאריות של התוחלת. נגדיר

$$X = \sum_{\substack{|T|=k \\ T\subseteq[n]}} X_T$$ כאשר

$$X_T(G) = \begin{cases} 1 & \text{T קליקה ב-G} \\ 0 & \text{אחרת} \end{cases}$$

$$\Rightarrow \mathbb{E}X = \sum_{|T|=k} \mathbb{E}X_T = \binom{n}{k} Pr\left(\substack{\{1,\dots,k\} \text{ היא} \\ \text{קליקה ב-G}}\right) = \binom{n}{k}\cdot\left(\frac{1}{2}\right)^{\binom{k}{2}}$$

כי אין משנה איזה לכל T כנ"ל

כי לכל זוג ב- [k] חייבים להגרל צלע והשאר לא משנה...

כעת נראה ש- $\binom{n}{k}(\frac{1}{2})^{\binom{k}{2}} < \frac{1}{2}$. ידוע ש- $\binom{n}{k} \lessapprox (\frac{ne}{k})^k$.

אז השאלה הוא אם $(\frac{ne}{k})^k < 2^{\frac{k(k-1)}{2}-1}$, כלומר אם

אז לאפשר אם $\frac{ne}{k} < 2^{\frac{k-1}{2}-\frac{1}{k}}$

$$n < \frac{1}{e} k \cdot 2^{\frac{k-1}{2}-\frac{1}{k}}$$

ועכשיו זה כבר ברור. נשים ♡ שוב כאן רמאות קטנה כי לא אמרנו ש- $\left(\frac{ne}{k}\right)^k$ הוא חסם מלמעלה של $\binom{n}{k}$ אלא ידוע ש- $\binom{n}{k} < k\left(\frac{ne}{k}\right)^k$ (בדוק!) ואפשר להשתמש גם בזה ועדיין מתקבל את שצריך. (:

<u>אי שוויון צ'בישב</u>: יהי $X$ מ"מ עם תוחלת $\mu$ ושונות $Var X = \sigma^2$. אז $Pr(|X-\mu| \geq c\sigma) \leq \frac{1}{c^2}$ (*) לכל $0<c$.

כלומר במילים: יש רק סיכוי קטן שהמ"מ יסטה הרבה מהתוחלת שלו. ומה זה "הרבה"? איך מודדים סטיה כזו? מסתבר שיחידת המדידה הנכונה בהקשר זה היא ביחידות של סטיית התקן $\sigma$.

למשל, מקרה פרטי של (*) הוא $Pr(X=0) \leq \frac{Var X}{(\mathbb{E}X)^2}$

למה? נשים לב שהמאורע $X=0$ גורר את המאורע $|X-\mu| \geq \mu$ ולכן

$$Pr(X=0) \leq Pr(|X-\mu| \geq \mu) \leq \frac{1}{\left(\frac{\mu}{\sigma}\right)^2} = \frac{Var X}{(\mathbb{E}X)^2} \quad (**)$$

אם $A \subseteq B$ אז $Pr(A) \leq Pr(B)$

$\mu = c\sigma \Rightarrow c = \frac{\mu}{\sigma}$

מה המשמעות של האי-שוויון הזה? אם השונות קטנה והממוצע גבוה אז הסיכוי ש- $X$ מתאפס קטן - כלומר, רוב הזמן הוא חיובי (אם מלכתחילה $X$ אי-שלילי).

$G(n,p)$ הוא מרחב הסתברות של כל הגרפים המתויגים בעלי $n$ קודקודים כאשר לכל זוג קודקודים $x,y$ יש צלע בין $x,y$ בהסתברות $p$ ובאופן בלתי תלוי בכל הצלעות האחרות. אם $e(G)=t$ ($e$ = מספר הצלעות) אז $Pr(G) = p^t(1-p)^{\binom{n}{2}-t}$

למה המודלים האלה מעניינים? הם יכולים לתאר כל מיני תופעות פיסיקליות - יש אובייקטים והם באופן בלתי תלוי יכולים ליצור קשר אחד עם השני. למשל במכניקה סטטיסטית בוחנים תופעות שברמת המיקרו אנחנו לא יכולים להבין אותן...

תופעה מעניינת ואופיינית מאוד במודל $G(n,p)$ היא של מעבר פאזה / תופעת סף. ספציפית, אנחנו נתעניין בשאלה אם יש או אין בגרף 4-קליקה.

<u>משפט</u>: הסף להופעת $K_4$ ב-$G(n,p)$ הוא ב-$p=n^{-2/3}$.
כלומר אם $p \ll n^{-2/3}$ אז כמעט בוודאות אין ב-$G$ $K_4$
ואם $p \gg n^{-2/3}$ אז כמעט בוודאות $G$ מכיל $K_4$.

יש פה כמה ניסוחים שדורשים הסבר ופירמול. מה הכוונה "כמעט בוודאות"?
הכוונה היא שאם $p \ll n^{-2/3}$ אז $\lim_{n\to\infty} \Pr(G \supseteq K_4) = 0$,
כלומר מדובר פה בתהליך אסימפטוטי. ואילו אם $p \gg n^{-2/3}$
אז $\lim_{n\to\infty} \Pr(G \supseteq K_4) = 1$.
ומה זה "קטן מאוד" / "גדול מאוד"?
$p \ll n^{-2/3}$ פירושו $p = o(n^{-2/3})$ כלומר $\frac{p}{n^{-2/3}} \xrightarrow[n\to\infty]{} 0$
$p \gg n^{-2/3}$ פירושו $p = \omega(n^{-2/3})$ כלומר $\frac{p}{n^{-2/3}} \xrightarrow[n\to\infty]{} \infty$

<u>הוכחה</u>:

(1) יהי $p = o(n^{-2/3})$ ונגדיר מ"מ $X$ על המרחב $G(n,p)$
$X(G) =$ מספר ה-4 קליקות ב-$G$ אז

$$X = \sum_{\substack{|T|=4 \\ T\subseteq[n]}} X_T \qquad X_T = \begin{cases} 1 & \text{4-קליקה } T \\ 0 & \text{אחרת} \end{cases}$$

$$\Rightarrow \mathbb{E}X = \sum_{\substack{|T|=4 \\ T\subseteq[n]}} \mathbb{E}X_T = \binom{n}{4} \Pr\left(\substack{\{1,2,3,4\} \\ \text{קליקה}}\right) = \binom{n}{4} p^6 = \underset{\text{כי } p=o(n^{-\frac{2}{3}})}{=} o(n^4 \cdot n^{-\frac{2}{3}\cdot 6}) = o(1)$$

$\Leftarrow$ $\mathbb{E}X = o(1)$, כלומר $\lim_{n\to\infty} \mathbb{E}X = 0$. לכן $X$ מקבל ער

⑧ הערכים $0,1,2,\dots$ . נובע לכן ש- $\Pr(X=0)\xrightarrow[n\to\infty]{}1$
(כי אם ל-X נכון $\mathbb{E}X=\varepsilon$ אז $\Pr(X=0)\geq 1-\varepsilon$ ואז ניקח גבול).

② נגדיר כמקודם את X כמ"מ על המרחב $G(n,p)$ אשר מונה את מספר ה- $K_4$ בגרף. כפי שראינו

$$\mathbb{E}X=\binom{n}{4}p^6\xrightarrow[n\to\infty]{}\infty$$

(הפעם השאיפה ל-$\infty$ כי $p=\omega(n^{-2/3})$. אני רוצה להשתמש ב-(**) (המקרה הפרטי של צ'בישב) תוך שנראה ש-

$$\operatorname{Var}X=o((\mathbb{E}X)^2)$$

נזכור ש- $\operatorname{Var}X=\mathbb{E}X^2-(\mathbb{E}X)^2$. נתחיל: נסמן $X=\sum_{\substack{|T|=4\\T\subseteq[n]}}X_T$ כמקודם. אז

$$\mathbb{E}X^2=\mathbb{E}\Big(\big(\sum_{|T|=4}X_T\big)^2\Big)=\mathbb{E}\Big(\sum_{|T|=4}X_T^2+2\sum_{S\neq T}X_SX_T\Big)=$$

$$=\sum_{|T|=4}\mathbb{E}X_T^2+2\sum_{S\neq T}\mathbb{E}(X_SX_T)$$

$$(\mathbb{E}X)^2=\big(\sum\mathbb{E}X_T\big)^2=\sum_{|T|=4}(\mathbb{E}X_T)^2+2\sum_{S\neq T}\mathbb{E}X_S\,\mathbb{E}X_T$$

$$\Rightarrow\operatorname{Var}X=\mathbb{E}X^2-(\mathbb{E}X)^2=\sum_{|T|=4}\big[\mathbb{E}X_T^2-(\mathbb{E}X_T)^2\big]+$$

$$+2\sum_{S\neq T}\underbrace{\big[\mathbb{E}(X_SX_T)-\mathbb{E}X_S\,\mathbb{E}X_T\big]}_{=\operatorname{Cov}(X_S,X_T)}$$

וזה מתאפס אם $S\cap T=\emptyset$ כי אז
המשתנים בלתי תלויים ולכן $\mathbb{E}(X_SX_T)=\mathbb{E}X_S\mathbb{E}X_T$ .
ונשים לב שאפילו אם יש קודקוד 1 משותף אז המ"מ $X_S,X_T$ הם ב"ת.

המחובר השני בסכום שווה ל-

$$2\sum_{\substack{S,T\\|S\cap T|=2}}\operatorname{Cov}(X_S,X_T)+2\sum_{\substack{S,T\\|S\cap T|=3}}\operatorname{Cov}(X_S,X_T)$$

עתה מספיק להראות ש-

(א) $\sum_T\big(\mathbb{E}X_T^2-(\mathbb{E}X_T)^2\big)\ll(\mathbb{E}X)^2$

(ב) $\sum_{|S\cap T|=2}\mathbb{E}X_SX_T-\mathbb{E}X_S\mathbb{E}X_T\ll(\mathbb{E}X)^2$

(ג) $\sum_{|S\cap T|=3}\mathbb{E}X_SX_T-\mathbb{E}X_S\mathbb{E}X_T\ll(\mathbb{E}X)^3$

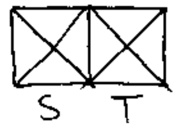
אז נראה.

(א) $X_T$ הוא מ"מ מציין ולכן $X_T^2=X_T$. לכן,

$$\sum_{|T|=4}\left(\mathbb{E}X_T^2-(\mathbb{E}X_T)^2\right) \le \sum_{|T|=4}\mathbb{E}X_T^2 = \sum_{|T|=4}\mathbb{E}X_T = \mathbb{E}X \longrightarrow \infty$$

$$\Rightarrow \frac{\sum\left[\mathbb{E}X_T^2-(\mathbb{E}X_T)^2\right]}{(\mathbb{E}X)^2} \le \frac{\mathbb{E}X}{(\mathbb{E}X)^2} = \frac{1}{\mathbb{E}X} \longrightarrow 0$$

(ב)



$$\sum_{|S\cap T|=2}\left[\mathbb{E}(X_SX_T)-\mathbb{E}X_S\mathbb{E}X_T\right] \le \sum_{|S\cap T|=2}\mathbb{E}X_SX_T =$$

$$= \binom{n}{4}\binom{4}{2}\binom{n-4}{2}p^{11} = \Theta(n^6p^{11}) \ll n^8p^{12}$$

קודקודי S, מי המשותפים, להשלמת T

כי $\frac{1}{n^2}<p$

$p=\omega(n^{-2/3})$

(ג)

באופן דומה לחלק (ב) מתקבל

$$\binom{n}{4}\binom{4}{3}\binom{n-4}{1}p^9 = \Theta(n^5p^8) \ll n^8p^{12}$$

$p=\omega(n^{-2/3})$ כי $\frac{1}{n}<p$

זורקים כדורים לתוך n כדים. יש כמה שאלות מעניינות.

① מתי צפוי שלראשונה יהיה בכד כלשהו יותר מכדור אחד? - $\sqrt{n}$

(פרדוקס יום ההולדת)

② מתי צפוי שלא יהיו יותר כדים ריקים? - $n\ln n$

(אספן הקופונים)

נניח שזורקים r כדורים ל-n כדים. מהי ההסתברות שכל שני כדורים נופלים לכדים שונים? - $1\cdot\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)$

לשני יש כד אחד שאסור לו להגיע אליו

לאחרון יש r-1 כדים שאסור לו להגיע אליהם

היינו רוצים להעריך את ההסתברות הנ"ל איכשהו. משהו שמאוד חשוב לזכור הוא ש- $e^{-x} \sim 1-x$ ולכן

$$1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{r-1}{n}\right)\sim$$

$$\sim e^{-\frac{1}{n}}\cdot e^{-\frac{2}{n}}\cdots e^{-\frac{r-1}{n}} = e^{-\sum_{j=1}^{r-1}\frac{j}{n}} = e^{-\frac{r(r-1)}{2n}}$$

לכן אם $r>>\sqrt{n}$ אז $Pr(\text{הכדורים בתאים שונים}) \xrightarrow[n\to\infty]{} 0$

ואילו אם $r<<\sqrt{n}$ אז $Pr(\text{הכדורים בתאים שונים}) \xrightarrow[n\to\infty]{} 1$

(המעבר הוא ב-$\sqrt{n}$ כי המונה מתנהג כמו $r^2$ והמכנה לינארי)

את השימוש של זה: יש אגם ובו דגים ורוצים להעריך כמה יש. אז נדוג דגים ונסמן אותם עד הפעם הראשונה שדגים דג שהוא כבר מסומן. נניח שזה קרה אחרי $n$ פעמים. אז הערכה טובה היא שיש $n^2$ דגים באגם...

<u>בעיה</u>: בוחרים מתוך עולם של $[n]$ שתי קבוצות באקראי:

$$|A|=k \qquad |B|=l$$

מהי ההסתברות ש- $A\cap B=\emptyset$ ?

<u>תשובה</u>: אם $n<<kl$ אז כמעט בוודאות $A\cap B\neq\emptyset$ ואם $kl<<n$ אז כמעט בוודאות $A\cap B=\emptyset$. ואילו אם $kl\sim n$ אז יש הסתברות חיובית הן לזרות והן לאי זרות.
(החישובים כאן נעשים באופן דומה לדוגמה הקודמת).

נחזור לבעיית אספן הקלפים. ראינו שאם $r>n(\ln n+\Delta)$ אז ההסתברות שתהיה עדיין איזשהו כד ריק היא $< e^{-\Delta}$

מהו מרחב ההסתברות שלנו? $\Omega=\{(n_1,n_2,\dots,n_r): 1\le n_i\le n\}$

ראינו המרחב הוא עוד סופי כי $r$ קבוע. אבל אם היינו זורקים כדורים עד שאין יותר כדים ריקים אז המרחב שלנו היה אינסופי - כי סדרת הזריקות עלולה להיות ארוכה כרצוננו.

יהי $X$ המ"מ המונה את מספר הכדים הריקים לאחר $r$ הטלות כדורים.

אפשר לרשום $X=\sum_{j=1}^{n}X_j$ כאשר $X_j=\begin{cases}1 & \text{הכד ה-}j\text{ ריק}\\ 0 & \text{אחרת}\end{cases}$

מאדיטיביות של התוחלת ומכך שלכל הכדים יש אותו מעמד נובע ש-

$$\mathbb{E}X = \sum_{j=1}^{n} \mathbb{E}X_j = n \cdot Pr\left(\text{הכד הראשון ריק לאחר r הטלות}\right)$$

נחשב אז:

$$Pr\left(\text{הכד הראשון עדיין ריק לאחר r הטלות}\right) = \left(1-\frac{1}{n}\right)^r < \left(1-\frac{1}{n}\right)^{n(\ln n+\Delta)} \leq$$

$$\leq e^{-(\ln n+\Delta)} = \frac{1}{ne^{\Delta}}$$

לכן, $\mathbb{E}X \leq n \cdot \frac{1}{ne^{\Delta}} = \frac{1}{e^{\Delta}}$. וכפי שכבר ראינו היום מספר פעמים (אופורמלית אפשר להשתמש במרקוב) נובע ש-

$$\cancel{Pr(X \geq 1)}$$

$$1 - Pr(X=0) = Pr(X \geq 1) \leq \mathbb{E}X$$

כלומר
$$Pr\left(\text{אין כדים ריקים}\right) \geq 1 - e^{-\Delta}$$

נראה עכשיו שתוחלת הזמן עד שלא יוותרו כדים ריקים היא

$$n \cdot H_n = n \cdot \sum_{k=1}^{n-1} \frac{1}{k} = n(\ln n + O(1))$$

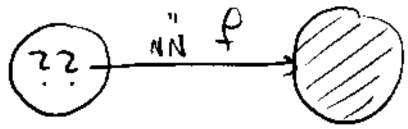
עכשיו מרחב ההסתברות שונה (שאינו מכוסה ע"י ההגדרה שנתנו - נדרש כאן קצת יותר) הוא המרחב של כל סדרות ההטלות $n_1, n_2, \ldots$ - נסמן ב- $Y_k$ את המ"מ המונה את מספר ההטלות עד שפגענו לראשונה בכד ה- $k$ (כאשר הכדים נספרים לפי סדר הופעתם - כלומר לפי סדר הפגיעה בהם)

נניח שיש $n$ כדים ומהם $k-1$ אינם ריקים. מהי ההסתברות לפגוע בכד חדש? - $p_k = \frac{n-k+1}{n}$. המ"מ המתאים הוא מ"מ מפולג גאומטרי עם $p_k = \frac{n-k+1}{n}$ ולכן תוחלת הזמן להצלחה של מ"מ כזה היא $\frac{1}{p_k}$ ולכן תוחלת זמן ההמתנה היא

$$\sum_{k=1}^{n} \frac{1}{p_k} = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \sum_{j=1}^{n} \frac{1}{j} = n H_n$$

10 18/11/08
טים מתמטיים

כדורים וכדים - אוסף הקופונים: ראינו בפעם הקודמת שתוחלת הזמן שיקח לאוסף הקופונים לאסוף את כל $n$ הקופונים השונים הוא $nH_n = n\sum_{k=1}^{n}\frac{1}{k} = n\ln n + O(n)$



כרגיל, בדיון בשאלות הסתברותיות יש לפנינו מ"מ $X$ והבעיה העיקרית היא להבין את התפלגות הערכים של $X$. כלומר, בהינתן $a<b$ ממשיים נרצה להבין/לחשב/להעריך את $Pr(a\le X\le b)$, או בהינתן $x\in\mathbb{R}$ את $Pr(X\ge x)$.

אז מה המאורע $a\le X\le b$? חשוב לזכור שתמיד מאחור עומד מרחב הסתברות, אז זה למעשה $\{x\in\Omega : a\le X(x)\le b\}\subseteq\Omega$

זה די נדיר שביכולתנו לענות על השאלה הזו בשלמות ולרוב אנו מסתפקים ב:

1. חישוב מומנטים (תוחלת, שונות, ...)
2. הערכות לגדלים הנ"ל

בבעיית אוסף הקופונים $\Omega=\{(n_1,n_2,\dots n_k): k\in\mathbb{N}, \forall i\ n_i\in[n]\}$

ואז במקרה של אוסף $\omega\in\Omega$ $X(\omega)=\min\{t: |\{n_1,\dots,n_t\}|=n\}$

למה אנחנו רוצים לדעת את $Pr(X\ge x)$. נניח שאנחנו באים לביטוח לאומי ואומרים לנו שזמן ההמתנה הממוצע הוא 10 דקות. זה נשמע טוב, אבל תמיד צריך לזכור את מקרה האגם עם העומק הממוצע 10 ס"מ. הרבה יותר מרגיע היה אילו היו אומרים לנו שהסיכוי שנחכה יותר מ- 12 דקות הוא קטן מ- $\frac{1}{1000}$...

ז"א, כרגיל אנו זקוקים למידע מפורט יותר מאשר התוחלת של $X$. זאת ניתן להשיג ע"י משפטי ריכוז מידה כגון אי-שוויון צ'בישב.

נעשה זאת עבור המ"מ של בעיית אספן הקופונים. נסמן $X=\sum_{i=1}^{n} X_i$ כאשר $X_i$ הוא משך הזמן שלוקח לנו לעבור ממצב של $i-1$ תאים לא ריקים למצב של $i$ תאים לא ריקים. כפי שראינו, $X_i$ הוא מ"מ בעל פילוג גיאומטרי עם פרמטר $p_i=\frac{n-i+1}{n}$ ולכן

$$\mathbb{E}X_i=\frac{1}{p_i}=\frac{n}{n-i+1} \quad \text{ו-} \quad \mathbb{E}X=\sum_{i=1}^{n}\mathbb{E}X_i .$$

נזכיר שלמ"מ מפולג גיאומטרית עם פרמטר $p$ יש שונות $\frac{1-p}{p^2}$ (נשים לב ש- $\frac{1-p}{p^2}\le\frac{1}{p^2}$) היות ש- $X_i$ הם מ"מ ב"ת, נקבל:

$$\sigma^2=\mathrm{Var}X=\sum \mathrm{Var}X_i \le \sum_{i=1}^{n}\frac{n^2}{(n-i+1)^2}=\sum_{k=1}^{n}\frac{n^2}{k^2}\le n^2\sum_{k=1}^{\infty}\frac{1}{k^2}=n^2\frac{\pi^2}{6}$$

אז $\sigma\le\frac{\pi}{\sqrt{6}}n$. מאי שוויון צ'בישב

$$\Pr(|X-\mu|>c\sigma)\le\frac{1}{c^2}$$

ולכן לכל $a$,

$$\Pr(|X-nH_n|>an)\le\frac{k}{a^2}$$

למשל, אפשר לקחת $a=\sqrt{\ln n}$ וזה נותן תוצאות מעניינות.

<u>אי שוויון צ'רנוף (הופדינג)</u>: יהיו $X_1,\dots,X_n$ מ"מ מציינים ב"ת. יהיו $X=\sum X_i$, $\mu=\mathbb{E}X$. אזי:

א) לכל $0<\delta$
$$\Pr(X>(1+\delta)\mu)<\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

ב) לכל $0<\delta<1$
$$\Pr(X<(1-\delta)\mu)<\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$

החסם כאן יותר טוב מצ'בישב אבל נשים לב שזה תחת הנחות מסוימות של המ"מ $X$. בצ'בישב מדובר על סתם מ"מ וכאן מדובר במקרה מיוחד שבו $X$ סכום של מ"מ מציינים ב"ת. אבל שכבר ראינו מספיק דוגמאות שבהן אפשר לפרק את $X$ באופן זה ואכן אי שוויון צ'רנוף יכול להיות יעיל מאוד!

מקרה פרטי חשוב: <u>מהלך מקרי על הישר</u>

יש חלקיק הנע על המספרים השלמים. הוא יוצא בזמן 0 מהנקודה 0 ומדי צעד הוא נע ימינה או שמאלה יחידה אחת, שניהם בהסתברות שווה $\frac{1}{2}$.

סוג השאלות שנרצה לברר:

1. היכן צפוי החלקיק להיות בזמן $N$ (גדול)?
2. כמה ביקורים צפוי שהוא יערוך בראשית עד זמן $N$?

③ מהו המקום הימני ביותר שאליו צפוי החלקיק להגיע במהלך N הצעדים הראשונים?

④ שאלות דומות גם בממדים גבוהים יותר.

⑤ מה קורה כשיש הטיה כלשהי?

נגדיר $X_1,\dots,X_N$ מ"מ ב"ת כך ש- $X_i$ מקבל את הערכים $\pm 1$ (צעדנו ימינה/שמאלה בזמן $i$). אז $X=\sum_{i=1}^{N} X_i$ הוא המקום של החלקיק בזמן $N$. כמה מהשאלות שהצגנו מתבטאות כבעיה בדבר ההתפלגות של $X$.

<u>טענה:</u> $\Pr(X>a)\le e^{-\frac{a^2}{2N}}$

במקרה הפרטי שאנו עוסקים בו ההתפלגות של המ"מ $X$ ידועה לנו במלואה. המאורע $X=t$ (בזמן $N$): אם הלכנו $\ell$ צעדים שמאלה ואת $N-\ell$ צעדים ימינה אז בזמן $N$ אנחנו נמצאים בנקודה $N-2\ell=t$. כלומר $\ell=\frac{N-t}{2}$. לכן

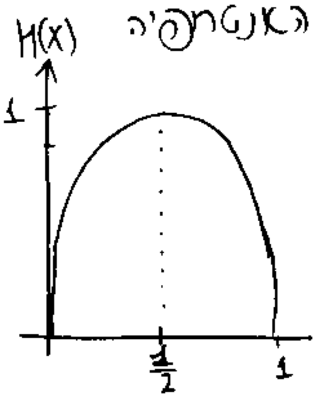
$$\Pr(X=t)=\frac{\binom{N}{\frac{N-t}{2}}}{2^N}$$

ז"א ההתפלגות של $X$ היא בינומית (אנו מתעלמים מבעיית הזוגיות/אי-זוגיות) (למשל אם $t=0$ אז ההסתברות להיות בראשית בזמן $N$ היא $\frac{\binom{N}{N/2}}{2^N}\approx\frac{c}{\sqrt{N}}$ (קירוב סטירלינג)) בפרט,

$$\Pr(X>a)=\frac{1}{2^N}\sum_{t>a}\binom{N}{\frac{N-t}{2}}$$

אז זאת נוסחה סגורה אלא שאנחנו לא מבינים אותה בכלל!!

<u>טענה מועילה:</u> לכל $0<\alpha<1$ $\binom{n}{\alpha n}=2^{n(H(\alpha)+o(1))}$ כאשר $H(x)=-x\log_2 x-(1-x)\log_2(1-x)$ פונקציית האנטרופיה




<u>הוכחת הטענה המועילה:</u> נניח כמובן ש- $\alpha n$ טבעי.

נראה ש- $$(*)\quad \frac{1}{n+1}\le\binom{n}{\alpha n}\alpha^{\alpha n}(1-\alpha)^{(1-\alpha)n}\le 1$$

נזכור ש- $1=(\alpha+(1-\alpha))^n=\sum_{k=0}^{n}\binom{n}{k}\alpha^k(1-\alpha)^{n-k}$

המחובר הגדול ביותר הוא המחובר שבו $k=\alpha n$ (כפי שנראה להלן) ולכן המסקנה נובעת:

בסכום $\sum_{k=0}^{n}\binom{n}{k}\alpha^k(1-\alpha)^{n-k}$ יש $n+1$ מחוברים ולכן המחובר הגדול ביותר $\binom{n}{\alpha n}\alpha^{\alpha n}(1-\alpha)^{(1-\alpha)n}$ הוא בין $1$ ל- $\frac{1}{n+1}$.

כלומר, ע"י העברת אגפים ב-(*) נקבל

$$\frac{1}{n+1}\cdot\frac{1}{\alpha^{\alpha n}(1-\alpha)^{(1-\alpha)n}} \le \binom{n}{\alpha n} \le \frac{1}{\alpha^{\alpha n}(1-\alpha)^{(1-\alpha)n}}$$

אבל $\frac{1}{\alpha^{\alpha n}(1-\alpha)^{(1-\alpha)n}} = \left(2^{H(\alpha)}\right)^n$ ולכן קיבלנו

בדיקה פשוטה מאוד

$$\frac{2^{nH(\alpha)}}{n+1} \le \binom{n}{\alpha n} \le 2^{nH(\alpha)}$$

$$2^{nH(\alpha)-\log_2(n+1)} = 2^{n\left(H(\alpha)-\underbrace{\frac{\log(n+1)}{n}}_{=o(1)}\right)}$$

$$\Rightarrow 2^{n(H(n)+o(1))} \le \binom{n}{\alpha n} \le 2^{nH(\alpha)}$$

זה מה שהיה צריך. נותר עוד להוכיח את ענין המחובר הגדול ביותר. מקום המחובר עבורו $k=\alpha n$ הוא הגדול ביותר בסכום $\sum_{k=0}^{n}\binom{n}{k}\alpha^k(1-\alpha)^{n-k}$ ? נבדוק מתי הסדרה עולה.

$$\binom{n}{j}\alpha^j(1-\alpha)^{n-j} \ge \binom{n}{j-1}\alpha^{j-1}(1-\alpha)^{n-j+1}$$

אמ"מ

$$\frac{n!}{j!(n-j)!}\,\alpha \ge \frac{n!}{(j-1)!(n-j+1)!}(1-\alpha)$$

אמ"מ $(n-j+1)\alpha \ge j(1-\alpha)$

אמ"מ $(n+1)\alpha - j\alpha \ge j(1-\alpha)$

אמ"מ $(n+1)\alpha \ge j$

כלומר האיברים בסכום עולים (ממש) עד המקום $j=(n+1)\alpha$ ואח"כ יורדים. טוב, יש פה איזו דקות עם ה- $\alpha n$ לעומת $(n+1)\alpha$, אבל סה"כ זה בסדר והטענה נכונה בכל זאת. ☺

הערכה $\binom{n}{\alpha n} = 2^{n(H(\alpha)+o(1))}$ מאפשרת לתת קירובים טובים ל- $\Pr(X>a)$ אבל לא נעשה זאת עכשיו. נראה אותה בדרך היותר מסורתית.

<u>הוכחת הטענה</u>: נשים לב שהמאורע $X\ge a$ הוא זהה למאורע $e^{tX}\ge e^{ta}$ לכל $0<t$ (כי $e^x$ מונוטונית עולה).

מאי שוויון מרקוב נובע ש-

$$\Pr(e^{tX} \ge e^{ta}) \le \frac{\mathbb{E}(e^{tX})}{e^{ta}}$$

נמצא את ה-$t$ שממזער את אגף ימין ונקבל חסם עליון טוב להסתברות של המאורע שלנו.

הפונקציה $e^{tX}$ נקראת הפונקציה יוצרת המומנטים (moment generating function) של $X$

$$\mathbb{E}e^{tX} = \mathbb{E}\left(1+tX+\frac{t^2X^2}{2!}+\frac{t^3X^3}{3!}+\dots\right) =$$
$$= 1+t\mathbb{E}X+\frac{t^2}{2!}\mathbb{E}X^2+\dots$$

את שאלות ההתכנסות אנחנו מטאטאים מתחת לשטיח

למה: לכל $t$ מתקיים: $\frac{e^t+e^{-t}}{2} \le e^{t^2/2}$

מראים ע"י פיתוח של האגפים לטור טיילור: שני האגפים הם פונקציות זוגיות והמקדמים באגף ימין גדולים מהמקדמים באגף שמאל.

$e^{tX_i}$ מקבל את הערכים $e^t, e^{-t}$ כל אחד בהסתברות $\frac{1}{2}$.

לכן $\mathbb{E}(e^{tX_i}) = \frac{e^t+e^{-t}}{2} \le e^{t^2/2}$ וכן

$$\mathbb{E}(e^{tX}) = \mathbb{E}\left(e^{t\sum_1^N X_i}\right) = \mathbb{E}\left(\prod_1^N e^{tX_i}\right) \underset{\text{ב"ת}}{=} \prod_1^N \mathbb{E}(e^{tX_i}) \le$$
$$\le \left(e^{t^2/2}\right)^N = e^{Nt^2/2}$$

אי שוויון מרקוב: אם $Z$ מ"מ אי שלילי אז לכל $0<c$

$\Pr(Z \ge c\,\mathbb{E}Z) \le \frac{1}{c}$ או לחילופין

$$\Pr(Z \ge M) \le \frac{\mathbb{E}Z}{M} \qquad \left(c := \frac{\mathbb{E}(Z)}{M}\right)$$

נפעיל את אי שוויון מרקוב על המ"מ האי שלילי $e^{tX}$:

$$\Pr(X \ge a) = \Pr(e^{tX} \ge e^{ta}) \le \frac{\mathbb{E}(e^{tX})}{e^{ta}} \le \frac{e^{Nt^2/2}}{e^{ta}}$$

כלומר

$$\Pr(X \ge a) \le e^{\frac{Nt^2}{2}-ta}$$

נרצה למזער את אגף ימין ע"י גזירה נקבל $Nt-a=0$

$\Leftarrow t=\frac{a}{N}$ ולכן

$$\Pr(X \ge a) \le e^{-a^2/2N}$$

<u>הוכחת אי שוויון צ'רנוף</u>: נוכיח את החלק הראשון והחלק השני אח"כ.

X הוא סכום של מ"מ ציינים ב"ת כך ש- $\mu = \mathbb{E}X$

צריך להראות שלכל $\delta>0$ מתקיים

$$\Pr(X > (1+\delta)\mu) \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \quad \left( \le e^{-\mu\frac{\delta^2}{3}} \right)$$

הטכניקה של ההוכחה תהיה דומה למקרה הפרטי.

ראו מקודם, נשתמש בהערכה $\Pr(X \ge a) \le \frac{\mathbb{E}(e^{tX})}{e^{ta}}$

נחשב חסם על $\mathbb{E}(e^{tX})$ ואז נמצער אותו ע"י בחירת $t$ מתאים.

ראו בהוכחה הקודמת - $\mathbb{E}(e^{tX}) = \prod \mathbb{E}(e^{tX_i})$

הסדרה $X_i$ מוגדרת בעצם ע"י סדרה $p_i$ כך ש-

$$\Pr(X_i=1)=p_i \qquad \Pr(X_i=0)=1-p_i$$

אז $e^{tX_i}$ מקבל את הערכים $e^t$ ו-$1$ בהסתברויות $p_i$ ו-$1-p_i$ בהתאמה. אז

$$\mathbb{E}(e^{tX_i}) = p_i e^t + (1-p_i) = 1+\underbrace{p_i(e^t-1)}_{x} \le e^{p_i(e^t-1)}$$

$e^x \ge 1+x$ לכל $x$

$$\Rightarrow \mathbb{E}(e^{tX}) = \prod \mathbb{E}e^{tX_i} \le e^{(e^t-1)\sum p_i} = e^{(e^t-1)\mu}$$

$$\Rightarrow \Pr(X \ge (1+\delta)\mu) = \Pr(X \ge a) \le \frac{\mathbb{E}(e^{tX})}{e^{t(1+\delta)\mu}} \le \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = e^{\mu(e^t-1-t(1+\delta))}$$

עכשיו אנחנו רוצים למצער את החסם הזה ע"י בחירת $t$. נגזור ונקבל $e^t = 1+\delta \Leftarrow t=\ln(1+\delta)$. הצבה נותנת:

$$\begin{aligned}\Pr(X \ge (1+\delta)\mu) &\le e^{\mu(e^{\ln(1+\delta)}-1-\ln(1+\delta)(1+\delta))} = \\ &= e^{\mu((1+\delta)-1-\ln(1+\delta)(1+\delta))} = \\ &= \left(e^{\delta-\ln(1+\delta)(1+\delta)}\right)^{\mu} = \\ &= \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}\end{aligned}$$

## משפט הגבול המרכזי: (Central Limit Theorem)

יהיו $X_1, \dots, X_n, \dots$ מ״מ iid (בת״ל ושווי התפלגות)
עם תוחלת $\mu$ ושונות $\sigma^2$. יהי $X = \sum_{i=1}^{N} X_i$. אזי לכל $a \in \mathbb{R}$

$$\lim_{N \to \infty} \Pr\left( \frac{X - N\mu}{\sqrt{N}\,\sigma} > a \right) = \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{t^2}{2}} \, dt$$

(שזו הזנב של ההתפלגות הנורמלית $N(0,1)$).

מתבקש לדעת הערכה של אחידות ההתכנסות. קיימות הערכות כאלה אבל זה מחוץ לתחום שלנו.

(14) 23.11.08
כלים מתמטיים

באי-שוויון צ'רנוף שהוכחנו בפעם הקודמת מופיע חסם מהצורה:

$$\Pr\left(X > (1+\delta)\mu\right) \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

כדי שזה מעניין בוודאי היינו רוצים לדעת שהחסם קטן מ-1.

$$(1+\delta)^{1+\delta} > e^{\delta} \quad \text{אמ"מ}$$

$$(1+\delta)\ln(1+\delta) > \delta$$

שני האגפים שווים ל-0 ב-$\delta=0$ אבל אם הנגזרת של אגף שמאל גדולה מזו של אגף ימין הטענה מוכחת לכל $\delta>0$

אז נגזור: $\ln(1+\delta)+1 \overset{?}{>} 1$

וזה בוודאי נכון כי $\delta>0$ ולכן $\delta+1>1$ ולכן $\ln(1+\delta)>0$.

אז עברנו sanity check - זה לא חסם מטופש. כעת נראה להעריך את הגודל הזה קצת יותר טוב.

נתבונן ב- $\ln\left(\frac{(1+\delta)^{1+\delta}}{e^{\delta}}\right)$. נפתח לפי טור טיילור

$$(1+\delta)\ln(1+\delta) - \delta = (1+\delta)\left(\delta - \frac{\delta^2}{2} + \frac{\delta^3}{3} - \dots\right) - \delta =$$

$$= \left(\delta - \frac{\delta^2}{2} + \frac{\delta^3}{3} - \dots\right) + \left(\delta^2 - \frac{\delta^3}{2} + \frac{\delta^4}{3} - \dots\right) - \delta =$$

$$= \frac{\delta^2}{2} - \frac{\delta^3}{6} + \dots$$

$$\Rightarrow \ln\left(\frac{(1+\delta)^{1+\delta}}{e^{\delta}}\right) = \frac{\delta^2}{2} + O(\delta^3)$$

$$\Rightarrow \frac{(1+\delta)^{1+\delta}}{e^{\delta}} = e^{\frac{\delta^2}{2} + O(\delta^3)}$$

בכמה מהשאלות שטופלו ראינו שהמושג של מרחב הסתברות בפועל כבר אינו מספק ואנו נזקקים לדיון מחודש בעצם ההגדרה של מרחב הסתברות.

עדיין יש לנו קבוצה $\Omega$ שהיא קבוצת מרחב ההסתברות. קודם לכל נקודה ולכל תת קבוצה של $\Omega$ נתנו הסתברות שהיא סכום ההסתברויות של הנקודות בקבוצה. אבל ממש במרחב לא בן-מנייה אי אפשר לעשות את זה. בכלל אי אפשר לדבר על סכומים.

במקרה הכללי, לא לכל תת קבוצה של $\Omega$ אנו מייחסים הסתברות.

יש אוסף $\mathcal{F}$ של תת קבוצות של $\Omega$, כלומר $\mathcal{F} \subseteq 2^{\Omega}$.

לקבוצה $A \in \mathcal{F}$ אני קוראים קבוצה מדידה (measurable)
האוסף $\mathcal{F}$ צריך להיות σ-אלגברה, כלומר הוא צריך לקיים את התנאים הבאים:

א. $\emptyset \in \mathcal{F}$

ב. אם $A \in \mathcal{F}$ אז $A^c \in \mathcal{F}$ (סגירות להשלמה)

ג. אם $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ סדרת קבוצות מדידות אז גם $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (סגירות לאיחוד בן-מניה)

מרחב מידה הוא שלשייה $(\Omega, \mathcal{F}, \mu)$ כאשר:

- $\Omega$ קבוצה
- $\mathcal{F}$ σ-אלגברה ב-$\Omega$
- $\mu$ פונקציית מידה, כלומר $\mu: \mathcal{F} \to \mathbb{R}$ כך ש-
  - $\mu(A) \geq 0$ לכל $A \in \mathcal{F}$
  - $\mu$ היא σ-אדטיבית, כלומר אם $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ סדרת קבוצות מדידות זרות בזוגות אז $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$
  (נשים לב ש-$\mu(\bigcup_{n=1}^{\infty} A_n)$ מוגדר היטב כי $\mathcal{F}$ היא סגורה לאיחוד בן-מניה)

במקרה המיוחד שבו $\mu(\Omega) = 1$ אנו קוראים ל-$(\Omega, \mathcal{F}, \mu)$ מרחב הסתברות. מרחב הסתברות בדיד הוא פשוט מקרה פרטי שבו $\Omega$ סופית ! $\mathcal{F} = 2^{\Omega}$.

דבר אפייני הוא ש-$\mathcal{F}$ אינה מתוארת בצורה מפורשת. אלא בוחרים בסיס $\mathcal{M}$ ו-$\mathcal{F}$ - אוסף של תת קבוצות של $\Omega$ - לכל $A \in \mathcal{M}$ מייחסים מידה $\mu(A)$. ואז $\mathcal{F}$ היא ה-σ-אלגברה הקטנה ביותר שמכילה את $\mathcal{M}$ = ה-σ-אלגברה הנוצרת ע"י $\mathcal{M}$ - היא החיתוך של כל ה-σ-אלגבראות של $\Omega$ אשר מכילות את $\mathcal{M}$ (קל לראות שחיתוך של σ-אלגבראות הוא σ-אלגברה וברור שחיתוך כזה הוא מינימאלי מבחינת הכלה).
הבעיה הטכנית היא איך לקבוע את $\mu(K)$ כאשר $K \in \mathcal{F}$ אך $K \notin \mathcal{M}$.

15

שיפורים:

מידה חיצונית $m^*(K) = \inf\{ \sum \mu(A_i) : A_i \in \mathcal{M}, \bigcup A_i \supseteq K \}$

מידה פנימית $m_*(K) = \sup\{ \sum \mu(A_i) : A_i \in \mathcal{M}, \bigcup A_i \subseteq K \}$.

ואז אם $m^*(K) = m_*(K)$ אז מגדירים [illegible] את הערך המשותף הזה להיות $\mu(K)$.

<u>דוגמה</u>: את $\Omega = [0,1]$ הופכים למרחב הסתברות ע"י $\mathcal{M} = \{(a,b) : a<b\}$ ; $\mu((a,b)) = b-a$

<u>דוגמה</u>: $\Omega = \{ (a_1, a_2, \dots) : 1 \le a_i \le k \}$

$\mathcal{M}$ = אוסף כל הצילינדרים ב-$\Omega$

(צילינדר הוא תת קבוצה של $\Omega$ אשר מוגדרת ע"י מספר סופי של תנאים מהסוג $a_{i_1} \in S_1, \dots, a_{i_t} \in S_t$)

לצילינדר כנ"ל מיחסים הסתברות $\prod_{j=1}^{t} \frac{|S_j|}{k}$

סיכום: בהרצאה סיסטמטית אנחנו נאלצים לוותר על היכולת להסתכל על כל תת הקבוצות של $\Omega$. במקרה כזה אפשר להסתכל על תת אוסף של $2^{\Omega}$ אבל הוא חייב להיות σ-אלגברה!

# אלגברה לינארית

תזכורת מהירה:

$\mathbb{F}$ שדה (השדות העיקריים בשימוש הם $\mathbb{F} = \mathbb{R}, \mathbb{C}$ וכן $\mathbb{F}$ שדה סופי).

מרחב וקטורי $V$ מעל השדה $\mathbb{F}$ זו קבוצה עם שתי פעולות: חיבור וקטורים וכפל בסקלר. $V$ הוא חבורה קומוטטיבית ביחס לחיבור. איברי $V$ נקראים וקטורים ואיברי $\mathbb{F}$ נקראים סקלרים. לכל $v \in V$ אם $\alpha \in \mathbb{F}$ מוגדר $\alpha v \in V$ ומתקיימים התנאים הבאים:

דיסטריבוטיביות $(\alpha+\beta)v = \alpha v + \beta v$, $\alpha(v+w) = \alpha v + \alpha w$

אסוציאטיביות $(\alpha\beta)v = \alpha(\beta v)$

שני סוגי דוגמאות חשובים במיוחד: $\mathbb{R}^n, \mathbb{C}^n$

$$V = \{ (x_1, \dots, x_n) : x_i \in F \}$$

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1+y_1, \dots, x_n+y_n)$$

$$\alpha (x_1, \dots, x_n) = (\alpha x_1, \dots, \alpha x_n)$$

אלו מרחבים ממימד $n$.

יש גם מרחבים אינסוף ממדיים חשובים מאוד בעולם המתמ',
למשל מרחבי פונקציות
- אוסף G הפולינומים הממשיים
- אוסף G הפונקציות הממשיות על הקטע $[0,1]$
המקיימות $\sup_{x\in[0,1]} f(x) < \infty$

בין מרחבים לינאריים מעניין להבין העתקות לינאריות: אם $U, V$ מרחבים לינאריים מעל אותו השדה, אז העתקה לינארית $T: U \to V$ מקיימת לכל $u_1, u_2 \in V$ ולכל $\alpha \in \mathbb{F}$

$$T(u_1+u_2) = Tu_1 + Tu_2$$

$$T(\alpha u_1) = \alpha T u_1$$

אומרים שוקטורים $v_1, \dots, v_n \in V$ הם תלויים לינארית אם יש סקלרים $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ שלא כולם אפס כך ש-

$$\alpha_1 v_1 + \dots + \alpha_n v_n = \sum \alpha_i v_i = 0$$

מגדירים את $\dim V$ כגודל המירבי של תת קבוצה בלתי תלויה לינארית ב-$V$.

במקרה ש- $U, V$ הם מרחבים בעלי מימד סופי, ניתן לתאר טרנספורמציה לינארית ביניהם בעזרת מטריצה. אם $\dim U = m$ ו- $\dim V = n$ אז העתקה לינארית מ- $U$ ל- $V$ מתוארת באמצעות מטריצה $A_{m\times n}$ ($m$ שורות ו- $n$ עמודות) כך ש- $Tx = xA$

$$(\dots x \dots)\left( A \right)_m^{\,n} = y \qquad y_i = \sum_j a_{ji} x_j$$

ערכים עצמיים - רוצים להראות שבהנחות קלות יש דרך נוחה מאוד להבין ולתאר טרנספורמציות לינאריות. התורה הספקטרלית (ספקטרום זה מילה מפוארת לאוסף ערכים העצמיים) מראה שבתנאים קלים ניתן לתאר טרנספורמציה לינארית רק: יש כיוונים מסויימים במרחב שבהם $T$ אינה אלא מתיחה (כפל בסקלר) ומזה ניתן להבין את ההתנהגות של $T$ בכלל. הכיוונים נקראים וקטורים עצמיים והסקלרים נקראים ערכים עצמיים.

אם $Ax_1 = \lambda_1 x_1, \dots, Ax_n = \lambda_n x_n$ קל לנו להבין איך מתנהגת ההעתקה הלינארית אם $x_1, \dots, x_n$ פורשים את $V$. בהינתן וקטור $x$ נציג אותו כצירוף לינארי $x = \sum \alpha_i x_i$ ואז

$$Ax = A(\sum \alpha_i x_i) = \sum \alpha_i A x_i = \sum \alpha_i \lambda_i x_i$$

אז זה מאוד נוח וזה התכון שלנו - למצוא בסיס של וקטורים עצמיים!

מקורסים באלגברה לינארית אנו מכירים את העובדה שיש מטריצות שבשבילן התהליך הנ"ל אינו אפשרי. למשל ל-$\begin{pmatrix}0&1\\0&0\end{pmatrix}$ אין בסיס של ו"ע, אבל האמת היא שזו לא בעיה מאוד חשובה. אפשר להתגבר על הקושי הזה כי המטריצות ה"טובות" הן צפופות במרחב של כל המטריצות.

יהיה מבחן בית

מדוע אין ל-$\begin{pmatrix}0&1\\0&0\end{pmatrix}$ בסיס של ו"ע? קל לוודא שהערך העצמי היחיד של המטריצה הזו הוא 0 בריבוי 2. מאידך, המרחב העצמי של 0 הוא חד מימדי (נפרש ע"י $(1,0)$) אבל לפי משפט מלינארית 2 אין בסיס של ו"ע.

איך מחשבים ע"ע? הדטרמיננטה $|A-\lambda I|$ היא פולינום ממעלה $n$ והע"ע של $A$ הם שורשי פולינום זה (נקרא הפולינום האופייני). במקרה שלנו

$$|\lambda I - A| = \begin{vmatrix}\lambda & -1\\ 0 & \lambda\end{vmatrix} = \lambda^2$$

לכל ע"ע $\lambda$ מתאים הריבוי האלגברי שלו כשורש של

הפולינום האופייני. הריבוי הגאומטרי הוא המימד של מרחב הו"ע עם ע"ע $\lambda$. תמיד ריבוי אלגברי $\geq$ ריבוי גאומטרי.
ושיש פגע את השיטה שתיארנו קודם אינה עובדת.
אבל התופעה הזאת (של אי שוויון בין הריבוי האלגברי והגאומטרי) היא נדירה. לחילופין, אפשר לחשוב על זה רק: אם נשנה באקראי ובמעט את $\varepsilon$ איברי המטריצה אז בהסתברות 1 נקבל מטריצה שאין לה הבעיה הזו. (נשים לב שזה לא אומר שבמציאות נקבל מטריצה לכסינה - כמו שהמידה של הרציונליים בקטע $[0,1]$ היא אפס ובכל זאת אם ניגרל מספר אקראי יכול להיות שנקבל מספר רציונלי).

בשימושים חשוב לנו הרבה לייחס לווקטורים ולמטריצות גודל.
איך אפשר למדוד את המרחק בין שתי נקודות ב-$\mathbb{R}^n$?
היינו רוצים לומר שהמרחק הוא ה"גודל" של $x-y$. לכן אני זקוקים לדרך שיטתית לייחס "גודל" לווקטורים.

אומרים ש- $\|\cdot\|: V \to \mathbb{R}$ היא נורמה של המרחב הווקטורי $V$ אם מתקיים:

- לכל $x \in V$ $\|x\| \geq 0$ ושוויון אמ"מ $x=0$
- אי שוויון המשולש: לכל $x,y \in V$ $\|x+y\| \leq \|x\|+\|y\|$
- הומוגניות: לכל $x \in V$, $\alpha \in \mathbb{F}$ $\|\alpha v\| = |\alpha|\|v\|$ $(\alpha \in \mathbb{R}, \mathbb{C})$

יש מבחר גדול של נורמות. למשל,

$L_1$ - $\|x\| = \Sigma |x_i|$

$L_\infty$ - $\|x\| = \max |x_i|$

$L_2$ - $\|x\| = \sqrt{\Sigma x_i^2}$

אלה הן שלוש הדוגמאות החשובות ביותר מתוך משפחה פרמטרית של נורמות הנקראות נורמות $L_p$:

$$\|x\|_{L_p} = \|x\|_p = \left(\Sigma |x_i|^p\right)^{1/p}$$

לא קשה להראות ש- $\lim_{p\to\infty} \|x\|_p = \max|x_i| = \|x\|_\infty$

באופן דומה אפשר להגדיר נורמות על מרחבי פונקציות :

$$\|f\|_2 = \sqrt{\int_0^1 f^2(x)\,dx}$$

$$\|f\|_\infty = \sup_{x\in[0,1]} f(x)$$

$$\|f\|_1 = \int_0^1 f(x)\,dx$$

$f:[0,1]\to\mathbb{R}$
אינטגרביליות

ומה לגבי "גודל" של מטריצות - נורמות מטריציוניות. יש שתי גישות שונות למדידת גודל של מטריצה.
משפחה אחת של נורמות של מטריצות היא כזו שבה אנו מתייחסים למטריצה $M\in A_{n\times m}$ כאל וקטור ב- $\mathbb{F}^{n\cdot m}$ ואז

$$\|A\|_{Frobenius} = \sqrt{\sum_{i,j} a_{ij}^2}$$

זו המקבילה לנורמת $L_2$. (משל, אפשר למדוד ככה כמה תמונות דומות זו לזו כמובן מפספס כמה דברים חשובים).
כשחושבים על מטריצה כמייצגת טרנספורמציה לינארית, נרצה לייחס לה גודל עם מידת המתיחה המירבית שהיא מותחת וקטורים.
כלומר, נניח $(X,\|\,\|_X)\xrightarrow{A}(Y,\|\,\|_Y)$ אז

$$\|A\|_{op} = \sup_{x\in X}\frac{\|Ax\|_Y}{\|x\|_X}$$

(18) 2/12/08
כלים מתמטיים

בפעם הקודמת ראינו איך ניתן לייחס "אורך" או "גודל" לווקטורים במרחב וקטורי. המושג המתאים הוא נורמה וקטורית. למרחב וקטורי המצויד גם בנורמה קוראים מרחב נורמי או (במקרה הסוף-ממדי) מרחב בנך (Banach).

מושג כללי עוד יותר ממושג הנורמה הוא מושג המטריקה. $(X,d)$ הוא מרחב מטרי אם $d: X\times X \to \mathbb{R}^+$ רק ש-

① $d$ סימטרית - $d(x,y)=d(y,x)$

② $d$ אי שלילית - $d(x,y)\ge 0$ ושוויון אמ"מ $x=y$

③ אי שוויון המשולש - $d(x,y)\le d(x,z)+d(z,y)$

<u>הגדרה</u>: יהי $X$ מרחב וקטורי. $\|\cdot\|$ היא נורמה של $X$ אם מתקיים:

① $\|\cdot\|$ אי שלילית - $\|x\|\ge 0$ ושוויון אמ"מ $x=0$

② לינאריות - לכל $\alpha\in F$ $\|\alpha x\|=|\alpha|\|x\|$

③ אי שוויון המשולש - $\|x+y\|\le\|x\|+\|y\|$

קל לראות שנורמה משרה מטריקה ע"י $d(x,y)=\|x-y\|$
ז"א אם נתונה נורמה $\|\cdot\|$ על $X$ ומגדירים $d(x,y)$ כנ"ל אז $d$ מטריקה של $X$.

דוגמאות חשובות למרחבים נורמיים:

• המרחב $\ell_p$: $X=\mathbb{R}^n$ והנורמה $\|x\|_p=\sqrt[p]{\sum_{i=1}^{n}|x_i|^p}$ $1\le p\le\infty$
(אנחנו לא נבדוק את מה שצריך, אבל זה נכון. ולמעשה האי-שליליות והלינאריות הן טריוויאליות. אי שוויון המשולש מצריך הוכחה. הוא לא טריוויאלי אבל גם לא מאוד קשה. אנחנו בכל אופן לא נוכיח)

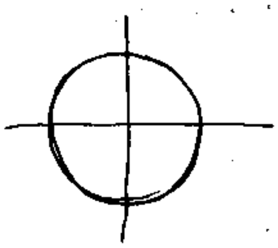
• המקרים החשובים ביותר של $\ell_p$ הם עבור $p=1,2,\infty$:

$$\|x\|_1=\sum|x_i|$$
$$\|x\|_2=\sqrt{\sum x_i^2}$$
$$\|x\|_\infty=\max|x_i|$$

• יש גם אנלוגים חשובים במרחבי פונקציות. למשל $X=L_2([0,1])$ (הכולל את כל הפונקציות הממשיות האינטגרביליות על הקטע $[0,1]$ עם הנורמה $\|f\|_2=\sqrt{\int_0^1 f^2(x)\,dx}$.

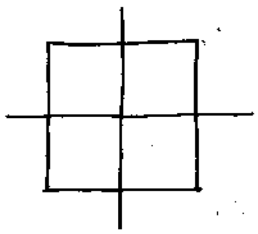
אם נורמה $\|\cdot\|$ מתאים כדור היחידה $B=\{x: \|x\|\le 1\}$
נראה שלא כי בלבד שהנורמה מגדירה את $B$, גם ניתן להסיק מהי $\|\cdot\|$ מתוך ידיעת $B$.

כדור היחידה של $\ell_2$:



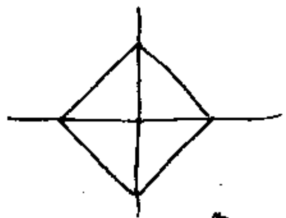
$B_2=\{x\in\mathbb{R}^n : \sqrt{\sum x_i^2}\le 1\}$

כדור היחידה של $\ell_\infty$:



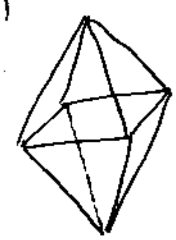
$B_\infty=\{x\in\mathbb{R}^n : -1\le x_i\le 1\}$

כדור היחידה של $\ell_1$:



$B_1=\{x\in\mathbb{R}^n : \sum|x_i|=1\}$

אוקטהדר
cross-polytope



ב- 3D זה
ויש לו 8 פאות.

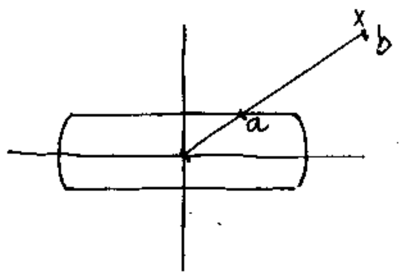
קוביה ב-$\mathbb{R}^n$ היא הקמור של $2^n$ נקודות (קודקודי הקוביה) $(\varepsilon_1,\dots,\varepsilon_n)$ כאשר $\varepsilon_i=\pm 1$.

לאוקטהדר יש רק $2n$ קודקודים והם $\pm e_i$ כאשר $e_i$ הוא וקטור היחידה הסטנדרטי בקואורדינטה ה-$i$.

נבונן בנורמה כלשהי $\|\cdot\|$ ובכדור היחידה המתאים לה $B_{\|\cdot\|}$.
זו קבוצה המכילה את הראשית, קמורה וסימטרית מרכזית.

<u>הגדרה</u>: אומרים שקבוצה $S\subseteq\mathbb{R}^n$ היא <u>קמורה</u> (convex)
אם לכל $x,y\in S$ ולכל $0\le\alpha\le 1$ מתקיים $\alpha x+(1-\alpha)y\in S$.
כלומר, הקטע שמחבר את $x$ ו-$y$ הוא מתוך $S$.
<u>קמור</u> (convex hull) של קבוצת נקודות $A$ הוא
הקבוצה הקמורה המינימלית שמכילה את $A$. למעשה זה
חיתוך כל הקבוצות הקמורות שמכילות את $A$ (קל לראות
שחיתוך של קמורות הוא קבוצה קמורה).
קבוצה $A$ היא <u>סימטרית מרכזית</u> אם לכל $x\in A$ גם $-x\in A$.

אז נניח שיש לנו את כדור היחידה שהוא קמור וסימטרי מרכזית,
אנחנו יודעים למדוד את הנורמה המקורית. כל נקודה שהיא על
שפת $B$ הנורמה שלה היא 1. שאר הנקודות מתקבלות ע"י מתיחה
מתאימה.




אם $t=\dfrac{\text{האורך של } b}{\text{האורך של } a}$ אז נגדיר
$\|x\|_B=t$. באופן פורמלי:

$$\|x\|_B=\inf\{t:\ \tfrac{x}{t}\in B\}$$

את הקמירות והסימטריות המרכזית צריך כדי שזה יהיה מוגדר היטב.
וכן הקמירות נותנת את אי שוויון המשולש והסימטריות נותנת את האי שליליות והליניאריות.

שניים ממושגי היסוד החשובים ביותר בעיסוק במרחבים נורמיים הם
המושגים של פונקציונל ליניארי ונורמה דואלית.

אם $X,Y$ מרחבים ליניאריים נורמיים ואם $A:X\to Y$ העתקה
ליניארית אז הגדרנו את הנורמה האופרטורית של $A$ ע"י

$$\|A\|_{op}=\sup_{x\in X}\frac{\|Ax\|_Y}{\|x\|_X}$$

יש סוג חשוב של העתקות ליניאריות ל-$\mathbb{R}$: $u: X\longrightarrow\mathbb{R}$, $x\longmapsto\langle x,u\rangle$
כאשר $u\in X$; $\langle\,,\rangle$ מ"פ.
הנורמה של $u$ נקראת <u>הנורמה הדואלית</u> והיא

$$\|u\|^*=\sup\frac{\langle u,x\rangle}{\|x\|}$$

פונקציונל לינארי הוא העתקה לינארית $f: X \to \mathbb{R}$. אם X סוף ממדי אז כל פונקציונל לינארי הוא מהצורה $u: X \to \mathbb{R}$, $x \mapsto \langle x,u\rangle$

אז אם $(X, \|\ \|)$ מרחב נורמי ממד סופי אז מוגדר המרחב הנורמי הדואלי $(X, \|\ \|^*)$ כאשר

$$\|u\|^* = \sup_{x\in X} \frac{\langle u,x\rangle}{\|x\|}$$

לא קשה לבדוק שאכן מתקיימות שלוש התכונות הנדרשות מנורמה.

<u>דוגמה</u>: מהי הנורמה הדואלית ל-$\ell_1^n$? אנחנו מתחילים ב-$(\mathbb{R}^n, \|\cdot\|_1)$. נרצה לחשב את $\|u\|^*$ של u.

$$\|u\|^* = \sup_{x\in\mathbb{R}^n} \frac{\langle u,x\rangle}{\|x\|_1} = \sup_{\|x\|_1=1}\langle u,x\rangle$$

כי גם המונה וגם המכנה לינאריים בכפולה של x בסקלר חיובי.

עכשיו צריך לחשב. נניח ש- $u=(u_1,\dots,u_n)$ נתון. כדי לחשב את $\|u\|^*$ עלינו לפתור את בעיית האופטימיזציה הבאה: מיצאו $x_1,\dots,x_n$ רק ש- $\sum|x_i| = 1$ וכך ש- $\sum x_iu_i$ יהיה מירבי. ברור שכדי ש- $\sum x_iu_i$ יהיה מירבי נרצה ש- $\operatorname{sgn}x_i = \operatorname{sgn}u_i$. לכן ניתן להניח ש-$u$ וקטור חיובי ואז גם $x$ בהכרח יהיה חיובי. ואז התנאי שלנו הוא ש- $\sum x_i = 1$. כמובן הבחירה הטובה ביותר היא לתת את הערך $x_j = 1$ ו- $x_i = 0$ לכל $i \neq j$ כאשר $u_j$ הוא הגדול מבין $u_i$.

מצאנו אם כך שהנורמה הדואלית לנורמת $\ell_1$ היא נורמת $\ell_\infty$ כי בבחירה כזו של $x$ מתקיים $\|x\|_1 = 1$ ; $\langle u,x\rangle = \|u\|_\infty$.

בדיוק באותו אופן הנורמה הדואלית לנורמת $\ell_\infty$ היא $\ell_1$.

<u>דוגמה</u>: מהי הנורמה הדואלית ל- $\ell_2$? באופן דומה, כאן עלינו לפתור את בעיית האופטימיזציה הבאה: בהינתן $u\in\mathbb{R}^n$, מצא וקטור יחידה $\|x\|_2=1$ רק ש- $\langle u,x\rangle$ מירבי.

הבחירה הטובה ביותר היא $x = \frac{u}{\|u\|_2}$ ואז $\langle u, x\rangle = \langle u, \frac{u}{\|u\|_2}\rangle = \frac{1}{\|u\|_2}\|u\|_2^2 = \|u\|$

למה זו הבחירה הטובה ביותר? באופן כללי $\langle s,t\rangle = \|s\|_2\|t\|_2\cos\varphi$

הזווית בין s ו-t

אם נרצה לבחור אותה רק ש- $\cos\varphi$ מקסימלי (רג במקרה שלנו $\|s\|=\|t\|=1$ ) אם נרצה $\varphi=0$.

באופן כללי ניתן להראות (ההוכחה נמצאת בסיכומים) שהמרחבים הנורמים $\ell_p$ ו- $\ell_q$ מדויק כאשר $\frac{1}{p}+\frac{1}{q}=1$. (גם $p=1$ ו- $q=\infty$ מקיימים תכונה זו). הדבר העיקרי בהוכחה ש- $(\ell_p)^*=\ell_q$ הוא אי שוויון Hölder שהוא הכללה של אי שוויון קושי שוורץ: לכל $x,y\in\mathbb{R}^n$ ולכל $p,q$ כך ש- $\frac{1}{p}+\frac{1}{q}=1$ מתקיים $\langle x,y\rangle \le \|x\|_p\|y\|_q$ (אי שוויון קושי-שוורץ מתקבל עבור $p=q=2$ )

ההיבט הגיאומטרי של הנורמה הדואלית: בהינתן נורמה $\|\|$ על $\mathbb{R}^n$ ראינו שכדור היחידה המתאים הוא $B_{\|\|}=\{x\in\mathbb{R}^n : \|x\|\le 1\}$ מעניין להבין מהו כדור היחידה של הנורמה הדואלית. לא קשה לוודא שזהו הגוף הקוטבי (Polar body) של $B$ המוגדר כך:

$$B^\circ=\{x\in\mathbb{R}^n : \forall y\in B \quad \langle x,y\rangle\le 1\}$$

למשל, הגוף הקוטבי של כדור היחידה של $\ell_2$ זה הוא עצמו. האוקטהדר והקובייה קוטביים זה לזה.

מה ניתן לומר על העתקה לינארית מרחב $\mathbb{R}^n,\mathbb{C}^n$ לעצמו שהם שומרת על נורמת $\ell_2$? העתקה $x\mapsto Ux$ היא כזו אמ"מ היא אורתוגונלית (במקרה של $\mathbb{R}^n$) או אוניטרית (במקרה של $\mathbb{C}^n$).

<u>הגדרה</u>: העתקה $\varphi$ בין מרחבים מטריים $(X,d)\xrightarrow{\varphi}(Y,\rho)$ נקראת <u>איזומטריה</u> אם היא שומרת מרחקים. כלומר לכל $x_1,x_2\in X$

$$d(x_1,x_2)=\rho(\varphi(x_1),\varphi(x_2))$$

<u>הגדרה</u>: אומרים שמטריצה ממשית $U$ היא אורתוגונלית אם $UU^t=I$ (ואז גם $U^tU=I$).

אם $U$ מרוכבת אומרים ש- $U$ אוניטרית אם $UU^*=I$ (כאשר $U^*=\bar{U}^t$) (ואז גם $U^*U=I$).

<u>טענה</u>: ההעתקה הלינארית $x \mapsto Ux$ היא איזומטריה של $\mathbb{R}^n$ אמ"מ $U$ אורתוגונלית.

<u>הערה</u>: אם $U$ איזומטריה לינארית של $\mathbb{R}^n$ אז $U$ שומרת גם על זוויות: $\|U(x+y)\|^2 = \|x+y\|^2 \quad \Leftarrow$

$$\|Ux\|^2+\|Uy\|^2+2\langle Ux,Uy\rangle = \|x\|^2+\|y\|^2+2\langle x,y\rangle$$

$$\Leftarrow \quad \langle Ux,Uy\rangle = \langle x,y\rangle$$

$$\Leftarrow \quad \|Ux\|\,\|Uy\|\cos(Ux,Uy) = \|x\|\,\|y\|\cos(x,y)$$

$\Leftarrow \cos(Ux,Uy)=\cos(x,y)$. עכשיו היינו רוצים להגיד שזה אומר שהזוויות גם שוות. אבל זה לא נובע באופן ישיר. יש פה איזה פגם אבל הטענה נכונה בכל אופן.

<u>הוכחת הטענה</u>: נניח ש-$U$ אורתוגונלית. אז

$$\|Ux\|^2 = \langle Ux,Ux\rangle = (Ux)^t Ux = x^t U^t U x = x^t I x = x^t x = \langle x,x\rangle = \|x\|^2$$

$$\Leftarrow \quad \|Ux\| = \|x\|,$$

מצד שני, אם $U$ איזומטריה אז

$$\|x\|_2^2 = \|Ux\|_2^2 = \langle Ux,Ux\rangle = \langle U^tUx, x\rangle \le$$

$$\underset{\text{קושי-שוורץ}}{\le} \|U^tUx\|_2\|x\|_2 \underset{\text{נורמה שמורה ע"י } U}{=} \|x\|_2^2$$

$\Leftarrow$ הא־שוויון הוא שוויון. אבל זה מתקיים רק כששני הווקטורים הם באותו כיוון. $\Leftarrow \quad U^tUx = \alpha x$ עבור $\alpha\in\mathbb{R}$. אבל

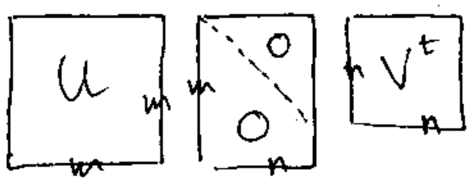
טוב, משהו כאן חורק. הטענה נכונה ואנחנו נחזור להוכחה בהמשך.

# פירוק ערכים סינגולריים SVD

תזכורת של ערכים עצמיים, לכסון וכו':

<u>משפט</u>: הפירוק הספקטרלי למטריצות סימטריות ממשיות:
לכל מטריצה סימטרית ממשית $A$ יש "פירוק" $A=VDV^t$
כאשר $V$ אורתוגונלית ו- $D$ אלכסונית. איברי $D$ נקבעים ביחידות.
השוויון הזה שקול לכך ש- $AV=VD$. העמודות של $V$ הן הע"ע
של $A$ ואיברי האלכסון של $D$ הם הע"ע המתאימים של $A$.

<u>משפט ה-SVD</u>: לכל $A\in M_{m\times n}(\mathbb{R})$ יש "פירוק" כ-
$A=UDV^t$ כאשר $U_{m\times m}$, $V_{n\times n}$ אורתוגונליות
ו- $D_{m\times n}$ "אלכסונית". כלומר אם למשל $m\ge n$ אז

$A =$ 

ב"אלכסון" של $D$ רשומים מספרים אי שליליים $\sigma_1\ge\sigma_2\ge\dots\ge\sigma_n\ge 0$
והמספרים אלה קרויים הערכים הסינגולריים של $A$ (Singular Values)

נשחק קצת עם המשפט:

$$AA^t = UDV^t\cdot VD^tU^t = U\underbrace{DD^t}U^t$$ (מטריצה סימטרית ומוגדרת אי-שלילית)

$$\begin{array}{cc} m & \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots \ \sigma_n^2 & \\ 0 & & 0 \end{pmatrix} \\ & m \end{array}$$

אז $AA^t$ דומה למטריצה אלכסונית, ו- $\sigma_1^2,\dots,\sigma_n^2$
הם ערכים עצמיים של המטריצה $AA^t$.
כלומר האיברים באלכסון של $D$ הם השורשים החיוביים של הע"ע
החיוביים של $AA^t$.

ל- SVD יש המון שימושים מעשיים.
אם המטריצה $A$ היא מערך של נתונים (תמונה למשל) אז נרצה במקרים רבים: ① לפתח שיטות לניקוי רעש
② לדחוס את הנתונים
מהינתן $A$ רוצים למצוא $B$ "פשוטה" המקרבת את $A$ היטב.
אז מה זה "פשוטה" ומה זה "קירוב טוב"?
"פשוטה" נוכל לממש ע"י רק שנדרוש ש- $rank(B)$ קטנה
וקירוב טוב פירושו ש- $\|A-B\|$ קטנה.
מסתבר ש- SVD במכה אחת פותר את שתי השאלות
האלה. שאנחנו יודעים לפתור בצורה יעילה שתי גרסאות קונקרטיות של השאלה:
$$\min_{rank(B)\leq k} \|A-B\|$$
עבור שתי נורמות: - הנורמה האופרטורית $\ell_2 \to \ell_2$
- פרובניוס

22 9/12/08
כום
מתמטיים

תזכורת: $(X,d)$ מרחב מטרי אם $X$ קבוצה ; $d: X\times X\to\mathbb{R}$ - מטריקה, כלומר:

א) $d$ חיובית: $d(x,y)\ge 0$ ושוויון אמ"מ $x=y$

ב) $d$ סימטרית: $d(x,y)=d(y,x)$

ג) אי שוויון המשולש: $d(x,y)\le d(x,z)+d(z,y)$

העתקה $\varphi:(X,d)\to(Y,\rho)$ בין מרחבים מטריים נקראת איזומטריה אם היא שומרת מרחקים, כלומר $\rho(\varphi(x_1),\varphi(x_2)) = d(x_1,x_2)$

מתעוררת השאלה: אילו העתקות לינאריות $T:X\to X$ כש-$X$ הוא המרחב הנורמי $\ell_2$ הן איזומטריות? כלומר, מה מאפיין את המטריצות $T$ רק ש- $\|Tx\|_2=\|x\|_2$ לכל $x\in X$.

טענה: הטרנספורמציה הלינארית המתאימה למטריצה $T_{n\times n}$ היא איזומטריה אמ"מ $T$ אורתוגונלית (כשעובדים מעל $\mathbb{R}$) או אוניטרית (כשעובדים מעל $\mathbb{C}$)

תזכורת: $A$ אורתוגונלית - $A^tA=AA^t=I$

$A$ אוניטרית - $A^*A=AA^*=I$ $\quad A^*=\bar{A}^t$

לכאורה יש שני תנאים לאורתוגונליות, אבל אם $A^tA=I$ כבר נובע אוטומטית ש-$AA^t=I$ ולהפך. העניין הוא שמטריצות אורתוגונליות הן הפיכות ומטריצה הופכי מימין הוא גם הופכי משמאל.

כמו כן, נשים לב שמה שכתוב במקום ה-$ij$ ב-$AA^t$ הוא מכפלה פנימית של השורה ה-$i$ והשורה ה-$j$. אז השוויון $AA^t=I$ אומר ממש שהשורות של $A$ הן בסיס אורתונורמלי למרחב ואילו השוויון $A^tA=I$ אומר את אותו הדבר לגבי העמודות... גם ההפך הוא כמובן נכון.
אם השורות (עמודות) של $A$ הן בסיס אורתונורמלי אז ברור ש-$A$ אורתוגונלית.

הוכחה: נראה למקרה האוניטרי. אצל $\mathbb{R}$ ההוכחה דומה.

נראה תחילה שאם $M$ אוניטרית אז ההעתקה הלינארית $x \mapsto Mx$ היא איזומטריה. ידוע ש- $\langle x, Qy\rangle = \langle Q^*x, y\rangle$. נזכור, $\langle x, x\rangle = \|x\|_2^2$. לכן, עלינו להוכיח ש- $\langle Mx, Mx\rangle = \langle x, x\rangle$.

ואכן - $\langle Mx, Mx\rangle = \langle M^*Mx, x\rangle = \langle Ix, x\rangle = \langle x, x\rangle$

לכן, אם $M$ אוניטרית אז $M$ איזומטריה.

להפך, רוצים להראות שאם $x \mapsto Mx$ איזומטריה אז $M$ אוניטרית.

נשים לב שאם $M$ הינו איזומטריה ב- $\ell_2$ אז היא שומרת לא רק אורכים, אלא גם זוויות. שהרי, $4\langle x,y\rangle = \|x+y\|_2^2 - \|x-y\|_2^2 =$

$$= \|M(x+y)\|_2^2 - \|M(x-y)\|_2^2 = \|Mx+My\|_2^2 - \|Mx-My\|_2^2 =$$

$$= 4\langle Mx, My\rangle$$

$\|x+y\|_2^2 - \|x-y\|_2^2 = 4\langle x,y\rangle$ קל לבדוק

אזי $\|Mx\|_2 = \|x\|_2$ ; $\|My\|_2 = \|y\|_2$ לכן גם הזווית שבין $x,y$ שווה לזווית שבין $Mx, My$.

כעת, אם $M$ איזומטריה אז לכל $x,y$

$$\langle x,y\rangle = \langle Mx, My\rangle = \langle x, M^*My\rangle$$

לכן $\langle x, y - M^*My\rangle = 0$. נקבע $y$. אם $y - M^*My \neq 0$ אז יש $x$ בשבילו המכפלה הפנימית הזו שונה מ-0 לכן $y = M^*My$ לכל $y$, כלומר $M^*M = I$

# פירוק ערכים סינגולריים

משפט 8: לכל מטריצה ממשית $A_{m\times n}$ יש הצגה $A = UDV^t$ כאשר $U_{m\times m}$, $V_{n\times n}$ אורתוגונליות! $D_{m\times n}$ "אלכסונית"

$$\left[\; D = {}_m\overbrace{\begin{pmatrix} \sigma_1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & \sigma_m & 0 \end{pmatrix}}^{n} \quad \text{או} \quad D = {}_m\overbrace{\begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \\ 0 & \cdots & 0 \end{pmatrix}}^{n} \;\right]$$

כאשר איברי האלכסון אי שליליים, מסודרים מהגדול לקטן ונקבעים ביחידות.

הערכים $\sigma_i$ נקראים הערכים הסינגולריים (singular values) של $A$.

ננסה להבין מהם הערכים הסינגולריים. נביט במטריצה $AA^t$ (זו מטריצה מוגדרת אי שלילית אם כל הערכים העצמיים שלה אי-שליליים)

אם יש ל-$A$ הצגה כנ"ל $A=UDV^t$ אז

$$AA^t = UDV^t(UDV^t)^t = U D \underbrace{V^t V}_{=I \text{ כי } V \text{ אורתוגונלי}} D^t U^t = UDD^tU^t = U\begin{pmatrix}\sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_n^2 & \\ & & & 0 \end{pmatrix}U^t$$

אז ל-$AA^t$ יש דמיון אורתוגונלי למטריצה $\begin{pmatrix}\sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_n^2 & \\ & & & 0\end{pmatrix}$ אך יש רק אחת כזו. לא קיבלנו את המסקנה שאם משפט ה-SVD נכון אז האיברים שב-$D$ הם השורשים החיוביים של הע"ע של $AA^t$.

אינטרפרטציה אפשרית אחרת למשפט $A=UDV^t$ - מהגדרת המכפלה בין מטריצות יוצא ש- $A=\sum_i \sigma_i u_i\otimes v_i$ (*) ($u_i\otimes v_i$ מטריצה מדרגה 1 [מטריצה $x\otimes y$ שבמקום ה-$ij$ שלה כתוב $x_iy_j$] $u_i,v_i$ הן העמודות ה-$i$ ב-$U,V$ בהתאמה).

המסקנות ממשפט ה-SVD שנוכיח אח"כ מראות שהקירוב הטוב ביותר ל-$A$ ע"י מטריצה מדרגה $\le k$ מתקבל ע"י קטיעת הסכום (*) במקום ה-$k$.

<u>רעיון ההוכחה</u>: יש פירוק $A=UDV^t$ כנ"ל אמ"מ $U^tAV=D$ אז אנחנו רוצים "ללכסן" את $A$. ואת המציאה של $U,V$ נעשה בשלבים. במתמטיקה שלבים זה אינדוקציה.

נחפש תחילה מטריצות אורתוגונליות $P,Q$ רק ש-

$$P^tAQ=\begin{pmatrix}\psi & 0 \cdots & 0\\ 0 & & \\ \vdots & B & \\ 0 & & \end{pmatrix}$$

ל-$B$ נוכל להפעיל את הנחת האינדוקציה ונמצא $R,S$ אם הן אורתוגונליות רק ש- $R^tP^tAQS=R^t\begin{pmatrix}\psi & 0\\ 0 & B\end{pmatrix}S$ "אלכסונית"

תזכורת: $\|A\|_{2\to 2} = \max_x \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2$

הוכחה: נגדיר $\sigma_1 = \|A\|_{2\to 2}$

ויהיו $u_1, v_1$ שני וקטורי יחידה כך ש- $Av_1 = \sigma_1 u_1$

(קיימים ראה מהגדרת הנורמה).

נשלים את $u_1$ לבסיס אורתונורמלי ונבנה את המטריצה האורתוגונלית $U = {}_n(\overset{1}{u_1} \mid \overset{n-1}{U_2})$. באותו אופן נבנה $V = {}_n(\overset{1}{v_1} \mid \overset{n-1}{V_2})$

נחשב: נעבוד בבלוקים:

$$U^tAV = \left(\frac{u_1^t}{U_2^t}\right)\left(A\right)\left(v_1 \mid V_2\right) =$$

$$= \left(\frac{u_1^t}{U_2^t}\right)\left(Av_1 \mid AV_2\right) = \left(\begin{array}{c|c} u_1^tAv_1 & u_1^tAV_2 \\ \hline U_2^tAv_1 & U_2^tAV_2 \end{array}\right) =$$

$$\underset{Av_1=\sigma_1u_1}{=} \left(\begin{array}{c|c} u_1^t\sigma_1u_1 & u_1^tAV_2 \\ \hline U_2^t\sigma_1u_1 & U_2^tAV_2 \end{array}\right) = \left(\begin{array}{c|c} \sigma_1 & u_1^tAV_2 \\ \hline 0 & \\ \vdots & U_2^tAV_2 \\ 0 & \end{array}\right)$$

כי $U$ מטריצה אורתוגונלית ולכן $u_1^tu_1 = 1$ וכל עמודה $u_j$ של $U_2$, $u_j^tu_1 = 0$.

ראינו אם כן ש- $UAV^t = \left(\begin{array}{c|c} \sigma_1 & w \\ \hline 0 & B \end{array}\right)$. מטרתנו הראשונה להראות ש- $w=0$. נרצה להראות שאם $w \neq 0$ אז יש וקטור אחר ש-$A$ מאריכה יותר מכדי $\sigma_1$. היות ש- $U, V$ אורתוגונליות, ל- $UAV^t$ ול- $A$ יש אותה נורמה אופרטורית. צ"ל שנרצה למצוא (בהנחה ש- $w \neq 0$) וקטור ש- $\begin{pmatrix} \sigma_1 & w \\ 0 & B \end{pmatrix}$ מאריכה פי יותר מ- $\sigma_1$: נביט בווקטור $(\sigma_1 \ w)$:

$$\left(\begin{array}{c|c} \sigma_1 & w \\ \hline 0 & B \end{array}\right)\begin{pmatrix} \sigma_1 \\ w \end{pmatrix} = \begin{pmatrix} \sigma_1^2 + \|w\|_2^2 \\ Bw^t \end{pmatrix}$$

$\Rightarrow \|UAV^t\| \geq \sqrt{\sigma_1^2 + \|w\|_2^2}$ סתירה. $\Rightarrow w = 0$.

(24)

וכן

$$U^t A V = \left(\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & B \end{array}\right)$$

אפשר להראות ש- $\|B\|_{2\to2} \le \sigma_1$

היות ש- B קטנה במימדיה מ- A ניתן להפעיל (באינדוקציה) עליה את המשפט שלנו. כלומר, $B = \tilde{U}_{m-1} \tilde{D} \tilde{V}^t_{n-1}$. לכן,

$$A = \underbrace{\left(u_1 \middle| U_2\right)}_{\substack{\text{אורתוגונלית} \\ \text{לפי הגדרה}}} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & \tilde{U}_{m-1} \end{pmatrix}}_{\substack{\text{אורתוגונלית} \\ m\times m}} \underbrace{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \tilde{D}_{(m-1)\times(n-1)} \end{pmatrix}}_{\text{״אלכסונית״}} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & \tilde{V}^t_{n-1} \end{pmatrix}}_{\substack{\text{אורתוגונלית} \\ n\times n}} \underbrace{\left(\begin{array}{c} v_1^t \\ \hline V_2^t \end{array}\right)}_{\substack{\text{אורתוגונלית} \\ \text{לפי הגדרה}}}$$

וזה מה שהיה צריך להוכיח.

מהי הנחת האינדוקציה? המקרה ש- A הוא וקטור. אז זה טריוויאלי.

את היחידות של D הראינו מכוח שהאלכסון הוא הע״ע של $AA^t$.

**משפט:** תהי $A_{m\times n}$ מטריצה ממשית ונחפש מטריצה B מדרגה $\text{rank}(B) \le k$ כך ש- $\|A-B\|_?$ מינימלי. ראשית במקום ? ניתן לרשום נורמה אופרטורית $2\to2$ וכן נורמת פרובניוס.

בשני המקרים B האופטימלית היא $B_{opt} = \sum_{i=1}^{k} \sigma_i u_i \otimes v_i$

ומתקיים

$$\|A - B_{opt}\|_{2\to2} = \sigma_{k+1}$$

$$\|A - B_{opt}\|_F = \sqrt{\sum_{j=k+1}^{n} \sigma_j^2}$$

<u>הערה</u>: לפי שראינו, לכל מטריצה M ולכל P, Q אורתוגונליות

$$\|M\|_{2\to2} = \|PMQ\|_{2\to2}$$

ומה בדבר נורמת פרובניוס? גם היא נשמרת בהכפלה במטריצה אורתוגונלית. $\|M\|_F^2 = \|MQ\|_F^2$. כדי להראות זאת נשים לב שלכל מטריצה A, $\|A\|_F^2 = \sum \|a_i\|_2^2$ כאשר $a_i$ השורה ה-i ב-A. לכן

$$\|M\|_F^2 = \sum \|m_i\|_2^2 \underset{\text{כי } Q \text{ איזומטריה}}{=} \sum \|Qm_i\|_2^2 = \|QM\|_F^2$$

באותו אופן מראים להכפלה מימין.

<u>הוכחת המשפט:</u> אם אנו נגדיר $B=\sum_i^k \sigma_i u_i\otimes v_i$ אז בוודאי $\|A-B\|_{2\to 2}=\sigma_{k+1}$ כי זהו הערך הסינגולרי הגדול ביותר של המטריצה $A-B$ (ראינו בהוכחת של SVD שערך זה הוא הנורמה האופרטורית $2\to 2$ של המטריצה המפורקת).

אז עלינו להראות שאם $\operatorname{rank} C\le k$ אז $\|A-C\|_{2\to 2}\ge \sigma_{k+1}$ נוכיח זאת. יהיו $v_1,\dots,v_n$ השורות של המטריצה $V$ (מהפירוק של $A$). מכיוון ש-$\operatorname{rank} C\le k$ אז לגרעין של $C$ יש חיתוך לא ריק עם $\operatorname{Span}\{v_1,\dots,v_{k+1}\}$ (נובע מספירת ממדים בעזרת משפט הממדים). אז יש וקטור יחידה $z$ בגרעין של $C$, כלומר $Cz=0$ ו- $z=\sum_{i=1}^{k+1}\alpha_i v_i$. נטען ש- $\|(A-C)z\|_2\ge\sigma_{k+1}$. זה חשבון די קל. נוכיח אותו:

$Cz=0 \Rightarrow (A-C)z=Az$. כעת,

$$(A-C)z=Az=\left(\sum_i^n \sigma_i u_i\otimes v_i\right)z=$$
$$=\sum_i^n \sigma_i u_i\langle v_i,z\rangle \underset{\downarrow}{=} \sum_{i=1}^{k+1}\sigma_i u_i\langle v_i,z\rangle$$

$z\in Sp(v_1,\dots,v_{k+1})$ אך המכפלה עם האחרים נותנת $0$ (כי $v_i$ אורתונורמליים)

$$\Rightarrow \|(A-C)z\|_2^2 \underset{\downarrow}{=} \sum_{i=1}^{k+1}\sigma_i^2\langle v_i,z\rangle^2 \ge$$

כי $\{u_i\}$ אורתונורמליים

$$\ge \sigma_{k+1}^2\sum_{i=1}^{k+1}\langle v_i,z\rangle^2 \underset{\downarrow}{=} \sigma_{k+1}^2\|z\|_2^2=\sigma_{k+1}^2$$

כי $z$ כאן מפותח בבסיס האורתונורמלי $v_1,\dots,v_{k+1}$

וזה בדיוק מה שרצינו. ☺

(25) 16/12/08
סים מתמטיים

<u>משפט ה- SVD</u>

לכל מטריצה ממשית $A_{m\times n}$ יש הצגה כ- $A=UDV^t$ כאשר
$U_{m\times m}$, $V_{n\times n}$ אורתוגונליות ! - $D_{m\times n}$ "אלכסונית". איברי $D$
הם $\sigma_1 \ge \dots \ge 0$ והם נקבעים ביחידות (ממש, רשומים השורשים החיוביים
של הערכים העצמיים של $AA^t$). הדרגה (rank) של $A$ היא
מס' ה- $\sigma_i$ החיוביים. כמו כן, $\min_{\operatorname{rank} B\le k}\|A-B\|$ מתקבל ע"י
$B=UD^{(k)}V^t$ כאשר $D^{(k)}$ מתקבלת מ- $D$ ע"י החלפת
המספרים $\sigma_{k+1},\dots,$ באפסים. הטענה הזו תקפה הן כש- $\|\cdot\|$
היא נורמת פרובניוס והן כשזו הנורמה האופרטורית $\ell_2\to\ell_2$. כמו כן,
$\sigma_1$ היא הנורמה האופרטורית של $A$.

אומרים שמטריצה ממשית סימטרית $A$ היא מוגדרת חיובית (מוגדרת
אי-שלילית) אם כל הערכים העצמיים שלה חיוביים (אי-שליליים).

מוגדרת חיובית PD - positive definite

מוגדרת אי-שלילית PSD - positive semi-definite

<u>משפט</u>: תהי $A$ מטריצה סימטרית ממשית. אזי התנאים הבאים שקולים:

(א) $A$ היא PSD (הערכים העצמיים אי-שליליים)

(ב) ניתן לרשום $A=MM^t$ לאיזו מטריצה ממשית $M$

(ג) לכל וקטור ממשי $x$ מתקיים $xAx^t\ge 0$.

<u>הוכחה:</u>

(א ⇐ ב) מהמשפט הספקטרלי למטריצות ממשיות סימטריות
ניתן לרשום $A=V\Lambda V^t$. אם נגדיר את המטריצה האלכסונית
$D$ כך $d_{ii}=\sqrt{\lambda_{ii}}$ (וזה מוגדר וממשי כי $\lambda_{ii}\ge 0$ היות ש-
$A$ PSD). אז ניקח $M=VD$ ונקבל $MM^t=V\underbrace{DD^t}_{\Lambda}V^t=A$

(ב ⇐ ג) אם $A=MM^t$ אז $xAx^t=xMM^tx^t=\|xM\|_2^2\ge 0$

(ג⇐א) יהי $x$ ו"ע $Ax^t=\lambda x^t$. נכפיל משמאל בווקטור $x$ ונקבל $\lambda\|x\|_2^2 = xAx^t$. עפ"י ההנחה $xAx^t\ge 0$ והרי $\|x\|_2^2\ge 0$ $\Rightarrow$ $\lambda\ge 0$.

☺

<u>ההיבט הגאומטרי</u> - נניח ש-$A$ מטריצה PSD. אז הקבוצה $\{x\in\mathbb{R}^n : xAx^t\le 1\}$ היא אליפסואיד שמרכזו בראשית. כיוון שעפ"י מה שראינו ניתן לרשום $A=MM^t$, את התנאי $xAx^t\le 1$ ניתן לרשום כתנאי $\|xM\|_2\le 1$.
לכן אליפסואיד שמרכזו בראשית הוא התמונה של כדור היחידה ע"י העתקה לינארית ואליפסואיד כללי הוא התמונה של כדור היחידה ע"י העתקה אפינית.

<u>הגדרה</u>: קונוס $n$-ממדי: זו תת קבוצה $C\subseteq\mathbb{R}^n$ כך ש-
(א) $C$ סגורה לחיבור: $x,y\in C \Rightarrow x+y\in C$
(ב) $C$ סגורה לכפל בסקלרים אי שליליים: $\alpha\in\mathbb{R}^{\ge 0},\ x\in C \Rightarrow \alpha x\in C$.

אנו נסמן ב-$PSD_n$ את הקונוס של מטריצות PSD $n\times n$. זה אכן קונוס. קל להיווכח שאם $A,B$ הן PSD ו- $\alpha\ge 0$ אז $\alpha A$, $A+B$ הן PSD. (בעזרת התנאי ש- $xMx^t\ge 0$).

$\varphi$ $x\mapsto xAx^t$ התבנית הריבועית של $A$

$(x,y)\mapsto xAy^t$ התבנית הבילינארית של $A$

<u>הגישה הווריאציונית לערכים עצמיים</u>: בהינתן מטריצה ממשית סימטרית מתבקש להבין מהו $\max_{\|x\|=1} xAx^t$ או $\min_{\|x\|=1} xAx^t$

מעניין גם להבין את הבעיה הזו. מהו הווקטור $x$ [illegible] מתחשבים.

**משפט Rayleigh-Ritz** תהי $A_{n\times n}$ מטריצה ממשית סימטרית.

יהיו $\lambda_1 \ge \dots \ge \lambda_n$ הע"ע של $A$ ו- $v_1, \dots, v_n$ וע"ע מתאימים. אורתוגונליים

אז $\lambda_1 = \max \frac{xAx^t}{\|x\|_2^2}$ (*) ובאופן כ':

$$\lambda_{k+1} = \max_{x \perp v_1, \dots, v_k} \frac{xAx^t}{\|x\|_2^2} \quad (**)$$

כמו כן, $\lambda_{t-1} = \min_{x \perp v_t, \dots, v_n} \frac{xAx^t}{\|x\|_2^2}$ , $\lambda_n = \min \frac{xAx^t}{\|x\|_2^2}$

> תמיד אפשר למצוא כאלה. אם $v_i, v_j$ מתאימים לע"ע שונים אז הם בהכרח אורתוגונליים. אחרת, אפשר לעשות גרם-שמידט.

**הוכחה**: נתחיל ב-(*). הוכחה זו ניתן להרחיב גם ל-(**) תוך שימוש בכופלי לגרנז' אבל לא נעשה זאת.

ניגש להוכחה ע"י חישוב $\frac{\partial}{\partial x_i} \frac{xAx^t}{\|x\|_2^2}$ ואיפוסה.

נזכור ש- $\left(\frac{f}{g}\right)' = 0$ אמ"מ $fg' = g'f$.

אז נחשב-

$$\|x\|_2^2 = \sum x_i^2 \Rightarrow \frac{\partial}{\partial x_i} \|x\|_2^2 = 2x_i$$

$$xAx^t = \sum a_{ij} x_i x_j \Rightarrow \frac{\partial}{\partial x_i}(xAx^t) = 2(Ax)_i$$

$\Leftarrow$ בכל נקודה קריטית מתקיים $x_i(xAx^t) = (Ax)_i \|x\|^2$

לכל $i$. ובכתיב וקטורי $x \underbrace{(xAx^t)}_{\in \mathbb{R}} = Ax \cdot \underbrace{\|x\|^2}_{\in \mathbb{R}}$

כלומר $x$ ו- $Ax$ פרופורציוניים, כלומר $x$ וע"ע. אז המקסימום (שקיים כי זו פונקציה רציפה ואנחנו מחפשים מקסימום בתחום קומפקטי) מתקבל עבור וע"ע $x$, כלומר $Ax = \lambda x$. לכן $\frac{xAx^t}{\|x\|^2} = \lambda$

והמקסימום הוא לכן עבור $\lambda_1 = \lambda$ (כי $\lambda_1$ הוא הע"ע המקסימלי)

ניתן להמשיך ולהוכיח את המשפט (**) בדרך דומה.

את $\lambda_2$ נמצא כך: נפתור בעזרת כופלי לגרנז' את הבעיה $\max_{x \perp v_1} \frac{xAx^t}{\|x\|^2}$ וכך הלאה.

נשתמש שוב במשפט הספקטרלי ונרשום $A = V\Lambda V^t$ , $V$ אורתוגונלית ו- $\Lambda$ אלכסונית.

$$xAx^t = xV\Lambda V^t x^t = \sum_i \lambda_i (xV)_i^2 \le$$

$$\max \lambda_i \cdot \sum (xV)_i^2 = \lambda_1 \|xV\|_2^2 = \lambda_1 \|x\|_2^2$$

כי $V$ אורתוגונלית.

כנ"ל לכל $x$, $xAx^t \le \lambda \|x\|_2^2$.
אבל אם $x=v_1$ אז שני האגפים שווים לזה $\lambda_1\|v_1\|^2 = v_1Av_1^t$.
לכן הוכחנו שלם את (*). נוכיח את (**).
נזכור שבהצגה $A=V\Lambda V^t$ העמודות של $V$ הן הו"ע של $A$, אורתונורמליים. אם $x\perp v_1,\dots,v_k$ אז $(xV)_i=0$ לכל $1\le i\le k$. ועכשיו,

$$xAx^t = \sum_i \lambda_i (xV)_i^2 = \sum_{i\ge k+1}\lambda_i (xV)_i^2 \le$$
$$\le \max_{i>k}\lambda_i \|xV\|_2^2 = \lambda_{k+1}\|x\|_2^2$$

זה מראה גם כן שאם $x\perp v_1,\dots,v_k$ אז $xAx^t \le \lambda_{k+1}\|x\|^2$
אבל אם ניקח $x=v_{k+1}$ נקבל שוויון:
כמובן, ממש באופן דומה נקבל את מה שצריך למינימום.

$$\lambda_{k+1} = \min_{x\perp v_k,\dots,v_n}\frac{xAx^t}{\|x\|^2} \qquad \lambda_n = \min\frac{xAx^t}{\|x\|^2}$$

<u>משפט Courant-Fischer</u>: תהי $A$ מטריצה ממשית סימטרית עם ע"ע $\lambda_1\ge\dots\ge\lambda_n$ אז

$$\lambda_{i+1} = \min_{\dim V=i}\ \max_{x\perp V}\frac{xAx^t}{\|x\|^2}$$

(כאשר המינימום נלקח על כל התת מרחבים $V$ כך ש- $\dim V=i$ -
זה יכול להיות לא נעים כי יש הרבה כאלה...)
ובאופן דומה,
$$\lambda_i = \max_{\dim W=n-i}\ \min_{x\perp W}\frac{xAx^t}{\|x\|^2}$$

<u>מסקנה: משפט השירוג</u> (interlacing)
תהי $A$ ממשית סימטרית עם ע"ע $\lambda_1\ge\dots\ge\lambda_n$ ותהי $B$ מטריצה המתקבלת מ-$A$ ע"י מחיקת השורה ה-$i$ והעמודה ה-$i$-ית שלה ($B$ ממשית סימטרית גם) ויהיו $\mu_1\ge\dots\ge\mu_{n-1}$ הע"ע של $B$. אז
$$\lambda_1\ge\mu_1\ge\lambda_2\ge\mu_2\ge\dots\ge\mu_{n-1}\ge\lambda_n$$

<u>כמה מילים על ההוכחה:</u>
$$\mu_1 = \max\frac{yBy^t}{\|y\|^2} = \max_{x_i=0}\frac{xAx^t}{\|x\|^2}$$
$$\lambda_1 = \max\frac{xAx^t}{\|x\|^2} \ \ge$$

ובאופן דומה לשאר.

(24)

# גרפים מרחיבים (Expander Graphs)

אלה גרפים ללא "צווארי בקבוק".

<u>הגדרה</u>: נאמר שגרף $G=(V,E)$ הוא $\varepsilon$-מרחיב-צלעות
(ε-edge-expanding) אם

$$\forall S\subseteq V \quad \left(|S|\le \frac{|V|}{2} \Rightarrow e(S,\bar{S})\ge \varepsilon|S|\right)$$

מספר הצלעות בין $S$ ל-$\bar{S}$.

מתברר שחישוב את $\varepsilon$ בהינתן $G$ זו בעיה קשה (co-NP complete). מאידך, ניתן להשיג הערכות של $\varepsilon$ בעלות חישוב ע"ז. מקובל לסמן ב- $h(G)$ את קבוע ההרחבה הצלעית של $G$, ר"ל הערך המירבי של $\varepsilon$ כך ש- $G$ $\varepsilon$-מרחיב-צלעית (קבוע cheeger). השאלה המעניינת העיקרית תהיה האם ניתן להשיג $h(G)$ גדול אף כי אין ב- $G$ יותר מדי צלעות.

ספציפית, ניתן להתעניין בגרפים $d$-רגולריים ולשאול: בהינתן $d\ge 3$ קבוע, האם יש $\varepsilon>0$ וגרפים $d$-רגולרים גדולים כרצוננו $G$ כך ש- $h(G)\ge\varepsilon$.

עם גרף ($d$-רגולרי) $G$ אנחנו מתאימים את מטריצת השכנות $A_G$ (זו מטריצה ממשית סימטרית). יהיו $\lambda_1\ge\lambda_2\ge\dots\ge\lambda_n$ הע"ע של המטריצה $A_G$.

היות ש- $G$ $d$-רגולרי קל לראות ש- $A_G\cdot\vec{1}=d\cdot\vec{1}$ לכן $d$ הוא ע"ע של $A$. ממשפט Perron Frobenius (שאותו נוכיח בשיעור הבא) נוכל לבדוק ש- $\lambda_1=d$. להפרש $\lambda_1-\lambda_2=d-\lambda_2$ קוראים <u>הפער הספקטרלי של $G$</u> וידוע שפער ספקטרלי והרחבה צלעית שקולים זה לזה במובן כלשהו.

<u>משפט</u>: יהי $G$ גרף $d$-רגולרי ותהי $A$ מטריצת השכנות של $G$. יהיו $d=\lambda_1\ge\lambda_2\ge\dots\ge\lambda_n$ הע"ע של $A$. אזי:

$$\frac{d-\lambda_2}{2} \overset{(*)}{\le} h(G) \overset{(**)}{\le} \sqrt{(d+\lambda_2)(d-\lambda_2)}$$

משמעות: חסמנו את $h(G)$ ע"י גדלים שאנחנו יודעים לחשב בצורה יעילה וזה טוב.

הוכחת (*): אפשר להתבונן על (*) כחסם תחתון של $\lambda_2$. אכן

$$\lambda_2 = \max_{x \perp \vec{1}} \frac{xAx^t}{\|x\|^2} = \max_{\sum x_i = 0} \frac{xAx^t}{\|x\|^2}$$

לכן ניתן באופן עקרוני להוכיח חסם תחתון על $\lambda_2$ ע"י הצגת וקטור $x$ המקיים $\sum x_i = 0$ וחישוב מנת ריילי שלו.

המשך שבוע הבא...

(28) 23|12|08

כלים מתמטיים

הוכחנו את האפיון הוריאציוני של ערכים עצמיים (משפט Rayleigh-Ritz): אם $A$ מטריצה ממשית סימטרית עם ע"ע $\lambda_1 \ge \dots \ge \lambda_n$ עם ו"ע $u_1,\dots,u_n$ אז לכל $1\le i\le n$

$$\lambda_i = \max_{x\perp u_1,\dots,u_{i-1}} \frac{xAx^t}{\|x\|^2}$$

אמרנו שגרף $G=(V,E)$ עם $n$ קודקודים הוא $\varepsilon$-מרחיב-צלעות אם לכל $S\subseteq V$ כך ש $|S|\le\frac{n}{2}$ מתקיים $e(S,\bar{S})\ge\varepsilon|S|$

ובהתאם מגדירים את $h(G)$ - קבוע ההרחבה הצלעית של $G$ - כ-

$$h(G)=\min_{\substack{S\subseteq V\\ |S|\le\frac{n}{2}}}\frac{e(S,\bar{S})}{|S|}$$

הערנו שאם $G$ גרף $d$-רגולרי ואם $A=A_G$ מטריצת השכנות של $G$ אז $d$ הוא ע"ע של $A$ עם ו"ע $(1,\dots,1)$. כפי שנראה (ע"פ משפט Perron-Frobenius) $d=\lambda_1$ הוא הע"ע הגדול ביותר של $A$. יהי אם כן $G$ כנ"ל והע"ע שלו

$$d=\lambda_1\ge\dots\ge\lambda_n$$

ההפרש $d-\lambda_2=\lambda_1-\lambda_2$ נקרא <u>הפער הספקטרלי</u> (Spectral gap) של $G$ ויש לו קשר הדוק למקדם ההרחבה הצלעית של $G$ כדלקמן:

<u>משפט</u>: יהי $G$ גרף $d$ רגולרי ויהיו $d=\lambda_1\ge\dots\ge\lambda_n$ הע"ע של $G$. אזי

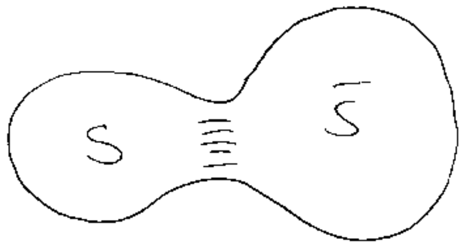
$$\frac{d-\lambda_2}{2}\overset{(*)}{\le}h(G)\overset{(**)}{\le}\sqrt{(d+\lambda_2)(d-\lambda_2)}$$

<u>הערה</u>: זה משפט טוב כי החישוב של $h(G)$ הוא בעיה co-NP-קשה ואילו את הערכים העצמיים אנחנו יכולים לחשב בלי בעיה אז המשפט נותן לנו איזה חסם על $h(G)$.

<u>הוכחת (*)</u>: נשים לב ש-(*) בעצם חוסם את $\lambda_2$ מלמטה. כדי לחסום את $\lambda_2$ מלמטה די (בגלל האפיון הוריאציוני) למצוא

וקטור $x$ רק ש- $x \perp \vec{1}$ (הו"ע הראשון $\vec{1}$) ורק ש- $\frac{xAx^t}{\|x\|^2}$ די גדול.

נניח מקבוצה $S \subseteq V$, $|S| \le \frac{n}{2}$ שבשבילה מושג הערך $h(G)$, כלומר $e(S,\bar{S}) = h(G)|S|$




$$xAx^t = \sum a_{ij}x_ix_j =$$

$$= \sum_{(i,j)\in E(G)} x_ix_j$$

כי A מטריצה של גרף

את זה אנחנו רוצים למקסם ⇐ נשאף שכל $x_i,x_j$ מאמפות היו שוני סימן. אי אפשר לקחת את כולם חיוביים כי אז לא יהיה $x \perp \vec{1}$. אז בחירה טבעית היא לדאוג שב-$S$ יהיו ערכים שליליים וב-$\bar{S}$ ערכים חיוביים. אבל צריך שהסכום יהיה $0$, אז לקואורדינטות של $S$ ניתן $-|\bar{S}|$ ולקואורדינטות של $\bar{S}$ ניתן $+|S|$:

$$x_i = \begin{cases} -|\bar{S}| & i \in S \\ |S| & i \in \bar{S} \end{cases}$$

בחירה כזו מבטיחה ש- $\langle x, \vec{1} \rangle = \sum x_i = 0$.

נחשב את מנת ריילי: נתחיל במכנה שהוא קל

$$\|x\|^2 = \sum x_i^2 = |S||\bar{S}|^2 + |\bar{S}||S|^2 =$$

$$= |S||\bar{S}|\left(|\bar{S}| + |S|\right) = n \cdot |S||\bar{S}|$$

המונה קצת יותר מסובך. יש שלושה סוגי צלעות - אלה שכולן מתוך $S$, אלה שכולן בתוך $\bar{S}$ ואלה שיש להן קודקוד ב-$S$ וקודקוד ב-$\bar{S}$.

$$(\star) \quad xAx^t = \sum_{(i,j)\in E} x_ix_j = 2\Big[\underbrace{e(S)|\bar{S}|^2}_{\text{כל } i,j\in S \text{ : } (i,j)\in E} + e(\bar{S})|S|^2 - \underbrace{e(S,\bar{S})}_{=h(G)|S|}|S||\bar{S}|\Big]$$

כי כל הזוגות (נספרים פעמיים)

יש $e(S)$ מחוברים כאלה והתרומה של כל אחד $|\bar{S}|^2$

G הוא d-רגולרי, לכן אם נסכם את הדרגות נקבל

$$d|S| = \sum_{v\in S} d(v) = 2e(S) + e(S,\bar{S})$$

כי צלעות בתוך S מוסיפות שתי דרגות ואילו את הצלעות שיוצאות מ-S צריך לספור רק פעם אחת.

באופן דומה $d|\bar{S}| = 2e(\bar{S}) + e(S,\bar{S})$

נציב ב-★ ונצמצם קצת ונקבל

$$xAx^t = d|S||\bar{S}|\cdot n - n^2|S|h(G)$$

כעת, לפי Rayleigh-Ritz נקבל ש-

$$\lambda_2 \geq \frac{xAx^t}{\|x\|^2} = \frac{d|S||\bar{S}|n - n^2|S|h(G)}{n|S||\bar{S}|} = d - \frac{nh(G)}{|\bar{S}|} \geq$$

$$\geq d - 2h(G)$$

כי $2|\bar{S}| \geq n$ שהרי $|\bar{S}|$ הוא הגדול בין השניים

מתברר שלמטריצות חיוביות (מטריצות שאיבריהן חיוביים) יש תכונות מעניינות.

<u>משפט Perron-Frobenius</u>: תהי $M_{n\times n}$ מטריצה ממשית שאיבריה אי-שליליים (זהו מטריצה אי-שלילית). נסמן ב-$\rho(M)$ את הרדיוס הספקטרלי של M: $\rho(M) = \max\{|\lambda| : \lambda \text{ ע"ע של } M\}$ (נשים לב שהע"ע יכולים להיות מרוכבים). ניחס למטריצה M גרף מכוון G באופן הבא: זהו גרף מכוון בעל n קודקודים שבו $(i,j)\in E(G)$ אמ"מ $m_{ij} > 0$.

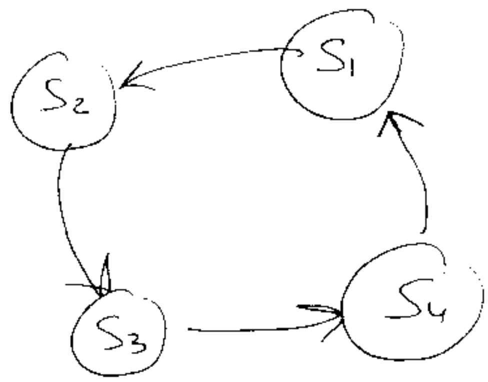
נניח ש-G קשיר חזק ובלתי מעגלי. אזי $\rho(M)$ הוא ע"ע פשוט של M והו"ע המתאימים לו חיוביים. אם $x_r, x_\ell$ הם הו"ע השמאלי והימני המתאימים לע"ע $\rho(M)$ אזי

$$\lim_{m\to\infty}\left(\frac{M}{\rho(M)}\right)^m = \frac{x_r \otimes x_\ell}{\langle x_r, x_\ell\rangle}$$

אנחנו לא נוכיח את המשפט המלא אלא רק את המקרה ש-M סימטרית. אבל קודם, נבין מה המשפט אומר.

קצת הסברים:

- גרף מכוון הוא קשיר חזק (strongly connected) אם לכל זוג סדור של קודקודים $(i,j)$ יש מסלול מכוון מ-$i$ ל-$j$
- גרף מכוון הוא מעגלי אם אפשר לחלק את קבוצת הקודקודים $V(G) = \dot{\bigcup}_{i=1}^{k} S_i$ כך שכל הצלעות ב-$G$ הולכות מקודקוד ב-$S_i$ לקודקוד ב-$S_{i+1}$ (החיבור מודולו $k$)




לפעמים קוראים לשני התנאים הללו (קשירות חזקה + אי-מעגליות) ארגודיות (מי שיש לו רקע בפיסיקה אולי יבין למה)

- ע"ע פשוט הוא שורש פשוט (בריבוי 1) של הפולינום האופייני
- גבול של מטריצה זה גבול איבר-איבר ומה שמתברר הוא שהגבול של המטריצות שם הוא מטריצה מדרגה 1 !

## שרשראות מרקוב

שרשרת מרקוב היא מערכת (חסרת זיכרון) עם מספר סופי של מצבים שהמעברים ביניהם נקבעים על סמך הכלל הפשוט הבא:
אם המערכת נמצאת במצב $i$ בזמן $t$ אז הסיכוי למצוא אותה בזמן $t+1$ במצב $j$ הוא $p_{ij}$

שרשרת מרקוב מתוארת אם כן באופן מלא ע"י מטריצה $P$ (הנקראת מטריצת המעברים של השרשרת transition matrix) כך ש- $P_{ij} = p_{ij}$. מטריצה כזו נקראת מטריצה סטוכסטית

$p_{ij} \geq 0$

$\forall i \; \sum_j p_{ij} = 1$

יש המון מערכות שמתוארות ככה.

קל לראות ש- $\rho(P)=1$ כי הווקטור $\vec{1}$ הוא ו"ע חיובי ימני
של P. לפי משפט PF יוצא ש- 1 הוא הע"ע הגדול ביותר
של P (לע"ע המקסימלי יש ו"ע חיוביים. אבל ו"ע שמתאימים לע"ע
שונים הם מאונכים ומאחר ש- $\vec{1}$ ו"ע לא יכול להיות עוד ו"ע חיובי
שמתאים ל- ע"ע גדול יותר).

<u>התפתחות בזמן של שרשרת מרקוב</u>: בזמן 0 אנחנו מניחים שיש
איזושהי התפלגות $\pi_0$ על מצב השרשרת. (למשל, אם מתחילים
ממצב i אז $\pi_0 = e_i$). $\pi_1$ זהו וקטור ההתפלגות בזמן 1, כלומר
הקואורדינטה ה-i של הווקטור $\pi_1$ היא ההסתברות להימצא במצב i
בזמן 1. מההגדרות נובע ש- $\pi_1 = \pi_0 P$. באופן דומה מגדירים
את $\pi_t$ כווקטור שהקואורדינטה ה-i שלו היא ההסתברות להימצא
במצב i בזמן t ומתקיים $\pi_t = \pi_0 P^t$. באופן טבעי
מעניין אותנו לדעת את ההתנהגות האסימפטוטית של השרשרת -
מהי ההתנהגות של השרשרת על פני פרקי זמן גדולים. כלומר,
האם קיים הגבול $\lim_{t\to\infty} \pi_t$ ואם כן, מהו. אבל

$$\lim_{t\to\infty} \pi_t = \lim_{t\to\infty} \pi_0 P^t = \pi_0 \lim_{t\to\infty} P^t$$

ומשפט PF בדיוק עונה לנו על השאלה הזאת כי $\rho(P)=1$. אז
רק צריך להבין מהם $x_r$ ו- $x_\ell$ - הו"ע השמאלי והימני המתאימים
לע"ע $1=\rho(P)$. את $x_r$ אנחנו יודעים - $x_r = \vec{1}$. לגבי $x_\ell$
אין נוסחה פשוטה, אבל הנקודה החשובה היא ש- $x_\ell = \pi_0 \cdot \lim_{t\to\infty} P^t$ אם
אם מנרמלים $\sum x_\ell^i = 1$
(וזה חשוב מהו $\pi_0$):

$$\pi_0 \cdot \lim_{t\to\infty} P^t \overset{PF}{=} \pi_0 \cdot \frac{x_r \otimes x_\ell}{\langle x_r, x_\ell \rangle} = \frac{\langle \pi_0, x_r \rangle \cdot x_\ell}{\langle x_r, x_\ell \rangle} = \frac{\sum \pi_0^i \cdot x_\ell}{\sum x_\ell^i} = \frac{1 \cdot x_\ell}{1} = x_\ell$$

הו"ע השמאלי המתאים לע"ע 1 נקרא גם התפלגות הגבול של
השרשרת וגם נקרא ההתפלגות הסטציונרית שלה. $x_\ell = P x_\ell$
כי $x_\ell$ היא (ה)התפלגות על מצבי השרשרת אשר איננה
משתנה במעבר של יחידת זמן.

נעיר עוד שאם P היא מטריצה סטוכסטית אז משפט PF
אומר ש- $\lim_{t\to\infty} P^t = \begin{pmatrix} — x_\ell — \\ \vdots \\ — x_\ell — \end{pmatrix}$

מקרה פרטי חשוב של שרשרת מרקוב הוא הילוך מקרי על קודקודים של גרף. אז המצבים הם קודקודי הגרף. והמעבר בין קודקוד לשכניו הוא בהתפלגות אחידה. במקרה זה התפלגות הגבול היא פרופורציונית לדרגות (תרגיל (חמד)) כלומר

$$\chi_e(u) = \frac{deg_G(u)}{2|E|}$$

פשוט צריך לבדוק שזה גם הוקטור שמאלי מתאים.

מעניין בהקשרים רבים לדעת לא רק מהו הגבול (מה שנותן לנו משפט PF) אלא גם מהי מהירות ההתכנסות. במקרה של הילוך מקרי על גרף רגולרי מתברר שקצב השאיפה לגבול נשלט ע"י הפער הספקטרלי. בפרט, בגרפים עם קבוע הרחבה גדול השאיפה היא המהירה ביותר האפשרית.

מהי דרישת הארגודיות במקרה של גרף לא מכוון?

| מכוון | | לא מכוון |
|---|---|---|
| קשירות חזקה | $\longrightarrow$ | קשירות |
| אפריודית | $\longrightarrow$ | דו-צדדיות |

. Perron הוכיח את המשפט כש-M חיובית; Frobenius- הראה איך להרחיב את הטיפול למקרה האי-שלילי.

איך מבצעים את הרדוקציה מהמקרה האי-שלילי למקרה החיובי? הרעיון הוא כזה בהינתן $M \geq 0$ וארגודית אם ידוע לנו שיש מספר טבעי $k$ כך ש- $M^k > 0$ אז ניתן לבצע את הרדוקציה הרדוקציה מפני ש-$u$ ו"ע של $M$ עם ע"ע $\lambda$ אמ"מ הוא ו"ע של $M^k$ עם ע"ע $\lambda^k$.

אז הם מצמצמים לטענה הבאה: אם $M \geq 0$ וארגודית אז יש $k$ כך ש- $M^k > 0$.

זוהי הטענה הקומבינטורית הבאה: אם $G$ גרף מכוון קשיר חזק אפריודי אז יש $k$ כך שלכל $i,j$ יש מסילה מכוונת ב-$G$ מהקודקוד ה-$i$ לקודקוד ה-$j$ שאורכה היא $k$.

יש להוכחה שני שלבים: נמצא ראשית מספר טבעי $t$ ככה שב-
$M^t$ כל איברי האלכסון חיוביים. ז"א, זה המקרה הפרטי $i=j$.
ז"א צריך להראות שיש $k$ כך שלכל קודקוד $i$ יש מסילה סגורה
באורך $k$ מ-$i$ ל-$i$. ניתן לרשום $M^t = D + Q$
כאשר $D$ אלכסונית חיובית ו-$Q$ מטריצה סימטרית עם אפסים
באלכסון. נתאים ל-$Q$ את הגרף $H$ + לולאות בכל קודקוד.
מעתה, כדי להוכיח ש- $(M^t)^l > 0$ לאיזה $l$ די להביט במהלכים
מאורך $\geq l$ בגרף $H$ כי הלולאות מאפשרות לנו להישאר
בקודקוד שבו היינו. וזה נכון מפני שבגרף קשיר יש לכל שני
קודקודים מסלה ביניהם באורך $\geq n$.
מכיוון ש-$G$ גרף ללא קודקודים מבודדים אז ב-$M^2$ כל איברי
האלכסון חיוביים כי אפשר לצאת מקודקוד על צלע ולחזור
אליו בחזרה בצעד הבא.

עוד כמה מילים על הוכחת PF במקרה ש-$M$ סימטרית:

<u>טענה</u>: תהי $M$ מטריצה ממשית סימטרית חיובית ויהי $v$ וע"ע המתאים
לע"ע הגדול ביותר של $M$. אזי $v \geq 0$.
<u>הוכחה</u>: ממשפט ה-SVD יוצא שהקירוב הטוב ביותר (בנורמת
Frobenius) למטריצה $M$ ע"י מטריצה מדרגה 1 היא
המטריצה $\lambda v \otimes v$. אילו היו ב-$v$ גם קואורדינטות שליליות,
נחליף את $v$ בווקטור $u$ המוגדר ע"י $u_i = |v_i|$ לכל $i$.
לפי המשפט $\|M - \lambda v \otimes v\|_F \leq \|M - |\lambda| u \otimes u\|_F$
$\Leftarrow$ $\sum (m_{ij} - \lambda v_i v_j)^2 \leq \sum (m_{ij} - |\lambda||v_i||v_j|)^2$
אבל $m_{ij} > 0$ ואם $(m-x)^2 \leq (m-|x|)^2$ לכל $x$
ובפרט כש- $i=j$ האי שוויון הוא חריף והמסקנה נובעת.

☺

(32) 30.12.08
כלים מתמטיים

ברמה האיכותית יש שקילות בין התכונות הבאות של גרף רגולרי:

$$\frac{d-\lambda_2}{2} \le h \le \sqrt{d^2-\lambda_2^2}$$

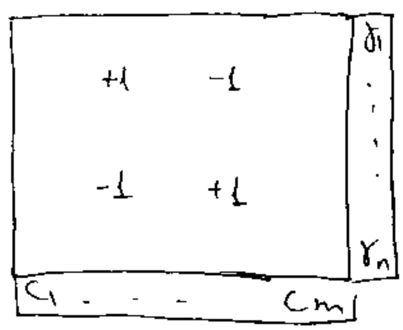
הרחבה
אין צווארי בקבוק

פער ספקטרלי
$\lambda_2$ מופרד מ- $d=\lambda_1$

המהלך המקרי על הגרף
מתכנס "מהר" להתפלגות
היציבה (=ההתפלגות האחידה)

למה זה טוב?

- דגימה ממרחב מאוד גדול: נבנה גרף (שקוונה שהוא מרחיב) שקודקודיו הם איברי המרחב שממנו נרצה לדגום. נריץ עליו מהלך מקרי לדגימה אחת או כמה מהלכים מקריים כדי לקבל כמה דגימות
- contigency table: יש טבלה של מספרים טבעיים $\ge 0$ אבל המספרים לא ידועים. מה שידוע זה סכומי השורות וסכומי העמודות. את המרחב שלנו הוא כל המטריצות מהצורה הזאת - איך אפשר לעשות מהלך מקרי על מטריצות אלה? (נגריל באקראי תת מטריצה 2x2 ואז נשנה את המספרים רק שהסכומים יישמרו:



חזרה להוכחת משפט Perron Frobenius למטריצות ממשיות סימטריות. המאפיינים העיקריים במשפט הם:

אם $M$ מטריצה אי-שלילית ארגודית אז הע"ע הגדול ביותר בערך מוחלט $\lambda$ הוא ממשי, מתאימים לו ו"ע חיוביים פשוטים $x_l, x_r$ ומתקיים $\left(\frac{M}{\lambda}\right)^k \xrightarrow[k\to\infty]{} \frac{x_r \otimes x_l}{\langle x_r, x_l \rangle}$

במקרה של המקרה הממשי הסימטרי הנ"ל הרי חשוב וחזק שיש לנו את המשפט הספקטרלי. ניתן לרשום $M = V \Lambda V$ כש-$V$ אורתוגונלית, $\Lambda$ אלכסונית, שורות $V$ הם הו"ע והערכים באלכסון של $\Lambda$ הם הע"ע של $M$.

אז $M^k = (V\Lambda V^t)\dots(V\Lambda V^t) = V\Lambda^k V^t$

$VV^t = I$ כי $V$ אורתוגונלית.

נניח שהאלכסון ב-$\Lambda$ מסודר מהגדול לקטן אז

$$\left(\frac{M}{\lambda}\right)^k = V\left(\frac{\Lambda}{\lambda}\right)^k V^t \xrightarrow[k\to\infty]{} V\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & & & \\ \vdots & & O & \\ 0 & & & \end{pmatrix} V^t$$

הפשטות היא חיונית: $\lambda$ הוא הע"ע הגדול ביותר והוא היחיד! לכן כשמחלקים את האלכסון של $\Lambda$ ב-$\lambda$ האיבר הראשון הוא 1 והשאר קטנים ממש מ-1. ולכן מתקיים $q^k \xrightarrow[k\to\infty]{} 0$ $.q<1$

<u>טענה</u>: תהי $M$ מטריצה ממשית סימטרית ותהי $v$ הו"ע המתאים לע"ע הגדול ביותר (בערך מוחלט) של $M$. אז $v \ge 0$, $\lambda \ge 0$.

<u>הוכחה</u>: ממשפט ה-SVD ידוע לנו שהקירוב הטוב ביותר בנורמת Frobenius ל-$M$ ע"י מטריצה מדרגה 1 הוא ע"י $\lambda v \otimes v$. אם יש ב-$v$ גם קואורדינטות שליליות, יהי $u$ הווקטור כך ש- $u_i = |v_i|$ לכל $i$. עפ"י ההנחה

$$\|M - \lambda v\otimes v\|_F \le \|M - |\lambda|\, u\otimes u\|_F^2$$

$$\Rightarrow \sum_{ij}(m_{ij} - \lambda v_i v_j)^2 \le \sum_{ij}(m_{ij} - |\lambda||v_i||v_j|)^2$$

אנו נבדוק מאיבר כלשהו בסכום: הרי

$$|m_{ij} - \lambda v_i v_j|^2 \geq (m_{ij} - |\lambda||v_i||v_j|)^2$$

נסמן את $|\lambda||v_i||v_j| = a$. אז $\lambda v_i v_j = \pm a$. אם
$\lambda v_i v_j = a$ אז יש שוויון וזה בסדר. אבל אם $m$
חיובי ו- $a$ חיובי אז תמיד $(m+a)^2 \geq (m-a)^2$.
מהסתכלות במקרה $i=j$, $v_iv_j = |v_i||v_j|$ נובע ש- $\lambda \geq 0$.

<u>משפט</u> (המקרה הסימטרי חיובי ממשי של משפט PF):
תהי $M$ מטריצה ממשית סימטרית וחיובית. יהיה $\lambda$ הע"ע
הגדול ביותר ו- $v$ הו"ע המתאים. אם $\lambda$ פשוט, $v>0$
ו: $\left(\frac{M}{\lambda}\right)^k \xrightarrow[k\to\infty]{} v\times v$

<u>הוכחה</u>: יש כאן שתי טענות שעדיין לא הוכחנו:

(א) $v>0$ (למעשה $v\geq 0$): שוב, לפי משפט ה-SVD ידוע
לנו ש- $\lambda v\otimes v$ הוא הקירוב הטוב ביותר ל- $M$ מדרגה
1. אם $v_i=0$ נגדיר $u = v+\varepsilon e_i$. יש ערך חיובי
קטן מספיק של $\varepsilon$ שבשבילו $\lambda u\otimes u$ קירוב טוב יותר של
$M$ מאשר $\lambda v\otimes v$.

(ב) $\lambda_n > -\lambda$: כאן נשתמש בעיקרון הווריאציוני.

$$\lambda = \max_{\|x\|=1} xMx^t =$$

$$\lambda_n = \min_{\|y\|=1} yMy^t$$

נניח כי $y$ המשיג את $\lambda_n$ ונגדיר $x_i=|y_i|$ לכל $i$.
כדי להראות ש- $\sum m_{ij}x_ix_j + \sum m_{ij}y_iy_j > 0$. ואכן,
כל מחובר בסכום $\sum m_{ij}(x_ix_j+y_iy_j)$ הוא אי שלילי וכש- $i=j$
הוא חיובי.

# אופטימיזציה

בעיית אופטימיזציה: יש קבוצה $D$ (תחום ההגדרה של הבעיה)
ופונקציה $f: D \to \mathbb{R}$. מחפשים את $\max_{x\in D} f(x)$ (או $\min_{x\in D} f(x)$)
(פונקציית המטרה, objective function)

מקרה בסיסי במיוחד של הבעיה הכללית הוא כזה: $D \subseteq \mathbb{R}^n$
ומוגדרת ע"י משוואות ואי-שוויונים. $f$ גם היא פונקציה לינארית,
כלומר יש וקטור $c$ כך ש- $f(x)=\langle x,c\rangle$ (למקרה הפרטי
הזה קוראים בעיית תכנון לינארי (Linear Programming = LP)

פוליהדר (polyhedron) זוהי תת קבוצה של $\mathbb{R}^n$
המוגדרת ע"י משוואות לינאריות ואי שוויונים לינאריים.

אם $D$ פוליהדר לא חסום אז קורה אחד מן השניים:
(א) פונקציית המטרה אינה חסומה על $D$ (וזה לא מעניין)
(ב) יש פוליהדר אחר $D' \subseteq D$ חסום כך ש- $\max_{x\in D} f(x) = \max_{x\in D'} f(x)$

לכן רוב העניין שלנו יהיה בפוליהדרים חסומים. לפוליהדר חסום
קוראים פאון (polytop). לרוב בעיות ה-LP שבהן
נתעניין תהיינה מהצורה $\max_{x\in D} f(x)$ כש- $f$
לינארית ו- $D$ פאון.

<u>בעיית התזונה של הפרה:</u>
יש כמה סוגי אוכל שבהם ניתן להזין פרות: תירס, קש, תחמיץ,
כוספה, ...
יש כמה אבות מזון הדרושים לבריאות הפרה: חלבונים,
שומן, ויטמינים, ברזל, ...
לכל דבר יש יחידות מוגדרות משלו (למשל תירס נספר בחבילות,
ויטמינים במ"ג וכו').

נרשום את כל האינפורמציה בטבלה:

| | חלבונים | שומן | ויטמינים | ... |
|---|---|---|---|---|
| תזיר | | | | |
| קש | | | | |
| תחמיץ | | | | |
| ⋮ | | | | |

m סוגי אוכל

n אבות מזון

$a_{ij}$ = מספר היחידות של אב מזון j ביחידת אוכל אחת מטיפוס i.

כדי להבטיח את בריאותה של הפרה, יש לספק לה לפחות $b_j$ יחידות של אב המזון ה-j לכל j. נסמן ב-b את הווקטור שהקואורדינטה ה-j שלו הוא צריכת המינימום הנ"ל מאב מזון j.

הווקטור c מוגדר כך ש-$c_i$ הוא המחיר של יחידת אוכל אחת מסוג i.

<u>הבעיה:</u> איך להזין במחיר מזערי פרה כך שצרכי התזונה שלה יסופקו?

ננסה לבנות לפרה "תפריט הזנה". נחפש וקטור x כך ש-$x_i$ מציין כמה יחידות של סוג אוכל i ניתן לפרה.

הטבלה היא $A_{m\times n}$

$c\ ,\ x\in\mathbb{R}^m$

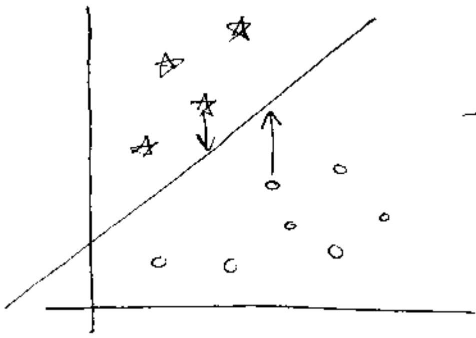
$b\in\mathbb{R}^n$

$\Leftarrow$ הבעיה היא מזעור המחיר:

$$\min \langle x,c\rangle$$
s.t.
$x\geq 0 \leftarrow$ אנחנו נותנים לפרה ולא לוקחים ממנה.
$$\underbrace{xA\geq b}_{\text{דרישות המינימום התזונתיות}}$$

בעיות שהן בעיות LP, אך כי לא רואות זאת במבט ראשון:

- נניח שנתונות לנו במישור שתי קבוצות סופיות של נקודות "כחולות" ו"אדומות". היינו רוצים, אם אפשר, להפריד את שתי הקבוצות ע"י קו ישר וכן להגדיל את "שולי הביטחון" ככל האפשר.

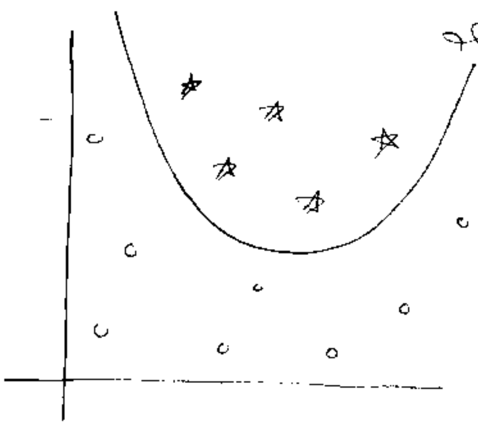


ישר מוגדר כמובן ע"י $y=ax+b$ .
יהיו $(u_1,v_1),\dots,(u_k,v_k)$ הכוכבים
$(\alpha_1,\beta_1),\dots,(\alpha_l,\beta_l)$ העיגולים

הבעיה שנפתור:

$$\max \varepsilon$$
$\forall i\ \ v_i \geq au_i+b+\varepsilon$ — ה-$y$ של הכוכבים $\varepsilon$ מעל הישר
$\forall j\ \ \beta_j \leq a\alpha_j+b-\varepsilon$ — ה-$y$ של העיגולים $\varepsilon$ מתחת לישר
$$\varepsilon \geq 0$$

כאן המשתנים הם למעשה $a$ ו-$b$ ו-$\varepsilon$.



אבל לא תמיד יש הפרדה ע"י ישר. אולי אפשר להפריד ע"י פרבולה.
גם כאן ניתן לנסח את הבעיה כ-LP. המשתנים הם $a,b,c,\varepsilon$ והבעיה לינארית במשתנים אלה.

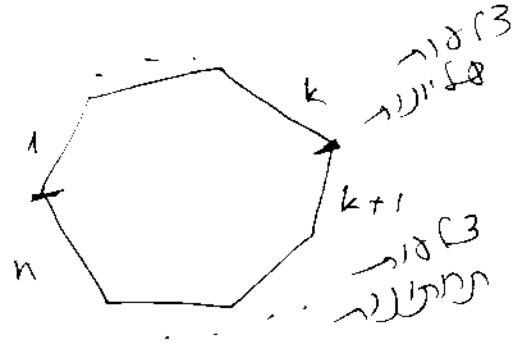
$$\max \varepsilon$$
$$\forall i\ \ v_i \geq au_i^2+bu_i+c+\varepsilon$$
$$\forall j\ \ \beta_j \leq a\alpha_j^2+b\alpha_j+c-\varepsilon$$

- נתון מצולע במישור. מחפשים את המעגל ברדיוס הגדול ביותר המוכל במצולע קמור.

תזכורת מגאומטריה אנליטית: אם $l$ הוא הישר $y=ax+b$ אז מרחקה של הנקודה $(x_0,y_0)$ מ-$l$ הוא

$$\pm\frac{y_0-ax_0-b}{\sqrt{1+a^2}}$$ הסימן נקבע עפ"י אם הנקודה מעל/מתחת לישר.

נמצא את המעגל המוכל ברדיוס מרבי. הרעיון: נמצא נקודה בתוך המצולע שמרחקה מספיק מכל אחת מן הצלעות שבמצולע.
נרשום המשתנים יהיו $u,v,R$ כאשר $(u,v)$ המרכז ו-$R$ הרדיוס.




$\max R$

$\forall i=1,\dots,k \quad R \le \frac{(v-a_i u-b_i)}{\sqrt{1+a_i^2}}$

$\forall j=k+1,\dots,n \quad -R \ge \frac{v-a_i u-b_j}{\sqrt{1+a_i^2}}$

## תוכניות ליניאריות בשלמים

ראינו כבר ש- 
$$\min\langle c,x\rangle \quad x\ge 0,\ Ax\ge b$$
היא LP.

לפעמים חשוב לנו שהקואורדינטות ב-x יהיו מספרים שלמים (או אפילו רק 0/1). לבעיה זו קוראים תכנון ליניארי בשלמים (ILP) והינו NP-קשה.

למשל, ידוע ש- 3-SAT היא NP-קשה. זו בעיה שניתן לבטא כ- ILP.

$x_i\in\{0,1\}$ במשמעות ש- $x_i=1$ אומר שהמשתנה ה-i הוא True ו- $x_i=0$ פירושו ש- False.

נסמן ב- $z_j$ את המשתנים הבוליאניים שבנוסחה. נניח ש- $F=\wedge C_i$ כשכל פסוקית $C_i$ היא מהצורה $z_{i_1}^{(\varepsilon)} \vee z_{i_2}^{(\varepsilon)} \vee z_{i_3}^{(\varepsilon)}$ כאשר $(\varepsilon)$ מאפשר שלילה. למשל, $z_5 \vee \bar{z}_7 \vee z_{12}$.

יותר קל להתחיל מ סימון $y_i\in\{-1,1\}$ (1 - true) כך שלעבור ש- $x_i\in\{0,1\}$ ע"י $x_i=\frac{1+y_i}{2}$

איך נבטא את הדרישה ש- $z_5 \vee \bar{z}_7 \vee z_{12} = \text{True}$

דרך ש- $y_5-y_7+y_{12} \ge -1$

$\Leftarrow \quad (2x_5-1)-(2x_7-1)+(2x_{12}-1) \ge -1$

ז"א זוהי בעיה NP-קשה להכריע אם מערכת של אי-שוויונים ליניאריים ניתנת להיפתר ע"י משתנים $x_i$ המקבלים את הערכים 0,1.

(36) 6/11/09
ספוממטיים

<u>שינויים בלוחות הזמנים:</u>
אין שיעור בשבוע הבא (13/11). יתכן שהתרגול ייערך בשעת השיעור הרגילה - פרטים יינתנו ע"י אילן.
שיעור השלמה - יום ו' 23/11 9-12 פרטים בהמשך.

---

תכנון ליניארי: $\max f(x)$ או $\min f(x)$
כאשר $x \in D$ והקבוצה $D$ מוגדרת ע"י משוואות ואי שוויונים ליניאריים.
נוח במיוחד ואפשרי להביא תוכניות ליניארית נתונה לאחת משתי הצורות הבאות (הנקראות קנונית וסטנדרטית)

$$\begin{aligned} &\max \langle c,x\rangle \\ &\text{s.t } Ax\le b \\ &x\ge 0 \end{aligned} \qquad\qquad \begin{aligned} &\max \langle c,x\rangle \\ &\text{s.t } Ax=b \\ &x\ge 0 \end{aligned}$$

מטרתנו היא שכל תוכנית ליניארית אפשר לייצג בצורה שקולה בכ"א משתי הצורות הנ"ל.

1. איך להיפטר ממשוואות? $\langle \alpha,x\rangle=\beta$

$\langle \alpha,x\rangle=\beta$ אומר $\langle \alpha,x\rangle\le\beta$, $\langle \alpha,x\rangle\ge\beta$ אומר $\langle \alpha,x\rangle\le\beta$, $-\langle \alpha,x\rangle\le-\beta$

2. איך להכריח את התנאי שהקואורדינטות ב-x אי-שליליות?
נאמר שאין על $x_i$ תנאי כזה. אז נגדיר
$x_i=y_i-z_i$ כאשר $y_i, z_i\ge 0$

3. איך להחליף אי שוויון במשוואה?
נניח שמופיע האי שוויון $\langle \alpha,x\rangle\ge\beta$. נהפוך אותו למשוואה + תנאי אי שליליות: $\langle \alpha,x\rangle-z=\beta$
$z\ge 0$

---

ט קבוצה המוגדרת ע"י רשימה סופית של משוואות

ואי שוויונים לינאריים נקראת <u>פוליהדר</u>. פוליהדר חסום נקרא <u>פאון</u>.

ייתכן שבבעיית LP נתונה פונקציית המטרה איננה חסומה - זה מקרה פחות מעניין. אבל אם האופטימום חסום אז ניתן להניח ש- D (תחום ההגדרה של הבעיה) הוא פאון (לא נראה איך, אבל זה לא מאוד קשה)

אחרי הרדוקציה הזאת בעיית ה- LP נראית גם רק ורק: $\max\{f(x) : x \in P\}$ כאשר $f$ לינארית ו- $P$ פאון.

אם $\alpha$ וקטור ו- $\beta$ מספר אז הקבוצה $\{x \in \mathbb{R}^n : \langle \alpha, x \rangle = \beta\}$ היא על מישור ב- $\mathbb{R}^n$ (כלומר זה תת מרחב אפיני ממימד $n-1$).

התמונה הגיאומטרית: מבטים במשפחה החד-פרמטרית של על-מישורים מקבילים $H_{\alpha,\beta} = \{x : \langle x, \alpha \rangle = \beta\}$ כאשר $\alpha$ קבוע ו- $\beta$ משתנה. מבין כל העל-מישורים האלה אשר חותכים את הפאון $P$, אנו מבקשים לדעת מהו $\beta$ המקסימלי שבשבילו $H_{\alpha,\beta}$ עדיין פוגש את $P$.

האינטואיציה הגיאומטרית אומרת שהאופטימום יתקבל בקודקוד של הפאון $P$. (האופטימום לא בהכרח מתקבל <u>רק</u> בקודקוד אבל <u>גם</u> בקודקוד).

האבחנה הזו מדגימה לנו שהבעיה $\max_{x \in P} f(x)$ הנראית לכאורה כבעיית אופטימיזציה רציפה ניתנת לתרגום לבעיית האופטימיזציה הבדידה הבאה

$$\max_{x \in vert(P)} f(x)$$

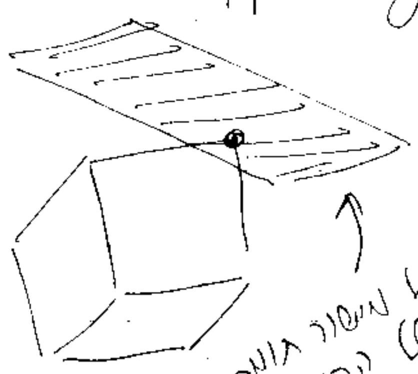
שזה טוב כי זה אומר שהבעיה הוא סופית!

## מעט יסודות בגאומטריה של פאונים:

- כאמור, פאון זו תת קבוצה חסומה של $\mathbb{R}^n$ אשר מוגדרת ע"י משוואות ואי שוויונים לינאריים.
- תת מרחב אפיני זוהי הזזה של תת מרחב לינארי. אז למשל ישר דרך הראשית זה תת מרחב לינארי ממימד 1 ואילו ישר כלשהו הוא תת מרחב אפיני. למשל תת מרחב אפיני הוא קבוצה מהצורה $L+x_0$ כאשר $L$ תת מרחב לינארי ו-$x_0\in\mathbb{R}^n$. את הממד של התת מרחב האפיני $L+x_0$ מגדירים ע"י $\dim L$.
- אם $H$ הוא על-מישור $H=\{x \mid \langle a,x\rangle=\beta\}$, הוא מחלק את $\mathbb{R}^n$ לשני חצאי מרחב:

$$H^+=\{x\mid\langle a,x\rangle\ge\beta\} \qquad H^-=\{x\mid \langle a,x\rangle\le\beta\}$$

- אם $P$ פאון ו-$H$ על-מישור אומרים ש-$H$ הוא **על-מישור תומך** (supporting hyper-plane) של $P$ אם $P\subseteq H^+$ או $P\subseteq H^-$ ו-$P\cap H\neq\emptyset$.




במקרה כזה קוראים ל-$P\cap H$ פאה (face) של $P$.

היות שכבר הגדרנו ממד של תת מרחב אפיני, מוגדר לנו המימד של הפאה: זה הממד של הנפרש האפיני (affine span) של הפאה. הנפרש האפיני של קבוצה $S\subseteq\mathbb{R}^n$ הוא

$$aff\text{-}span(S)=\{x=\sum\alpha_i x_i \mid \forall i\ x_i\in S,\ \sum\alpha_i=1\}$$

לפאה 0-ממדית אני קורא קודקוד

1-ממדית צלע

דיפן היא פאה ממימד שקטן ב-1 ממימד הפאון (facet)

**משפט**: כל פאון ב-$\mathbb{R}^n$ הוא הקמור של קודקודיו.
כלומר, ניתן לתאר פאון כקמור של קבוצה סופית של נקודות ב-$\mathbb{R}^n$.

**הערה**: לפי ההגדרה החדשה ברור שפאון הוא קמור, אבל ניתן לראות זאת גם מההגדרה הישנה. פאון מוגדר ע"י אוסף אי שוויונים $\Rightarrow$ הוא חיתוך של חצאי מרחב. אבל חצי מרחב הוא קמור וחיתוך של קבוצות קמורות הוא קבוצה קמורה לכן פאון הוא קמור.

כבר ממה שאמרנו יוצא שאכן כל פונקציה לינארית מקבלת את ערכי הקיצון שלה ב-$P$ בקודקוד של $P$. נביט במשפחת העל מישורים $H_\beta = \{x \mid \langle a, x \rangle = \beta\}$ כאשר $a$ קבוע ו-$\beta$ משתנה. יהי $\beta^*$ הערך המקסימלי של $\beta$ כך ש- $H_\beta \cap P \neq \emptyset$. רואים ש- $H_{\beta^*}$ הוא על מישור תומך של $P$ ולפיכך ב- $H_{\beta^*} \cap P$ יש קודקודים (מכח המשפט).

כלומר, אבחנה זו מאפשרת באופן עקרוני לפתור את בעיית התכנון הלינארי בתהליך חישוב סופי. כלומר, כדי למצוא את $\max_{x \in P} f(x)$ כאשר $f$ לינארית ו-$P$ פאון, נמצא את $\max_{x \in Vert(P)} f(x)$.
זה מביא למחשבה שניתן אולי לפתור את בעיית ה-$LP$ באופן הבא:

- צעד אתחול: מוצאים קודקוד $x_1$ של הפאון
- צעד כללי: עבור מ-$x_j$ הנוכחי (קודקוד של $P$) לקודקוד $x_{j+1}$ כך ש-
  - הקטע $[x_j, x_{j+1}]$ הוא פאה חד-ממדית (צלע) של הפאון $P$ (כלומר $x_j$ ו-$x_{j+1}$ שכנים)
  - ו- $f(x_{j+1}) \ge f(x_j)$

למעשה זו משפחה שלמה של אלגוריתמים הנקראים בשם הכולל אלגוריתם הסימפלקס.

כדי להגדיר את האלגוריתם בשלמות נדרשים שני פרטים:

א איך מוצאים את נקודת ההתחלה

ב בהינתן $x_j$ וקבוצת הקודקודים $y$ השכנים ל-$x_j$ כך ש- $f(y) \ge f(x_j)$, לאיזה מהם נעבור?

<u>שאלות מתבקשות:</u>

א האם אי אפשר "להיתקע"? כלומר האם אפשרי שיש ל-$P$ קודקוד $x$ כך ש- $f(x) \ge f(y)$ לכל קודקוד $y$ השכן ל-$x$ ובכל זאת $x$ אינו קודקוד אופטימלי?

ב איך לממש את האלגוריתם הזה הלכה למעשה (איך מחשבים קודקודים, שכנות וכו').

ג ועד כמה האלגוריתם יעיל? כלומר כמה זמן הוא רץ?

<u>תשובות:</u>

א בעיה כזו איננה קיימת. ובעצם התשובה הזו תקפה גם במקרה רחב יותר. מספיק ש-$P$ תהיה קבוצה קמורה (לאו דווקא פאון) ו-$f$ פונקציה קמורה (לאו דווקא לינארית). כלומר באופן כללי אם $P$ קמורה ו-$f$ קמורה ו-$x$ אופטימום מקומי, אז $x$ בהכרח אופטימום גלובלי.

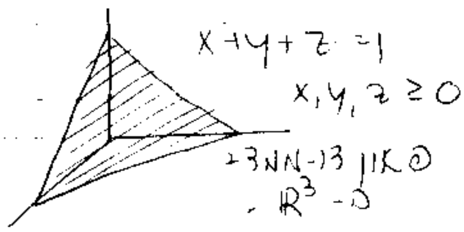
ג לא ידוע! האלגוריתם הומצא ב-1949. ב-1972 הראו שגרסה מסוימת של אלגוריתם הסימפלקס עשויה לקחת זמן מעריכי ב-$n$ (כאשר $A_{m\times n}$) (הגרסאות השונות של האלגוריתם אומרות באיזה שכן לבחור).

בעיה קשורה היא בעיית הקוטר של פאונים. לכל פאון $P$ מתאים באופן טבעי גרף $G$ שקודקודיו הם קודקודי $P$ ו-$xy$ הוא צלע ב-$G$ אמ"מ $[xy]$ צלע של $P$.

השאלה כמה גדול יכול להיות הקוטר של $G$ היא פתוחה!
אבל יש חסמים עליונים תת-מעריכיים...

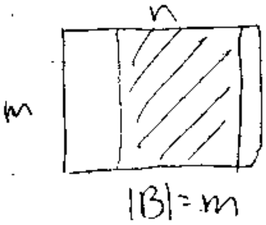
<u>מימוש</u>

איך נמצא קודקודים של $P$? נניח ש-$P$ מתואר ע"י
$$P=\{x\in\mathbb{R}^n : Ax=b,\ x\ge 0\}$$
(הממד של $P \le n$ כי אנחנו משתמשים בשיטה
להפיך אי-שוויונים לשוויונים ע"י הוספת משתנים).
נניח ש-$A_{m\times n}$ כאשר $m\le n$ (אחרת הפאון הוא
נקודה וזה לא מעניין). כמו כן, נניח ש-$\operatorname{rank}A=m$
(כלומר השורות של $A$ בת"ל). אם אכן $P\ne\emptyset$ אז
$\dim P=n-m$.




אם $B$ היא קבוצה של $k$
עמודות של $A$ ונתייחס לפעמים גם
ל-$B$ כמטריצה $m\times k$.

אם $x$ הוא וקטור $n$-ממדי אז התומך של $x$ הוא
$$\operatorname{supp}(x)=\{i: x_i\ne 0\}$$
אם $k=m$ והעמודות ב-$B$ הן בת"ל נקרא ל-$B$ בסיס.




אם $B$ בסיס אז כל עמודה ב-$A$
ניתן להציג באופן יחיד כצירוף לינארי של
עמודות מ-$B$.

אומרים ש-$x$ הוא פתרון מותר (feasible) של המערכת
$Ax=b$ אם $x\ge 0$.

אם $B$ בסיס אז יש $x$ אחד ויחיד כך ש- $Ax=b$
ו- $\operatorname{supp}x\subseteq B$ (כלומר $x_i\ne 0$ אם העמודה ה-$i$
בבסיס). לכל בסיס $B$ של המטריצה $A$ מוגדר
באופן יחיד הפתרון $x$ ל-$Ax=b$ ע"י $x"="B^{-1}b$
ושאר האינדקסים אפסים.

אומרים שפתרון $x$ ל- $Ax=b$ הוא בסיסי אם קבוצת העמודות המתאימות ל- $\mathrm{supp}\,x$ היא בת"ל.
אם כן, <u>פתרון בסיסי מותר</u> למערכת $Ax=b, x\geq 0$ הוא וקטור $x$ המקיים

(1) $x\geq 0$

(2) $Ax=b$

(3) קבוצת העמודות המצויינות ע"י $\mathrm{supp}\,x$ הן בת"ל.

<u>משפט</u>: יש התאמה חח"ע בין קודקודי $P$ לפתרונות הבסיסיים המותרים של המערכת $x\geq 0$, $Ax=b$.

בפרט, נובע מהמשפט שבפאון $\{x: 0\leq x,\ Ax=b\}$ יש לכל היותר $\binom{n}{m}$ קודקודים (כי קודקוד מגדיר בסיס ויש לכל היותר $\binom{n}{m}$ בסיסים).

<u>משפט</u>: נניח שפונקצית המטרה $x\mapsto\langle c,x\rangle$ מאולצים $x\geq 0$, $Ax=b$ היא חסומה. אז המקסימום של פונקצית המטרה מושג ע"י $x$ שהוא פתרון בסיסי מותר. (כלומר מקודקוד).

<u>הערה</u>: המקסימום אכן מתקבל כי פאון כזה הוא קבוצה קומפקטית, ופונקציה לינארית היא רציפה ופונקציה רציפה על קבוצה קומפקטית תמיד מקבלת מקסימום על הקבוצה.
למעשה צריך לדייק. הפאון לא בהכרח קומפקטי אבל היות שהפונקציה חסומה ניתן להוסיף עוד תנאים מהצורה $x_i\leq M\ \ \forall i$ רק שערכי הפתרונות לא יקטן. ולמעשה אנחנו מחפשים מקסימום של פונקציה רציפה על קבוצה קומפקטית (והוא אכן מושג.

<u>הוכחה</u>: יהי $x^*$ הפתרון האופטימלי. כלומר $x^*\geq 0$, $Ax^*=b$ ; $\langle c,x^*\rangle$ מקסימלי מבין מתאימים אלה.
נניח שיש $\bar{x}$ המקיים אותם אילוצים, $\langle c,x^*\rangle=\langle c,\bar{x}\rangle$ ו- $\bar{x}$ פתרון בסיסי מותר.

הקושי היחידי שצריך להתעורר הוא שהעמודות ב-A המתאימות ל-
$\mathrm{supp}\,x^*$ הן תלויות לינארית. אכן, יש וקטור $y$ כך ש- $Ax=0$
ו- $\mathrm{supp}\,y \subseteq \mathrm{supp}\,x^*$. (נמצא את $\bar{x}$ כך: מבין ה-$x$
המקיימים $x\geq 0$ , $Ax=b$ , $\langle c,\bar{x}\rangle \geq \langle c,x^*\rangle$ ; $|\mathrm{supp}\,\bar{x}|$
קטן ביותר. ($x^*$ מקיים את מה שצריך אז אנחנו מחפשים
מינימום של קבוצת טבעיים לא ריקה אז קיים מינימום).
ננסה $\bar{x} = x^* + \varepsilon y$ . אז

- $A\bar{x} = Ax^* + \varepsilon Ay = Ax^* + 0 = Ax^* = b$
- האם $0 \leq x^* + \varepsilon y$ ? כן! לפחות אם $|\varepsilon|$ קטן
מספיק כי $\mathrm{supp}\,y \subseteq \mathrm{supp}\,x^*$ .
- נגדיל את $|\varepsilon|$ המקסימלי כך שעדיין נשמר התנאי ש-
$x^* + \varepsilon y \geq 0$ אבל נוספה עוד קואורדינטה של אפס,
כלומר גודל התומך ירד. זה בסתירה לכך שבחרנו את
$x^*$ כבעל התומך הקטן ביותר האפשרי.

זה מחייב שהפסדנו בפונקציית המטרה. במקרה זה נהפוך
את סימנו של $\varepsilon$ ונקטין את $|\varepsilon|$ אם צריך כדי
להפוך את הסימן של $\varepsilon\langle c,y\rangle$ ובזה הגענו לסתירה.

(40) 20.01.09

[illegible]

דיברנו על אלגוריתם הסימפלקס - שם נוח לעבוד עם הצורה

$$(LP)\qquad \max \langle c, x\rangle \quad \text{s.t.}\quad Ax=b,\quad x\geq 0$$

$A_{m\times n}$ ממשית, $\mathrm{rank}(A)=m$, ובפרט $m\leq n$

הגדרנו פאון $P$, ופאה $d$-ממדית. בפרט עבור $d=0$ הפאה היא קודקוד של $P$. הגדרנו את המושג של פתרון בסיסי אפשרי (פב"א - bfs). במונחים האלגבריים אומרים ש-$x$ פתרון אפשרי אם $x\geq 0$ ; $Ax=b$. בסיסי פירושו שאם מסמנים $K=\mathrm{supp}\,x=\{i: x_i>0\}$ אז קבוצת העמודות ב-$A$ המצוינות ע"י $K$ היא בת"ל. ראינו שיש התאמה חח"ע בין קודקודי $P$ ופב"א של המערכת. כמו כן, ראינו שאם פונקציית המטרה $x\mapsto\langle x,c\rangle$ היא חסומה מלמעלה על $P$ אז האופטימום מושג בקודקוד של $P$, כלומר ע"י $x$ שהוא פב"א.

לא הוכחנו זאת, אבל קודקוד של פאון $P$ ניתן לאפיין כנקודה של $P$ שלא אפשר להציג כצירוף קמור ממש של נקודות שונות ב-$P$ (אם $x_1,\dots,x_t\in\mathbb{R}^n$ אומרים ש-$y$ צירוף קמור שלהן אם ניתן לרשום $y=\sum\alpha_i x_i$ כאשר $\alpha_i\geq 0$ ; $\sum\alpha_i=1$. הצירוף קמור ממש אם המקדמים הם לא $0,\dots,0,1,0,\dots,0$ - כלומר לא פשוט אחת הנקודות $x_i$ עצמן).

[מי שרוצה לקרוא עוד על הנושא יכול לעיין ב- Papadimitriou & Steiglitz (Combinatorial Optimization)]

מכאן נתאר האלגוריתם (המטופש ולא יעיל) ל-LP: עבור כל ה-$m$-יות $B$ של עמודות ב-$A$, אם ה-$m$-יה בת"ל מצאנו את ה-$x$ היחיד המקיים $Ax=b$ $\mathrm{supp}\,x\subseteq B$. אם $x\geq 0$ אז מצאנו פתרון בסיסי אפשרי. נחשב את הערך $\langle c,x\rangle$ בו. נבחר לבסוף את הערך הגבוה ביותר.

אלגוריתם הסימפלקס מממש את האלגוריתם הנ"ל בגישה
שיטתית יותר: עוברים מקודקוד של P לקודקוד שכן כזה
שאינו מקלקל את פונקציית המטרה.
בעצם מדובר לא באלגוריתם אחד, אלא במשפחה של
אלגוריתמים. יש לציין איך מתחילים ובעיקר - מבין G
השכנים של קודקוד נתון, לאיזה שכן עוברים.
(דרושים עוד: כללים למניעת לולאות אינסופיות (anticycling rules)
כי קורה שלא משפרים ממש את פונקציית המטרה אלא רק
שומרים על הערך שלה.
- איך מחשבים בפועל את המעבר מקודקוד (=פב"א)
לקודקוד שכן. (זה נקרא צעד ציר)
[מי שמעוניין בהקשר רחב יותר יכול לעיין ב.
[(Convex Optimization) Boyd Vanden Berghe

בסיס עמודות של $A$ זו קבוצה של $m$ עמודות בת"ל
ב-$A$. שני בסיסים הם שכנים אם ההפרש הסימטרי
ביניהם כולל בדיוק שתי עמודות. (כלומר השכן מתקבל
ע"י החלפת עמודה אחת בדיוק בבסיס).
מראים ששני קודקודים הם שכנים ב-$P$ (כלומר הקטע
ביניהם הוא פאה חד-ממדית של $P$) אמ"מ הבסיסים
המתאימים שכנים במובן הנ"ל.

קצת ידע על מצב המחקר: (עשו עבודות רבות שמנסות
להסביר מדוע באופן מעשי אלגוריתם הסימפלקס נוטה
לרוץ בזמן מהיר אף כי יש מקרים שבהם זמן הריצה
(לפחות של מספר וריאנטים של האלגוריתם) מעריכי.
עבודה חשובה בכיוון זה (של Spielman & Teng)
פיתחה סוג ניתוח הקרוי "ניתוח החלקה" (Smoothed Analysis)
הם הראו שאם LP (נאמר מהצורה $\max\langle c,x\rangle : Ax\le b,\ x\ge 0$)
משתנה באקראי במידה קטנה כרצוננו את איברי

$A, b, c$ נגיע לבעיית LP

$$\max \langle c', x\rangle \; : \; A'x \le b' \; , \; x \ge 0$$

את הבעיה החדשה, לאחר השינויים הקלים) ניתן להוכיח שכמעט בוודאות זמן הריצה של אלגוריתם הסימפלקס הוא $poly(m,n)$. אפשר להעריך מקלות את השינוי בפונקציית המטרה כתוצאה מהשינוי.

## דואליות LP

מהי הסיבוכיות החישובית של בעיית ה- LP?
הניתוח עם הבסיסים הראה: שהבעיה סופית - שזה דבר טוב.
אלגוריתם הסימפלקס יכול לרוץ בזמן מעריכי. אבל זה לא אומר שאין משהו טוב יותר.
האם הבעיה LP שייכת למחלקה NP? כדי שהשאלה תהיה טעם יש להתבונן בגרסת ההכרעה של LP (ולא בגרסת האופטימיזציה) - האם יש $x \ge 0$ כך ש-

$$Ax = b \; ; \; \langle c, x\rangle \ge T \quad (*)$$

ע"י חיפוש בינארי של ערכי T קל לראות ששתי הבעיות (גרסת ההכרעה וגרסת האופטימיזציה) הן שקולות.
כמובן, (*) הוא ב- NP כי x מהווה עד לקיום הפתרון.
האם LP (הכרעה) שייכת ל- co-NP? (נניח שאין x המקיים $x \ge 0$ כך ש- $Ax = b$ ; $\langle c, x\rangle \ge T$.
איך עשוי להיראות עד קצר לדבר?

ידוע לנו עוד מביה"ס שמותר לבצע צירופים לינאריים אחרים של משוואות ואי-שוויונים לינאריים. את המשוואות ניתן להכפיל במקדמים כלשהם (שליליים אפשר), את האי שוויונים במקדמים חיוביים בלבד. משפט הדואליות הוא משפט שלמות של השיטה הנ"ל. הוא אומר שכל מה שניתן להסיק ממשוואות ואי"ש לינאריים ניתן להסיק באופן זה.

במקרה שפניני התיאור הוא כדלקמן: אם אכן אין
פתרון למערכת $\langle c,x\rangle \geq T$ $Ax=b$ $x\geq 0$.
נניח שיש $y$ רק ש- $yA=c$ ; $\langle y,b\rangle < T$
נכפול מימין ב-$x$ ונקבל $\langle c,x\rangle = yAx = \langle y,b\rangle < T$
ואז במובן המערכת לא ספיקה.
אבל בעצם אין צורך ש- $yA=c$ ומספיק ש- $yA\geq c$
כדי שנוכל להסיק את המסקנה הרצויה (כי $x\geq 0$).

ראינו: אם יש וקטור $y$ המקיים $yA\geq c$ ;
$\langle y,b\rangle < T$ אז אין $x$ המקיים $Ax=b, x\geq 0$ $\langle c,x\rangle \geq T$
כי במקרה כזה נכפול את המשוואות ע"י
$y$ ונקבל לכל $x\geq 0$ $\langle c,x\rangle \leq yAx = \langle y,b\rangle < T$

כאן נוח במיוחד לעבוד עם LP בצורה
$$\begin{array}{l}\max\langle c,x\rangle\\ Ax\leq b\\ x\geq 0\end{array}\quad (LP)$$

ול-(LP) מתאימה תוכנית לינארית דואלית

$$(DLP)\qquad \begin{array}{l}\min\langle b,y\rangle\\ yA\geq c\\ y\geq 0\end{array}$$

<u>משפט הדואליות ב-LP</u>

הגרסה החלשה: לכל $x$ מותר ל-LP ולכל
$y$ מותר ל-DLP מתקיים $\langle c,x\rangle \leq \langle b,y\rangle$

גרסה החזקה: אם באחת מהשתיים מושג האופטימום אז

$$\begin{array}{l}\max\langle c,x\rangle\\ Ax\leq b\\ x\geq 0\end{array} = \begin{array}{l}\min\langle y,b\rangle\\ yA\geq c\\ y\geq 0\end{array}$$

הוכחת הגרסה החלשה: יהי $x$ וקטור מותר ל-LP ו-
$y$ ל-DLP אז

$$\langle c,x\rangle \leq \underset{0\leq}{y}\,\underbrace{\overset{\geq c}{A}\,\overset{\geq 0}{x}}_{\leq b} \leq \langle y,b\rangle$$

(42)

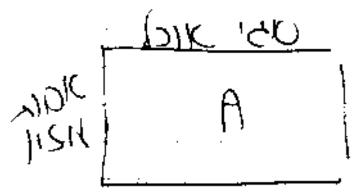
מדוע משקפת הגירסה החלשה של משפט הדואליות את עיקרון השמות שניסחנו מקודם?

הטענה: (1) יש $x \geq 0$ המקיים $Ax \leq b$ ; $\langle c,x \rangle \geq T$
היא ב-NP

אבל הטענה (2) לכל $0 \leq x$ המקיים $Ax \leq b$ מתקיים $\langle c,x \rangle \leq T$ היא גם ב-NP (כאן העד שלנו הוא ה-y האופטימלי של DLP). במילים אחרות, טענה (1) היא ב-co-NP ולכן $LP \in NP \cap co\text{-}NP$.

דוגמאות לשימושים של משפט הדואליות:

- בעיית התזונה של הפרה:




$a_{ij}$ = הכמות של אב מזון $i$ בסוג אוכל $j$

אנחנו רוצים למצוא את הדרך הזולה ביותר להזין פרה תוך עמידה בדרישות המינימום.

$$\min \langle c,x \rangle$$
$$x \geq 0$$
$$Ax \geq b$$

הקואורדינטה ה-$t$ בוקטור $x$ אומרת כמה יחידות מסוג אוכל $t$ נותנים לפרה.

$c_i$ - מחירה של יחידת אוכל מסוג אוכל $i$.

$b_j$ - הכמות המזערית המותרת של אב מזון $j$ בתזונת הפרה.

לפי משפט הדואליות

$$\min_{\substack{x \geq 0 \\ Ax \geq b}} \langle c,x \rangle = \max_{\substack{y \geq 0 \\ yA \leq c}} \langle y,b \rangle$$

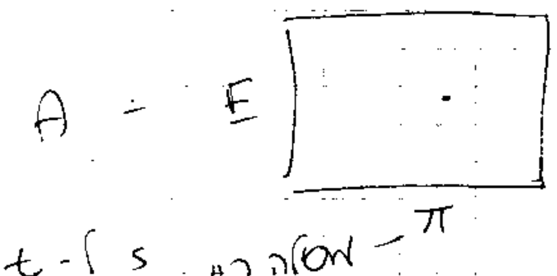
הבעיה הדואלית היא הבעיה של יצרן מזון סינתטי. הוא

רוצה לקבוע מחיר ליחידה אחת של b אם מצין. הרעיון
עשוי לא דווקא להצין את הפרה בגודל המספקת בדיוק את
צרכי הפרה. במקרה זה $\langle b,y \rangle$ הוא המחיר להצין
פרה בדיוק בדרישות המינימום התזונתיות שלה.
היצרן צריך להתחרות באפשרויות הזנה המסורתיות. לכן
אבטח התנאי $yA \le c$. הקואורדינטה ה-$i$ באי שוויון
הזה מבטאת את הדרישה שמחירי אבות המזון
המיוצרים הם כאלה שיעמדו בתחרות מול סוגי האוכל הקיימים.

• max flow - min cut

ראינו בקורס באלגוריתמים את משפט השטף והחתך.
נראה שזהו מקרה פרטי של משפט הדואליות.
יש לנו רשת זרימה, כלומר גרף לא מכוון $G=(V,E)$ עם
שני קודקודים מיוחדים: $s$ - המקור
$t$ - הבור
ויש גם קיבולים לצלעות $c: E \to \mathbb{R}^+$
זרימה זו פונקציה $f: E \to \mathbb{R}^+$ המקיימת:
1. שימור זרימה בכל קודקוד פרט ל- $s,t$
2. באף צלע לא מזרימים יותר מקיבולה.

רוצים למקסם את השטף מ-$s$ ל-$t$.
עובדה ידועה שנשתמש בה היא שכל זרימה ניתן לבטא כצרוף
אי-שלילי של מסלולים בין $s$ ל-$t$.




$c_e$ - הקיבול של הצלע $e$

$\pi$ - מסלול בין $s$ ל-$t$

$$A_{e\pi} = \begin{cases} 0 & \text{אחרת} \\ 1 & \text{הצלע } e \text{ במסלול } \pi \end{cases}$$

מגדירת המטריצה הזו נרצה לנסח את בעיית השטף המקסימלי ברשת הזרימה שלנו.

נסמן ב- $x_\pi$ את הזרם שעובר דרך $\pi$. $0 \le x_\pi$
אז התנאים הם: $0 \le x$ ; $Ax \le c$ (כלומר בכל
צלע אנחנו מזרימים לכל היותר את קיבולה).
והשטף הוא $\langle x, \vec{1} \rangle$ : אז מה שאנחנו צריכים
לפתור זה

$$\begin{aligned} \max\ & \langle x, \vec{1} \rangle \\ s.t\ & Ax \le c \\ & x \ge 0 \end{aligned} \qquad \text{max flow}$$

הבעיה הדואלית היא

$$\begin{aligned} \min\ & \langle y, c \rangle \\ s.t\ & yA \ge 1 \\ & y \ge 0 \end{aligned}$$

לפי משפט הדואליות הבעיות שוות.
אנחנו נטען שזה בדיוק min cut חיום.
ברור שב-$y$ אופטימלי כל הקואורדינטות הן בין 0 ל-1.
(נניח) רגע שכל הקואורדינטות מ-$y$ הן 0 או 1. במקרה זה
$y$ הוא הווקטור המציין של קבוצת צלעות $E \supseteq F$.
הדרישה $yA \ge 1$ אומרת ש-$F$ פוגשת כל מסלול מ-$s$
ל-$t$, אז $F$ כוללת חתך צלעי בין $s$ ל-$t$.
$\langle y, c \rangle$ זהו סכום הקיבולים של הצלעות ב-$F$. אז

$$\text{min cut} = \min_{\substack{y \ge 0 \\ yA \ge 1 \\ \forall e\ y_e \in \{0,1\}}} \langle y, c \rangle$$

ביום שלישי בבוקר נראה שהאופטימום מתקיים בווקטור כזה
ואז נסיים.

לעקוב אחרי הודעות לאתר - נפגשים ביום
שלישי ב- $8^{00}$

27/01/09

סום מגמטיים

סדר היום:

- עוד כמה מילים על "ישומי" משפט הדואליות
- למת פרקש ( + הוכחה)
- אלגוריתם האליפסואידים
- אנליזה הרמונית

**משפט הדואליות**: נביט בזוג התכניות הדואליות הבא:

$$(LP)\quad \max \langle c,x\rangle \quad Ax \le b \quad x \ge 0$$

$$\min\langle b,y\rangle \quad yA \ge c \quad y \ge 0 \quad (DLP)$$

אם אחת מהן ספיקה וחסומה אז גם השניה כנ"ל והערך האופטימלי בשתיהן שווה.

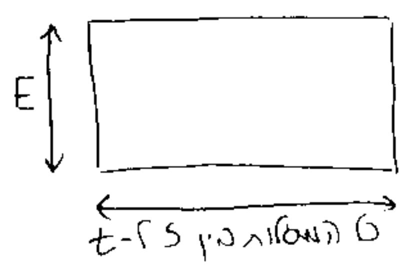
[הגרסה החלשה של משפט הדואליות אומרת שאם $x$ מותר לבעיה הפרימאלית ואם $y$ מותר לבעיה הדואלית מתקיים $\langle c,x\rangle \le \langle b,y\rangle$ ]

**למת פרקש**: המערכת $Ax = b$, $x \ge 0$ אינה פתירה אמ"מ יש $y$ כך ש-

(i) $\langle y,b\rangle < 0$

(ii) $0 \le yA$

**משפט השטף והחתך**




$x_\pi \ge 0$ = כמה הזרמנו על המסלול $\pi$

אם בעית הפרימאל ניסחנו כך:

$$\max \langle x, \vec{1}\rangle$$

s.t.

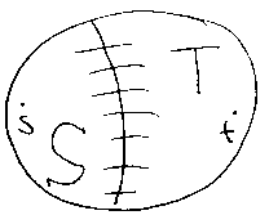
אסור לחרוג מהקיבולים המותרים $\quad Ax \le c$

$$x \ge 0$$

לפי משפט הדואליות בגרסתו החזקה המשקל המירבי שווה גם ל-

$$\begin{aligned} &\min \langle c,y\rangle \\ &\text{s.t} \\ &\quad yA \geq \vec{1} \\ &\quad y \geq 0 \end{aligned}$$

אם $y$ הוא וקטור ב- $\{0,1\}^m$ אז הוא וקטור מציין של קבוצת צלעות $F$ הפוגשות כל מסילה בין $s$ ל-$t$. וכרגע יש לה קיבול מזערי. כלומר, $y$ הוא הווקטור המציין של חתך צלעות מזערי.



למה? אם יש לנו קבוצת צלעות כזו $F$ אז נגדיר את $Z$ כרכיב הקשירות של $s$ בגרף $G \setminus F$.

מראים שכאן הפתרון האופטימלי של התכנית הדואלית הוא ב-$0|1$ משיקולים של אונימודולריות שלמה. (=כל מינור של המטריצה של דטרמיננטה $\pm 1$)

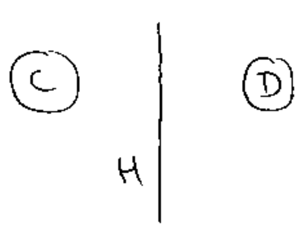
לא קשה להסיק את משפט הדואליות בשלמותו מתוך למת פרקש. לשם כך נוח להסב את בעיית האופטימיזציה LP לבעיית ספיקות.

$$\begin{aligned} &\max \langle c,x\rangle \\ &\text{s.t} \\ &\quad Ax \leq b \\ &\quad x \geq 0 \end{aligned} \quad \Leftarrow \quad \begin{aligned} &\text{האם יש } x \text{ המקיים} \\ &0 \leq x,\ Ax \leq b \ ? \\ &\langle c,x\rangle > T \end{aligned}$$

(אופטימיזציה) (ספיקות)

למת פרקש ודומה משפטי הדואליות נובעים ממשפטים גאומטריים יסודיים ואינטואיטיביים - משפטי הפרדה.

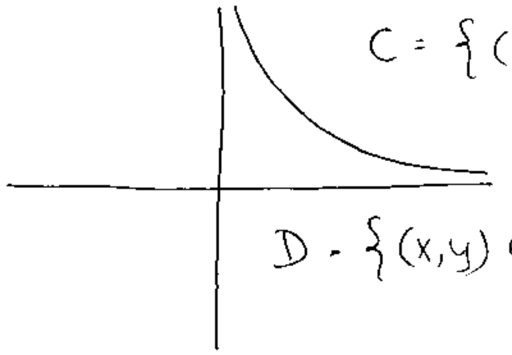
אם יש שתי קבוצות קמורות וזרות ב- $\mathbb{R}^n$ ניתן להפריד ע"י מישור



שאלות מעניינות:

- מה אם אחת הקבוצות היא נקודה?
- האם $C,D$ סגורות? חסומות?
- האם משטח ההפרדה חזק? [ $H=\{x:(a,x)=\alpha\}$ ואז ב- $D$ מתקיים $(a,x)\geq\alpha$ ובה- $C$ $(a,x)\leq\alpha$. ההפרדה חזקה אם האי שוויונים חזקים ]

דוגמה: יתכן להיות ש- $C,D$ קמורות וסגורות (אבל לא חסומות) אבל אין משטח הפרדה חזק:



$C=\{(x,y)\in\mathbb{R}^2 : x\geq 0,\ y\geq 0,\ xy\geq 1\}$

$D=\{(x,y)\in\mathbb{R}^2 : y\leq 0\}$

כאן מישור ההפרדה הוא ציר ה-$x$ אבל ברור שהוא לא חזק.

מצד שני, אם $C,D$ קמורות, סגורות וחסומות, יש משטח הפרדה חזק.

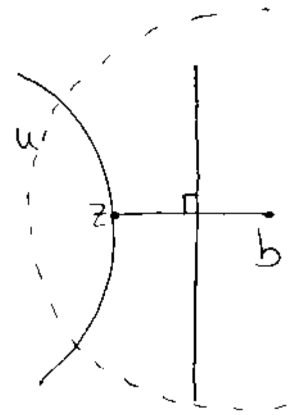
<u>הערה</u>: למת פרקש היא משפט הפרדה. קבוצת הווקטורים $z$ שניתן להציגם כ- $z=Ax$ עבור $x\geq 0$ כלשהו היא קונוס - הקונוס $C$ הנוצר ע"י העמודות של $A$. הלמה מדברת על הפרדת הנקודה $b$ מהקונוס $C$.

קונוס = קבוצה סגורה לחיבור וכפל בסקלר אי-שלילי

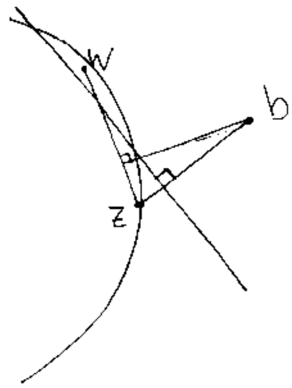
דרך ליצור קונוס (קונוס נוצר סופית - סימפליציאלי): לוקחים קבוצה סופית של וקטורים $u_1,\dots,u_k\in\mathbb{R}^n$ ומביטים בקבוצה $\{a_1u_1+\dots+a_ku_k : a_i\geq 0\}$.
קל לראות שזה אכן קונוס. במקרה שלנו, $C$ הוא הקונוס הנוצר (סופית) ע"י העמודות של $A$. התנאי "אין $x\geq 0$ המקיים $Ax=b$" שקול ל- $b\notin C$.
הלמה אם כן, מאפיינת אילו נקודות $b$ אינן נמצאות בקונוס $C$.
הלמה אומרת ש- $b\notin C$ אמ"מ יש מישור $H$ המפריד בין

b וקונוס C, $H = \{z : \langle y,z\rangle = 0\}$ כי $b \in H^-$
ו(חזק)! $C \subseteq H^+$.

<u>משפט ההפרדה</u>: תהי $C \subseteq \mathbb{R}^n$ קמורה וסגורה ותהי $b \notin C$
אז יש על מישור $H$ המפריד חזק בין $b$ ו-$C$, כלומר
קיימים $a, \alpha$ כך ש- $\langle a,b\rangle < \alpha$ ולכל $x \in C$
$\langle a,x\rangle > \alpha$.




<u>הוכחה</u>: קיימת נקודה $z \in C$ שממששת
את המרחק בין $b$ ו-$C$. בעיקר זה נובע
מכך שפונקציה רציפה מקבלת מינימום
בקבוצה קומפקטית. תהי $u \in C$ כלשהי.
נגדיר $D = \{v \in C : \|v-b\| \le \|u-b\|\}$
זו קבוצה קומפקטית (חיתוך של סגורה עם קומפקטית)
והפונקציה $t \mapsto \|b-t\|$ רציפה על $D$, לכן מקבלת מינימום
ב-$D$ וזה גם המינימום על $C$.
אז תהי $z$ הנקודה הקרובה ביותר ל-$b$ ב-$C$.
<u>למה</u>: לכל $w \in C$ מתקיים $\langle w-z, b-z\rangle \le 0$ (*)
יתר על כן, (*) מאפיין את הנקודה $z$.




<u>הוכחת הלמה</u>: נניח ש-(*) מתקיים. אז:
$$\|w-b\|^2 = \|(w-z)+(z-b)\|^2 =$$
$$= \|w-z\|^2 - 2\langle w-z, b-z\rangle + \|z-b\|^2 \ge$$
$$\ge \|z-b\|^2$$

לכן (*) מאפיין את $z$.
כעת נוכיח את (*): נניח שיש $w \in C$ המפר את (*),
כלומר $\langle w-z, b-z\rangle > 0$. נביט אז בנקודה
$z+t(w-z)$ עבור $0 \le t \le 1$. היות ש-$C$ קמורה,
זו נקודה ב-$C$. נראה שיש ערך $0<t<1$ שעבורו נקודה
זו קרובה ל-$b$ יותר מאשר $z$:
$$\|b-(z+t(w-z))\|^2 = \|z-b\|^2 - t\left(2\langle b-z, w-z\rangle - t\|w-z\|^2\right)$$

אם הנחנו ש- $\langle b-z, w-z\rangle > 0$ . רק יש בחירה של $t$ חיובי וקטן שעבורו האגף הזה קטן מ- $\|z-b\|^2$, בסתירה לבחירת $z$.

# אנליזה הרמונית

[מה הגננת שלנו עושה בלילות?]

נקודה $x\in\mathbb{R}^n$ היא למעשה פונקציה $x:\{1,\dots,n\}\to\mathbb{R}$
אבל $\{1,\dots,n\}$ הוא גם חבורה חיבורית (מודולו $n$).
אנחנו נתעניין באופן כללי בפונקציות $f:G\to\mathbb{C}$ כאשר $G$ חבורה חיבורית. החבורות $G$ שאנחנו נתעניין בהן הן:

- $(\mathbb{Z}_n,+)$
- $(\mathbb{R},+)$
- $(\mathbb{Z}_2^n,+)$

התורה מתחילה מן ההבחנה שאם על קבוצת האינדקסים שלנו יש מבנה חיבורי אז מתעורר באופן טבעי עניין בבסיס מיוחד ל- $\mathbb{R}^n$.
נחשוב למשל על $\{1,\dots,n\}$ כנקודות זמן. אז היינו מקווים שבבסיס שאיתו נעבוד יהיה קשר בין $f:[n]\to\mathbb{R}$ לבין $g:[n]\to\mathbb{R}$ המוגדרת ע"י $g(i)=f(i-1)$.
או למשל, בשימושים הקשורים לעיבוד אותות, אנחנו מסננים לעתים הרמות רעש ע"י פעולת קונבולוציה. למשל הדוגמה הכי פשוטה היא פשוט מיצוע - $g(i)=\frac{1}{4}f(i-1)+\frac{1}{2}f(i)+\frac{1}{4}f(i+1)$.

תזכורת: חבורה קומוטטיבית זו קבוצה $G$ יחד עם פעולה $+$, כך שלכל $x,y\in G$ מוגדר ביחידות $z\in G$ כך ש- $x+y=z$ ומתקיימות התכונות הבאות:

א' - יש איבר יחידה ב-$G$ שנסמנו ב-$0$ אשר לכל $x$, $x+0=x$

ב' - לכל $x\in G$ יש איבר הופכי שנסמנו $-x$ ומתקיים $x+(-x)=0$

ג' - הפעולה $+$ היא אסוציאטיבית: לכל $x,y,z\in G$

$$(x+y)+z = x+(y+z)$$

ד' - הפעולה $+$ היא קומוטטיבית. לכל $x,y\in G$

$$x+y=y+x.$$

דוגמאות:

- $\mathbb{Z}_n$ עם חיבור מודולו $n$
- $\{0,1\}^n$ עם חיבור לפי קואורדינטות מודולו 2
- $\mathbb{R}$ עם פעולת החיבור
- חבורת המעגל: $\mathbb{T}$ המספרים המרוכבים $z$ עם $\|z\|=1$ ופעולת הכפל של מרוכבים
אז האיבר הניטרלי הוא $1$ וההופכי של $\mathrm{cis}\theta$ הוא $\mathrm{cis}(-\theta)$ שגם הוא במעגל וכמובן יש סגירות לכפל כי $\mathrm{cis}\,\theta\cdot\mathrm{cis}\,\varphi=\mathrm{cis}(\theta+\varphi)$
חבורה זו נסמנים ב-$\mathbb{T}$.

**הגדרה**: קרקטר של חבורה קומוטטיבית $G$ הוא הומומורפיזם

$$\chi : G \longrightarrow \mathbb{T}$$

(כלומר לכל $x,y\in G$ $\chi(x+y)=\chi(x)\cdot\chi(y)$)

מיהם הקרקטרים של $\mathbb{Z}_n$? נניח $\chi:[n]\longrightarrow\mathbb{T}$

אם נדע מיהו $\chi(1)$ נדע את $\chi$ במלואו כי אז

$$\chi(j)=\chi(1+\dots+1)=\chi(1)\cdots\chi(1)=(\chi(1))^j$$

אם $y=0$ אז $\chi(x)=\chi(x+y)=\chi(x)\,\chi(y)$

$\Leftarrow$ בהכרח $\chi(0)=0$. ב-$\mathbb{Z}_n$, $n\equiv 0$

אבל $1=\chi(0)=\chi(n)=\chi(1)^n$ $\Leftarrow$ $\chi(1)$ ח"ב

להיות שורש יחידה n-י. נאמר $\chi(1)$ מוגדרת בצורה
$\chi(1) = e^{\frac{2\pi i}{n}k} \underset{\omega = e^{\frac{2\pi i}{n}}}{=} \omega^k$ ל-k כלשהו.

אלו אכן קרקטר שסמנו ב- $\chi_k$:

$$\chi_k(j) = (\chi_k(1))^j = e^{\frac{2\pi i}{n}jk}$$

לכל $k=0,\dots,n-1$ אלו אכן קרקטר:

$$\chi_k(a+b) = \omega^{(a+b)k} = (\omega^a)^k(\omega^b)^k = \chi_k(a)\chi_k(b)$$

אנליזה הרמונית על $\mathbb{Z}_n$ פירושה לקחת פונקציה כלשהי $f:\mathbb{Z}_n \longrightarrow \mathbb{C}$ ולהציג אותה כצירוף של הקרקטרים.

תכונה חשובה של הקרקטרים (כאן ב- $\mathbb{Z}_n$, אבל גם בכלל) היא שהם מהווים בסיס אורתונורמלי למרחב הפונקציות הממשיות על $\mathbb{Z}_n$

אם $f,g:\mathbb{Z}_n \longrightarrow \mathbb{C}$ אז המכפלה הפנימית שלהן היא

$$\langle f,g\rangle = \frac{1}{n}\sum_{x\in\mathbb{Z}_n} f(x)\overline{g(x)}$$

(ראה פשוט ההגדרה הסטנדרטית של מ"פ כשמסתכלים על הפונקציות כוקטורים). כעת

$$\langle \chi_k,\chi_l\rangle = \frac{1}{n}\sum_{t=0}^{n-1}\chi_k(t)\overline{\chi_l(t)} = \frac{1}{n}\sum_{t=0}^{n-1}\omega^{kt}\cdot\omega^{-lt} = \frac{1}{n}\sum_{t=0}^{n-1}\omega^{t(k-l)}$$

אם $k=l$ אז כל המחוברים הם 1 ; $\langle\chi_k,\chi_l\rangle = 1$

אם $k\neq l$ אז נניח לרגע ש-n ראשוני. היות ש- $k-l\neq 0$ אז המספרים $\{t(k-l)\}_{t=0}^{n-1}$ הם בדיוק כל המספרים $0,\dots,n-1$ בסדר כלשהו והסכום הוא לכן $\frac{1}{n}\sum_{s=0}^{n-1}\omega^s = \frac{1}{n}\cdot\frac{1-\omega^n}{1-\omega} = 0$

וכש-n פריק הסכום עשוי להיות מוצג כסכום של כמה מחוברים

שם אחד מהם $0$ (עפ"י המחלקים של $n$).

אז עכשיו אנחנו יודעים שזה בסיס אורתונורמלי. (כי ומרחב הפונקציות מ-$[n]$ ל-$\mathbb{C}$ יש ממד $n$ ויש גם $n$ קרקטרים)

שמחה וששון

אנחנו יודעים איך להציג פונקציה כצירוף ליניארי של קרקטרים!

$$f = \sum \alpha_j \chi_j$$

ובגלל ש-$\{\chi_j\}$ בסיס מתקיים

$$\alpha_j = \langle f, \chi_j \rangle$$

(לפי פירוש ההטלה על $\chi_j$).

הסימון המקובל ל-$\alpha_j$ הוא $\hat{f}(j)$. את $\hat{f}$ קוראים ~~טרנספורם~~ פורייה של $f$.

ניתן להמשיך ולרשום את $f$ ואת $\hat{f}$ כווקטורים $n$-ממדיים. ההעתקה $f \longmapsto \hat{f}$ היא ליניארית ומתקיים

$$\hat{f} = {}_j\left(\omega^{-jk}\right)^{k} f$$

המטריצה $M = (\omega^{-jk})$ היא אוניטרית (כלומר $MM^* = I$) שכן $\chi_k$ הם בא"נ. לכן ההעתקה $f \longmapsto \hat{f}$ היא איזומטריה וזו זהות Parseval: $\|\hat{f}\|_2 = \|f\|_2$

ע"י רישום זהות פרסבל ל-$f+g$ (ק) להסיק שבעצם נכון משהו כללי יותר - לכל $f,g:[n] \to \mathbb{C}$ מתקיים $\langle f,g \rangle = \langle \hat{f}, \hat{g} \rangle$ (ומשהו כזה נכון בשביל התכונה הכללית של טרנספורמציות אוניטריות של שמירה על זווית).

נעבור לדבר על המקרה $G = \mathbb{T}$ חבורת המעגל. עיקר העניין במקרה הזה הוא שניתן לחשוב על $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ ולכן פונקציות מרוכבות על $\mathbb{T}$ אפשר לראות כפונקציות מחזוריות על הישר (מחזור $2\pi$).

מהם הקרקטרים? אנחנו מצמצמים לקרקטרים שהם רציפים

(48)

והפונקציות מרוכבות של $\mathbb{T}$).

בניתוח דומה לזה שערכנו מקודם על $\mathbb{Z}_n$ מקבלים שהקרקטרים הרציפים היחידים הם $z \mapsto z^k$ לאיזה $k$ שלם.

מה שעלינו לבדוק הוא שאם $z_1, z_2 \in \mathbb{T}$ אז $\chi_k(z_1 z_2) = \chi_k(z_1)\chi_k(z_2)$ וזה בוודאי נכון. אז זה קרקטר רציף.

אמרנו שמרחב ההעתקות $\mathbb{T} \longrightarrow \mathbb{C}$ שקול למרחב ההעתקות המחזוריות על הישר, נאמר עם מחזור $1$

( $\mathbb{R} \longrightarrow \mathbb{T}$, $x \longmapsto e^{2\pi i x}$ למשל )

כאשר עושים את התהליך הדומה לזה שערכנו ב- $\mathbb{Z}_n$, מראים שהקרקטרים הם משפחה אורתונורמלית. תכונות השלמות שלה - לא בדיוק מידה נותן לומר שהצגנו פונקציה נתונה בבסיס של הקרקטרים - אלו שאלות עדינות ומורכבות שנבין מהן מעט.

התרגום הוא אם כן, כזה:

במרחב הפונקציות המרוכבות המחזוריות על הישר עם מחזור $1$, ישנו אוסף הפונקציות $x \longmapsto e^{2\pi i n x}$. הפונקציות האלה מהוות משפחה אורתונורמלית. לכל $f(x)$ היינו רוצים להתאים טור $\sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$ (כמובן נרצה שמושג ההתאמה הזו תהיה שוויון אותו אנחנו בכלל לא מגדירים עדיין פורמלית) אז הדבר הטבעי להגדיר היה ש- $\hat{f}(n) = a_n = \int_0^1 f(x) e^{-2\pi i n x}\, dx$

פורייה ראה עוד לדון בהתכנסות של הטור הנ"ל - האם בכלל יש התכנסות ואם כן, אז איזו?

Fejer שאל שאלה פשוטה יותר - האם המספרים $a_n = \hat{f}(n)$ מגדירים את $f$ (מהנחות קלות על $f$)? התשובה היא חיובית!

**משפט Fejer**: תהי $f:\mathbb{T}\to\mathbb{C}$ אינטגרבילית רימן ונניח ש-$f$ רציפה ב- $t\in\mathbb{T}$. אזי

$$\lim_{n\to\infty}\sigma_n(f,t)=f(t)$$

יתר על כן, אם $f$ רציפה אז ההתכנסות הנ"ל רציפה.

כאשר $\sigma_k(f,x)=\frac{1}{k}\left(S_1(f,x)+\dots+S_k(f,x)\right)$ !

סכום Cesaro

$$S_N(f,x)=\sum_{n=-N}^{N}\hat{f}(n)e^{2\pi inx}$$

(49) 3/2/09
טיפמתמטיים

שיעור אחרון ☺

# הרעיון היסודי של אנליזה הרמונית

יש הקשרים רבים שבהם אנו מעוניינים לחקור את מרחב הפונקציות הממשיות/מרוכבות $f: G \to \mathbb{R} | \mathbb{C}$ כאשר $G$ חבורה אבלית.
ספציפית, אם מוסיפים לקבוצה $[n]$ את המבנה החיבורי (מוד $n$) אז ניתן לראות את $\mathbb{R}^n, \mathbb{C}^n$ כאוסף כל הפונקציות $f: \mathbb{Z}_n \to \mathbb{R}^n | \mathbb{C}^n$.
יש למרחב הלינארי הנ"ל (של פונקציות $f: G \to \mathbb{R} | \mathbb{C}$) בסיס מיוחד של הקרקטרים. אם $G$ חבורה אבלית (אני עוסק רק בחבורות קומפקטיות מקומית) אז קרקטר $\chi$ של $G$ להיו הומומורפיזם $\chi: G \to \mathbb{T}$ כאשר $\mathbb{T} \subseteq \mathbb{C}$ זו חבורת המעגל, כ"א החבורה של כל המספרים המרוכבים בערך מוחלט 1, עם פעולת הכפל של מספרים מרוכבים.
התכונה הטובה הראשונה שגילינו בדבר הקרקטרים היא שהם מהווים מערכת אורתונורמלית. לכן, יש דרך פשוטה לבטא פונקציה כללית $f: G \to \mathbb{C}$ כ- $f = \sum \alpha_i \chi_i$ כאשר $\alpha_i \in \mathbb{C}$ ו- $\chi_i$ עוברים על כל הקרקטרים הרציפים. מקובל לסמן $\alpha_i = \hat{f}(i)$ ומכיוון ש-$\chi_i$ הם מע' אורתונורמלית, יש דרך פשוטה לחשב את $\alpha_i = \hat{f}(i)$ - $\hat{f}(i) = \langle f, \chi_i \rangle$.
כל מה שנאמר כאן הוא נכון ללא הסתייגות כש-$G$ סופית (למשל $G = \mathbb{Z}_n$ או $G = \{0,1\}^n$). במקרה הזה מספר הקרקטרים השונים של $G$ הוא בדיוק $|G|$ ולכן אוסף הקרקטרים מהווה בסיס אורתונורמלי למרחב הפונקציות $f: G \to \mathbb{C}$.
כשהחבורה $G$ אינסופית (בפרט $G = \mathbb{R}, \mathbb{T}$) יש כמה סיבוכים נוספים.
אולם הקרקטרים הרציפים של $G$ מהווים חבורה בפני עצמה שנקראת החבורה הדואלית. את החבורה הזו נהוג לסמן ב-$\hat{G}$.
מתכניים

$$\hat{\mathbb{T}} \cong (\mathbb{Z}, +)$$
$$\hat{\mathbb{R}} \cong (\mathbb{R}, +)$$

במקרה הסופי, אין קושי עם הטענה (*) $f = \sum \hat{f}(i)\chi_i$

אבל במקרה האינסופי עדיף ובעיתי $\sum \hat{f}(i)\chi_i$ הוא דבר לא

מובן מאליו. ויש צורך לברר באיזה מובן עשוי שוויון כמו (*) להיות

נכון.

במקרה הסופי $G=\mathbb{Z}_n$ ראינו שהקרקטרים הם $\chi_k:\mathbb{Z}_n\to\mathbb{T}$

כאשר $\chi_k(j)=\omega^{kj}$ ; $\omega=e^{2\pi i/n}$ שורש

היחידה ה-n. במקרה זה המעבר $f\to\hat{f}$ ניתן

ע"י הטרנספורמציה הלינארית $\hat{f}=\rho(\omega^{pq})f$ (זהו טרנספורם

פורייה הבדיד DFT).

באופן כללי, מכיוון שהקרקטרים מהווים בסיס אורתונורמלי

המטריצה $(\omega^{pq})$ או באופן כללי טרנספורם פורייה היא

איזומטריה ב-$\ell^2$ מתקבלת זהות Parseval שאומרת

$\|f\|_2=\|\hat{f}\|_2$. ובאופן כללי יותר, $\langle f,g\rangle=\langle \hat{f},\hat{g}\rangle$

התורה הזו החלה עם המקרה $G=\mathbb{T}$, (באופן שקול -

העתקות מחזוריות $\mathbb{R}\to\mathbb{C}$) והתורה של טורי פורייה ראשית,

במקרים כאלה זו בעיה לא פשוטה להבין מהי האינטרפרטציה

הנכונה של השוויון (*). כפי שראינו בפעם הקודמת, הקרקטרים

הרציפים של חבורת המעגל $\mathbb{T}$ הם $\chi_k(z)=z^k$ עבור $z\in\mathbb{T}$

ו-$k=0,\pm1,\pm2,\pm3,\ldots$. אם מעדיפים את נקודת המבט של פונקציות

מחזוריות על הישר מקבלים $\chi_k: x\mapsto e^{2\pi i k x}$

לאור הקשיים הכרוכים בניסוח נאות של השוויון (*) אפשר לשאול

שאלה נקודתית: האם $\hat{f}$ קובע את $f$?

<u>משפט</u>: אם $f,g:\mathbb{T}\to\mathbb{C}$ רציפות ואם לכל $k\in\mathbb{Z}$ $\hat{f}(k)=\hat{g}(k)$

אז $f=g$ (כאשר $\langle f,\chi_k\rangle=\hat{f}(k)=\frac{1}{2\pi}\int_{t=0}^{2\pi} f(t)e^{-2\pi i k t}\,dt$)

הרעיון הבסיסי של משפט Fejer הוא להחליף את הסכומים

$$S(f,n)(x)=\sum_{k=-n}^{n}\hat{f}(k)\chi_k(x)=\sum_{k=-n}^{n}\hat{f}(k)e^{-2\pi i k x}$$

בסדרה המתכנסת "יותר טוב" - סכומי Cesaro

<u>תזכורת</u>: אם הטור $\sum a_n = a$ מתכנס אז סדרת
הממוצעים $b_k = \frac{1}{k}(S_1+\dots+S_k)$ מתכנסת לאותו גבול $a$.
יש דוגמאות פשוטות שבהן הסדרה $S_N$ לא מתכנסת אבל
ש- $b_N$ כן מתכנסת..

<u>משפט Fejer</u>: אם $f:\mathbb{T}\to\mathbb{C}$ אינטגרבילית רימן
ואם $f$ רציפה בנקודה $t\in\mathbb{T}$ אז $\lim \sigma_n(f,t) = f(t)$
כאשר $\sigma_n(f,t) = \frac{1}{n+1}\sum_{l=0}^{n} S(t,l)(t)$

מרכיב חשוב נוסף בהוכחה ולהתעניין בה עתה $G \to \mathbb{C}$
הוא שניתן עתה להגדיר את: מושג הקונבולוציה שזו
פעולה חשובה בין פונקציות.
אם $f,g: G\to\mathbb{C}$, אז הקונבולוציה שלהן $h$ המסומנת
$h = f*g$ היא פונקציה $h: G\to\mathbb{C}$ המוגדרת ע"י

$$h(x) = \sum_{t\in G} f(t)\,g(x-t) = \sum_{t\in G} f(x-t)\,g(t)$$

למשל, אם $G=\mathbb{Z}_n$ ! -

$$g(x) = \begin{cases} 1/4 & x=-1 \\ 1/2 & x=0 \\ 1/4 & x=1 \end{cases}$$

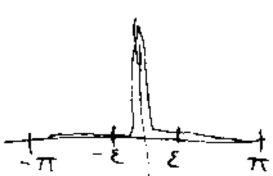
אז $h$ זו החלקה של $f$ המתקבלת ממיצוע עם שני השכנים.
תכונה חשובה היא $\widehat{f*g} = \hat{f}\cdot\hat{g}$.

כמה מילים על ההוכחה של משפט Fejer:
ע"י מניפולציות אלגבריות פשוטות מגיעים בחישוב ישיר לכך ש-

$$\sigma_n(f,t) = f * K_n$$

כאשר

$$K_n(s) = \begin{cases} \frac{1}{n+1}\left(\frac{\sin\frac{(n+1)s}{2}}{\sin\frac{s}{2}}\right)^2 & \text{אחרת} \\ n+1 & s=0 \end{cases}$$

הפונקציה הנ"ל נראית כך  והקונבולוציה עם
פונקציה כזאת.. השפט ונותנת את הערך הרצוי.

$$h(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)K(t)\,dt \approx \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} f(x-t)\cancel{K(t)}\,dt \approx f(x)\cdot\cancel{\frac{2\varepsilon}{2\varepsilon}}$$

מסקנה ממשפט Fejer היא משפט ויירשטראס:

לכל $f:[a,b]\to\mathbb{R}$ רציפה ולכל $\varepsilon>0$ יש פולינום $P$ כך

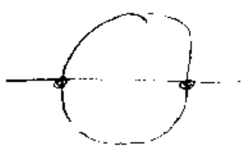
ש- $\max\limits_{x\in[a,b]} |f(x)-P(x)| = \|f-P\|_\infty < \varepsilon$

בה"כ $[a,b]=[-1,0]$. נשקף את $f$ לקטע $[0,1]$ ונמשיך את $f$ מחזורית. נקרב את הפונקציה המחזורית ע"פ משפט Fejer ונקבל את המשפט. (נזכור שהקרקטרים הם חזקות!)

נעבור לחבורה $G=\{0,1\}^n$ עם חיבור קואורדינטה קואורדינטה מודולו 2. אז $|G|=2^n$ ואנחנו מצפים שיהיו $2^n$ קרקטרים. ברור שבו למעשה $\mathbb{Z}_2^n$ או אנחנו גם מצפים שהקרקטרים של $G$ יהיו קשורים איכשהו לקרקטרים של $\mathbb{Z}_2$.

ב-$\mathbb{Z}_2$ יש שני קרקטרים - $\chi(x)=1$

או $\chi(x)=(-1)^x$



כדי לבנות את הקרקטרים של $\mathbb{Z}_2^n$ (נכיר את הקרקטרים החד-ממדיים האלה). יוצא שכדי להגדיר קרקטר של $\mathbb{Z}_2^n$ יש לציין רק בכל קואורדינטה באיזה משני הקרקטרים של $\mathbb{Z}_2$ בחרנו. לכן גם הקרקטרים יצוינו ע"י וקטורים ב-$\{0,1\}^n$ באופן הבא:

אם $u\in\{0,1\}^n$ אז $\chi_u(x)=(-1)^{\langle x,u\rangle}$

לפעמים מציינים את $u,x$ כתת-קבוצות של $[n]$ ואז $u$ מתואם ע"י $S$ ו-$x$ ע"י $T$ ואז $\chi_S(T)=(-1)^{|S\cap T|}$.

נתבונן בשאלה יסודית בתורה של קודים לתיקון שגיאות (Error Correcting Codes) ואיך ניתן, ע"י שימוש באנליזה הרמונית, לקבל חסמים עליונים על הגודל של קודים.

אנו משדרים על פני ערוץ רועש. היחידה הבסיסית שאנו משדרים היא מילה של ביטים. בגלל רעש בערוץ, אם שלחנו את $x\in\{0,1\}^n$, עלולה להגיע למקלט התשדורת $x\oplus e$ כאשר $e$ וקטור הרעש.

קוד לתיקון שגיאות זו קבוצה $C\subseteq\{0,1\}^n$. המרחק

של קוד רבה מוגדר כ-

$$dist(C) = \min_{\substack{x\neq y\\ x,y\in C}} \{ d_H(x,y) \}$$

מרחק Hamming (=כמות קואורדינטות הם שונים)

בהתקבל המילה $z$ במקום (פענוח) אותה

כ- $x\in C$ כך שהמרחק $d(x,z)$ מינימלי. כלומר

$$d(x,z) = \min_{y\in C} d(y,z)$$

קל לראות שאם $dist(C)=d$ או אנו נפענח נכון בכל מקרה שבו חלו פחות מ- $\frac{d}{2}$ שגיאות.

זה מעורר את השאלה הטבעית: לחשב או להעריך את הגודל המירבי של קוד $C\subseteq\{0,1\}^n$ עם מרחק $\geq d$.

$$A(n,d) = \max\{|C| : C\subseteq\{0,1\}^n,\ dist(C)\geq d\}$$

כפי שאמרנו, לכל פונקציה $f:\{0,1\}^n\to\mathbb{R}$ יש ייצוג מהצורה

$$f(x) = \sum_{u\in\mathbb{Z}_2^n}\hat{f}(u)\chi_u(x) = \sum_u \hat{f}(u)(-1)^{\langle u,x\rangle}$$

כאשר

$$\hat{f}(u) = \frac{1}{2^n}\sum_y f(y)(-1)^{\langle u,y\rangle}$$

נבדוק שאכן ה- $\chi_u$ מהווים בסיס אורתונורמלי וספציפית הנורמה כן הוא תקין.

$$\langle\chi_u,\chi_v\rangle = \frac{1}{2^n}\sum_y \chi_u(y)(-1)^{\langle v,y\rangle} =$$

$$= \frac{1}{2^n}\sum_y(-1)^{\langle u,y\rangle}(-1)^{\langle v,y\rangle} =$$

$$= \frac{1}{2^n}\sum_y(-1)^{\langle u+v,y\rangle} = \begin{cases}1 & u=v\\ 0 & u\neq v\end{cases}$$

כי אם $u=v$ אז $u+v=\vec{0}$. ואם $u\neq v$ אז $u+v=w\neq 0$. הסכום נוכל לחלק את ה-$y$-ים לזוגות זוגות כך שכל זוג שונה בקואורדינטה אחת. אז בכל זוג הסכום אפס ולכן כל הסכום אפס.

הקשר בין טרנספורם פוריה וקונבולוציה: $\widehat{f*g} = \hat{f}\cdot\hat{g}$

היות שאנו יכולים לפתח את $f,g$ בבסיס הקרקטרים

די יהיה להוכיח את הזהות הרצויה כש- $f=\chi_u,\ g=\chi_v$.

$$\hat{\chi}_u(w) = \begin{cases} 1 & w=u \\ 0 & \text{אחרת} \end{cases}$$

הטענה $\widehat{\chi_u*\chi_v} = \hat{\chi}_u\cdot\hat{\chi}_v$ אומרת למעשה ש.

$$\widehat{\chi_u*\chi_v} = \begin{cases} 0 & u\neq v \\ \hat{\chi}_u & \text{אחרת} \end{cases}$$

השאלה היא אם כן, מהו פיתוח פוריה של $\chi_u*\chi_v$.

$$\chi_u*\chi_v(x) = \frac{1}{2^n}\sum_t \chi_u(t)\,\chi_v(x+t) =$$

ב- $\mathbb{Z}_2$ + = - להם.

$$= \frac{1}{2^n}\sum_t (-1)^{\langle u,t\rangle}(-1)^{\langle v,x+t\rangle} =$$

$$= (-1)^{\langle v,x\rangle}\underbrace{\frac{1}{2^n}\sum_t (-1)^{\langle u+v,t\rangle}}_{\substack{=0 \text{ אם } u\neq v \\ =1 \text{ אם } u=v}} =$$

$$= \delta_{u,v}(-1)^{\langle v,x\rangle} = \delta_{u,v}\,\chi_u(x)$$

וזה מה שרצינו.

התכנית הלינארית של Delsarte. תהי $C\subseteq\{0,1\}^n$.

נסמן ב- $g=\mathbb{1}_C$ את הפונקציה האופיינית של $C$.

ונסמן $f=g*g$.

$$f(u) = \frac{1}{2^n}\sum_t g(t)\,g(t+u) = \frac{1}{2^n}\sum_{\substack{s,t\in C \\ s\oplus t=u}} 1 =$$

$$= \frac{1}{2^n}\left\{\begin{matrix}\text{מספר הזוגות של מילים} \\ s,t\in C \\ s+t=u \text{ רק ש- }\end{matrix}\right\}$$

$d_H(x,y)=|x\oplus y|$ אם כן

קיבלנו ש- $dist(C)\geq d$ אמ"מ $f(u)=0$ לכל $u$

שמשקל Hamming שלו הוא $1 \le |u| \le d-1$.

אבל $\hat{f} = g * \hat{g} = \hat{g} \cdot \hat{g} = |\hat{g}|^2 \ge 0$

מתקיים $f(0) = \frac{|C|}{2^n}$. נסמן $F = \frac{2^n}{|C|} f$. אז

$$\sum_u f(u) = \frac{1}{2^n} |C|^2$$

$$\sum F(u) = |C|$$

<u>מסקנה:</u> אם $C \subseteq \{0,1\}^n$; $dist(C) \ge d$ אז $|C|$ אינו גדול מן האופטימום של התכנית הלינארית הבאה:

$$|C| \le \max_{s.t} \sum F(u)$$

$$\hat{F} \ge 0$$

$$F(0) = 1$$

$$\forall\, 1 \le |x| \le d-1 \quad F(x) = 0$$

זה תנאי לינארי כי המעבר $F \to \hat{F}$ הוא לינארי.

לא קשה להראות שבה"כ $F(x)$ תלוי רק ב- $|x|$. זה מקטינים את ה- LP מגודל $2^n$ ל- $n+1$.

סוף