תורת הסיבוכיות – תרגול 13
הוכחות ניתנות לבדיקה הסתברותית

התרגול יוקדש כולו לנושא של **הוכחות ניתנות לבדיקה הסתברותית**, באנגלית **Probabilistically Checkable Proofs** (בקיצור **PCP**).

תזכורת

מוודא PCP הוא מ"ט הסתברותית פולינומית V בעלת גישה אקראית להוכחה π הכתובה ב**סרט הוכחה**, כאשר הגישה מתבצעת באמצעות **סרט שאלה** ו־ **סרט תשובה** באופן הבא: בהינתן קלט x, V בוחר קבוצת אינדקסים S שהיה רוצה לקרוא מתוך ההוכחה π, רושם את S בסרט השאלה, ומקבל בסרט התשובה את $\pi[i]$ לכל $i \in S$; על סמך הקלט ומקטע ההוכחה שנקרא, V מבצע חישוב שבסופו הוא מקבל או דוחה.

בהינתן שפה L ופונקציות $r, q \colon \mathbb{N} \to \mathbb{N}$, אנו אומרים כי V הוא $(r(n), q(n))$**־מוודא** עבור L אם V משתמש ב־ $O(r(n))$ ביטים של אקראיות, דוגם $O(q(n))$ ביטים מההוכחה, ומקיים:

- שלמות: אם $x \in L$ אז קיימת הוכחה π_x כך ש־ $\mathbf{Pr}[V^{\pi_x}(x) = \mathrm{acc}] = 1$.
- נאותות: אם $x \notin L$ אז לכל הוכחה π מתקיים $\mathbf{Pr}[V^{\pi}(x) = \mathrm{acc}] \leq \frac{1}{2}$.

נסמן ב־ $\mathsf{PCP}(r(n), q(n))$ את אוסף כל השפות עבורן קיים $(r(n), q(n))$־מוודא.

משפט (משפט ה־ PCP, 1992): $\mathsf{NP} = \mathsf{PCP}(\log n, 1)$.

משפט ה־ PCP נותן אפיון נוסף ל־ NP כאוסף כל השפות שניתן לבדוק את נכונותן באופן הסתברותי בזמן פולינומי תוך שימוש ב־ $O(\log n)$ ביטים אקראיים ודגימה של ההוכחה ב־ $O(1)$ מקומות בלבד.

המשפט, שנחשב לפורץ דרך והוא אבן יסוד של חקר הקושי של קירוב, הוא תוצר של שורת עבודות, והוכחתו ניתנה ב־ 1992 בזוג מאמרים שנכתבו על ידי Arora, Safra, Lund, Motwani, Sudan, ו־ Szegedy. בשנת 2006 Irit Dinur פרסמה הוכחה המפשטת ומקצרת את הוכחתו המקורית של המשפט.

וריאנטים של ההגדרה

מוודא אדפטיבי

ניתן להבחין בין מוודא PCP שאינו **אדפטיבי** ובוחר אילו ביטים לקרוא מההוכחה על סמך הקלט והאקראיות בלבד, לעומת מוודא אדפטיבי שיכול לבחור אילו ביטים לקרוא הלאה על סמך ביטים קודמים שנקראו.

למרות שכברירת מחדל אנו מניחים שהמוודא אינו אדפטיבי, נבחן את כוחו של וידוא אדפטיבי בתרגיל הבא.

תרגיל: הוכיחו כי לכל שפה L שיש עבורה מוודא PCP אדפטיבי שמשתמש ב־ r ביטים אקראיים וקורא q ביטים מההוכחה, יש עבורה גם מוודא PCP רגיל (לא אדפטיבי) שמשתמש ב־ r ביטים אקראיים וקורא 2^q ביטים מההוכחה.

פתרון: כל פעם שהמוודא האדפטיבי קורא ביט מההוכחה, ערך הביט קובע (יחד עם היסטוריית החישוב) מהו האינדקס של הביט הבא שיש לקרוא מההוכחה (בהינתן שקבענו מחרוזת אקראיות כמובן). מאחר ולכל ביט 2 ערכים אפשריים, ישנו אוסף אינדקסים מגודל לכל היותר 2^q שהמוודא האדפטיבי עלול לקרוא במהלך ריצתו.

אם כן, מוודא רגיל יכול תחילה לסמלץ את המוודא האדפטיבי מול כל מחרוזות תשובות אפשרית מאורך q, למצוא את כל האינדקסים שייתכן והמוודא האדפטיבי ירצה לקרוא, לקרוא אותם מסרט ההוכחה (של המוודא הרגיל), ולבסוף להריץ את המוודא האדפטיבי ולענות כמוהו.

השלמות והנכונות נובעים מאלו של המוודא האדפטיבי.

תנאי שונים לשלמות ונאותות

המחלקה $\mathsf{PCP}\left(r\left(n\right),q\left(n\right)\right)$ הוגדרה עם שלמות מלאה ונאותות עם הסתברות קבלה של לכל היותר $\frac{1}{2}$. בחלק זה נבחן מה קורה כאשר משנים את ספי הקבלה שבתנאי השלמות והנאותות.

לצורך כך נוסיף סימון: עבור קבועים $c,s\in[0,1]$ שמייצגים את דרישת השלמות והנאותות בהתאמה, נסמן ב־ $\mathsf{PCP}_{(c,s)}\left(r\left(n\right),q\left(n\right)\right)$ את מחלקת השפות עבורן קיים $\left(r\left(n\right),q\left(n\right)\right)$-מוודא עם הדרישות הבאות:

- <u>שלמות</u>: אם $x\in L$ אז קיימת הוכחה π_x כך ש־ $\mathbf{Pr}\left[V^{\pi_x}\left(x\right)=\mathrm{acc}\right]\geq c$.
- <u>נאותות</u>: אם $x\notin L$ אז לכל הוכחה π מתקיים $\mathbf{Pr}\left[V^{\pi}\left(x\right)=\mathrm{acc}\right]\leq s$.

<u>תרגיל</u>: הוכיחו את הטענות הבאות:

1. $\mathsf{MA}\subseteq\mathsf{PCP}_{\left(\frac{3}{4},\frac{1}{4}\right)}\left(\mathsf{poly}\left(n\right),\mathsf{poly}\left(n\right)\right)$.
2. $\mathsf{PCP}_{\left(\frac{3}{4},\frac{1}{4}\right)}\left(\mathsf{poly}\left(n\right),\mathsf{poly}\left(n\right)\right)\subseteq\mathsf{NEXP}$.
3. $\mathsf{IP}\left[2\right]\subseteq\mathsf{PCP}_{\left(1,\frac{1}{2}\right)}\left(\mathsf{poly}\left(n\right),\mathsf{poly}\left(n\right)\right)$.

<u>פתרון</u>:

1. תהא $L\in\mathsf{MA}$. אזי, קיים עבור L פרוטוקול מרלין-ארתור מתאים.

כזכור, פרוטוקול מרלין-ארתור עם סיבוב אחד מקיים שמרלין, המסומן ב־ M, הוא כל יכול, וארתור, המסומן ב־ A, הוא הסתברותי פולינומי, ועל קלט x הפרוטוקול מתנהל כך: מרלין שולח לארתור הודעה $y=M\left(x\right)$, ואז ארתור מחליט האם לקבל או לדחות על סמך חישוב הסתברותי $A\left(x,r,y\right)$ שתלוי בקלט, באקראיות, ובהודעה של מרלין; כמו כן, על המערכת לקיים את תנאי השלמות והנאותות הבאים:

$$x\in L\quad\Longrightarrow\quad\exists\,M:\ \mathbf{Pr}\left[A\leftrightarrow M\text{ accepts }x\right]\geq\frac{3}{4}$$
$$x\notin L\quad\Longrightarrow\quad\forall\,M':\ \mathbf{Pr}\left[A\leftrightarrow M'\text{ accepts }x\right]\leq\frac{1}{4}$$

רעיון ההוכחה פשוט: נבנה מוודא V PCP שמשתמש בהודעה של מרלין כהוכחה, שאותה הוא יקרא בכללותה בתחילת ריצתו. אם זמן הריצה של A חסום על ידי הפולינום $p\left(n\right)$, אז V יבקש לקרוא מההוכחה, שאורכה לכל היותר $p\left(n\right)$, את האינדקסים $1,\ldots,p\left(n\right)$ (כלומר הכל), ולאחר מכן ידמה את ריצת A כשהוא מציב ב־ y את תוכן הודעת המוודא.

קל לראות ש־ V פולינומי. כמו כן, V משתמש במספר פולינומי של ביטים אקראיים (לכל היותר $p\left(n\right)$), והוא קורא מספר פולינומי של ביטים מההוכחה. השלמות והנאותות של V נובעות באופן מיידי מהשלמות והנאותות של $\left(A,M\right)$.

לכן, לפי הגדרה, $L\in\mathsf{PCP}_{\left(\frac{3}{4},\frac{1}{4}\right)}\left(\mathsf{poly}\left(n\right),\mathsf{poly}\left(n\right)\right)$.

2. תהא $L\in\mathsf{PCP}_{\left(\frac{3}{4},\frac{1}{4}\right)}\left(\mathsf{poly}\left(n\right),\mathsf{poly}\left(n\right)\right)$, ויהי V מוודא PCP מתאים. נסמן ב־ $p\left(n\right)$ חסם פולינומי הן עבור זמן ריצת V והן עבור מספר הביטים האקראיים בהם V משתמש.

זה תרגיל טוב לנסות תחילה להבין מדוע הפתרון הבא לא עובד: ננסה להראות משהו חזק יותר ממה שנדרש בתרגיל, ש־ $L\in\mathsf{IP}\left[2\right]$. נציע פרוטוקול הוכחה אינטרקטיבי בעל שני סיבובים עם מוודא $\widetilde{V}$, שעל קלט x פועל כך: מוכיח הגון יוכל למצוא את π_x, ולכן $\widetilde{V}$ ישלח למוכיח שלו את רשימת האינדקסים ש־ V חפץ לקרוא מההוכחה, ואז $\widetilde{V}$ ידמה את יתר ריצת V ויענה כמוהו.

<u>רעיון ההוכחה</u>: ננחש הוכחה, נדמה את ריצת V על כל מחרוזות האקראיות האפשריות, ונקבל אם ורק אם V קיבל עבור מספר גדול מספיק של מחרוזות אקראיות.

באופן פורמלי, נבנה מ"ט א"ד M שעל קלט x נוהגת כך:

- מנחשת הוכחה (מחרוזת בינארית) π מאורך $2^{p(n)}$, ומאתחלת מונה $c\leftarrow0$.
- לכל מחרוזת אקראיות $r\in\{0,1\}^{p(n)}$:
  - מדמה את V עם אקראיות r, כאשר הביטים ש־ V רוצה לקרוא נלקחים מ־ π.
  - אם V קיבל, מעדכנת $c\leftarrow c+1$.
- מקבלת אם ורק אם $c\geq\frac{3}{4}\cdot2^{p(n)}$.

<u>זמן ריצה</u>: ניחוש π דורש זמן $O\left(\pi\right)=2^{O(p(n))}$. כמו כן, למעבר על כל מחרוזות האקראיות וסימולציית V נדרש זמן הנתון לפי $2^{p(n)}\cdot O\left(p\left(n\right)\cdot\log p\left(n\right)\right)=2^{O(p(n))}$. בסך הכל, זמן ריצת M אקספוננציאלי.

<u>נכונות</u>:

- נניח ש־ $x \in L$. אזי, קיימת הוכחה π_x שמשכנעת את V בהסתברות $\frac{3}{4}$ לפחות, ולכן קיים ניחוש של M שבו לפחות $\frac{3}{4}$ מהסימולציות של V תסתיימנה במצב מקבל, ולכן בתום הסימולציות יתקיים $c \geq \frac{3}{4} \cdot 2^{p(n)}$ ו־ M תקבל. לכן, קיים מסלול חישוב שבו M מקבלת.
- נניח ש־ $x \notin L$. אזי, כל הוכחה π משכנעת את V בהסתברות $\frac{1}{4}$ לכל היותר, ולכן בכל ניחוש של M לכל היותר $\frac{1}{4}$ מהסימוצליות של V תסתיימנה במצב מקבל, ולכן בתום הסימולציות יתקיים $c \leq \frac{1}{4} \cdot 2^{p(n)}$. לכן, M דוחה בכל מסלול חישוב שלה.

לכן, לפי הגדרה, $L \in \mathsf{NEXP}$.

3. נעיר תחילה שסעיף זה גורר את הסעיף הראשון, שכן:

$$\mathsf{MA} \subseteq \mathsf{AM} = \mathsf{IP}[2] \underset{.3}{\subseteq} \mathsf{PCP}_{\left(1,\frac{1}{2}\right)}(\mathsf{poly}(n), \mathsf{poly}(n)) \underset{\text{amplification}}{\subseteq} \mathsf{PCP}_{\left(\frac{3}{4},\frac{1}{4}\right)}(\mathsf{poly}(n), \mathsf{poly}(n))$$

תהא $L \in \mathsf{IP}[2]$, ויהי (V,P) פרוטוקול הוכחה אינטראקטיבית בעל שני סיבובים עבור L, עם שלמות מלאה ונאותות עם התסברות קבלה לכל היותר $\frac{1}{2}$. יהי $p(n)$ פולינום החוסם את זמן ריצת V.

בפרוטוקול כזה, V שולח למוכיח הודעה q, מקבל מהמוכיח תשובה a, ולבסוף בשביל להחליט אם לקבל או לדחות V מבצע חישוב שתלוי בקלט, באקראיות, ובתשובת המוכיח.

<u>רעיון</u>: חישוב פשוט מראה שעבור קלטים מאורך n סך כל הודעות האפשריות מאורך לכל היותר $p(n)$ ש־ V יכול לשלוח למוכיח שלו חסום על ידי $B \triangleq 2^{p(n)+1} - 1$. לכן, מוודא PCP יוכל לחשוב על ההוכחה π כעל טבלה עם B כניסות, כאשר כל כניסה היא מאורך $p(n)$ ומכילה את תשובת המוכיח בתגובה לשאילתה שמתאימה לכניסה. שימו לב שאורך הטבלה הוא $B \cdot p(n) = 2^{O(p(n))}$, ולכן ניתן לייצג כל אינדקס של הטבלה באמצעות $O(p(n))$ ביטים.

אם כן, נבנה מוודא PCP $\widehat{V}$ שעל קלט x פועל באופן הבא:

- מגריל מחרוזת אקראיות $r \in \{0,1\}^n$, ומסמלץ את $V(x,r)$.
- כאשר V שולח למוכיח הודעה q, $\widehat{V}$ יחשב עבור המחרוזת q את הכניסה המתאימה בטבלה i_q, ויבקש לקרוא מההוכחה את האינדקסים $i_q \cdot p(n)+1, \ldots, (i_q+1) \cdot p(n)$.
- מכאן $\widehat{V}$ ימשיך לסמלץ את V, ולבסוף יענה כמוהו.

קל לראות שזמן ריצת $\widehat{V}$ פולינומי (לכן גם מספר הביטים ש־ $\widehat{V}$ קורא מההוכחה פולינומי), ומספר הביטים האקראיים בהם הוא משתמש פולינומי. כמו כן, תנאי השלמות והנכונות נובעים מאלו של (V,P), ונשאיר את הפירמול שלהם כתרגיל.