שמות המגישים: רועי לייבוביץ' ; 206994840 ; roy.leibovici@mail.tau.ac.il
גולצמן ; 314645920 ; goltsman@mail.tau.ac.il

תרגיל בית 2 – מבוא לקריפטוגרפיה

Question 1

1. נגדיר את המרחב הבא:

$$S = \{0, \dots, 2k-1\}$$

$$A = \{CW, CCW\}$$

$$p^{(CCW)}_{ss'} = \begin{cases} 1, & s' = s+1 \bmod (2k) \\ 0, & s' = s-1 \bmod (2k) \end{cases}$$

$$p^{(CW)}_{ss'} = \begin{cases} 1, & s' = s-1 \bmod (2k) \\ 0, & s' = s+1 \bmod (2k) \end{cases}$$

$$R = \begin{cases} 1, & s = 0 \\ 0, & else \end{cases}$$

$$s_0 = k$$

2. המדינית האופטימלית π^* היא לבצע K פעולות CW או CCW עד להגעה למצב $s=0$ ולאחר מכן להתחיל לבצע CW אחר CCW או CCW אחר CW, ז"א לסרוגין לצאת ממצב $s=0$ וישירות לחזור אליו אחרי כן.

$$\pi^*(s=s') = \begin{cases} CW, & w.p.\ 0.5 \\ CCW, & w.p.\ 0.5 \end{cases}, \quad s' = \{s_0, 0\}$$

$$\pi^*(s=s') = CCW, \quad s' = \{s_0 + 1 \bmod (2K), \ldots, 2K-1\}$$

$$\pi^*(s=s') = CW, \quad s' = \{1, \ldots, s_0 - 1 \bmod (2K)\}$$

3. עבור $s = S \setminus \{0\}$ נקבל -

$$V_1(s) = \max_{a \in A} \{ \underbrace{r(s,a)}_{=0} + \gamma \sum_{s' \in S} p(s'|s,a) \underbrace{V_n(s')}_{=0} \}$$

$$V_1(s) = 0$$

עבור $s=0$ נקבל -

$$V_1(s=0) = \max_{a \in A} \{ \underbrace{r(0,a)}_{=1} + \gamma \sum_{s' \in S} p(s'|0,a) \underbrace{V_n(s')}_{=0} \}$$

$$V_1(s=0) = 1$$

- לכל $S = S \setminus \{0, 1, 2k-1\}$ מתקיים .4

$$V_2(s) = \max_{a \in A} \{ \underbrace{r(s,a)}_{=0} + \gamma \sum_{s' \in S} \underbrace{p(s'|s,a) V_n(s')}_{=0} \}$$

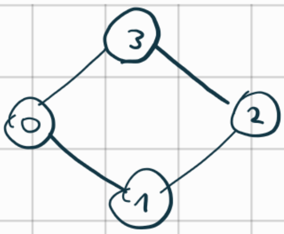
$$V_2(s) = 0$$

- לכל $S = \{0, 1, 2k-1\}$ מתקיים

$$V_2(s=0) = \max_{a \in A} \{ \underbrace{r(0,a)}_{=1} + \gamma \sum_{s' \in S} p(s'|0,a) \underbrace{V_n(s')}_{=0} \} = 1$$

$$V_2(s=1) = \max_{a \in A} \{ \underbrace{r(1,a)}_{=0} + \gamma \sum_{s' \in S} \underbrace{p(s'|1,a) V_n(s')}_{= \begin{cases} 1, & s'=0 \\ 0, & \text{o.w.} \end{cases}} \} = \gamma$$

$$V_2(s=2k-1) = \max_{a \in A} \{ \underbrace{r(2k-1,a)}_{=0} + \gamma \sum_{s' \in S} \underbrace{p(s'|2k-1,a) V_n(s')}_{= \begin{cases} 1, & s'=0 \\ 0, & \text{o.w.} \end{cases}} \} = \gamma$$



נבדוק 4 איטרציות ראשונות לפני האינדוקציה ואז נגלה:

$V_2(0) = 1$	$V_3(0) = 1+\gamma^2$	$V_4(0) = 1+\gamma^2$
$V_2(1) = \gamma$	$V_3(1) = \gamma$	$V_4(1) = \gamma(1+\gamma^2)$
$V_2(2) = 0$	$V_3(2) = \gamma^2$	$V_4(2) = \gamma^2$
$V_2(3) = \gamma$	$V_3(3) = \gamma$	$V_4(3) = \gamma(1+\gamma^2)$

$$V_n(0) = 1+\gamma\cdot V_{n-1}(1)$$

$$V_n(1) = V_n(3) = \gamma\cdot V_{n-1}(0)$$

$$V_n(2) = \gamma\cdot V_{n-1}(1)$$

סביר ש-$n\to\infty$ מתקבל כי

$$V_{n-1}(s) = V_n(s)$$

ולכן ניתן להציב ולקבל מערכת משוואות לינאריות.

$$V_n(0) = 1+\gamma V_n(1)$$

$$V_n(3) = V_n(1) = \gamma V_n(0) = \gamma(1+\gamma V_n(1))$$

$$V_n(2) = \gamma V_n(1)$$

$$\gamma+\gamma^2 V_n(1) = V_n(1)$$

$$V_n(1)\left[\gamma^2-1\right] = -\gamma$$

$$\boxed{V_n(1) = \frac{\gamma}{1-\gamma^2} = V_n(3)}$$

$$\boxed{V_n(2) = \frac{\gamma^2}{1-\gamma^2}}$$

$$\boxed{V_n(0) = 1+\frac{\gamma^2}{1-\gamma^2} = \frac{1}{1-\gamma^2}}$$

Question 2

1.

$$S=\left\{\vec{s}=(d_{N+1},d_N,\dots,d_0)^T \;\middle|\; 0\le i\le N,\ d_i\in\{0,1\},\ d_{N+1}\in\{0,\dots,9\}\right\}$$

כל מצב מיוצג מחרוזת בינרית המתארת אילו כלים של תא (0 - פנוי, 1 - מאוכלס) ואת הספרה שהתקבלה בהגרלה.

$$A=\{0,\dots,N\}$$

$$P^{(a)}_{ss'}=\begin{cases}\frac{1}{10}, & d'_j=\begin{cases}d_j, & \forall 0\le j\le N,\ j\ne a\\ 1, & j=a\\ k, & j=N+1\end{cases}\\ 0, & \text{else}\end{cases}\quad, a\in A,\quad k\in[0,\dots,9]$$

(k:) הערך שהתקבל בהגרלה למצב החדש s'.

$$r(s,a)=\frac{d_{N+1}\cdot 10^a}{9\cdot 10^N}=\frac{d_{N+1}}{9}\cdot 10^{a-N},\quad a\in A$$

הגמול נקבע ביחס ישר עם בחירת המיקום לספרה שהוגרלה.

בנוסף היא מנורמלת בבחירה המניבה גמול מירבי.

$$s_0=(0,\dots,0,k)^T,\ k\sim \text{Uniform}[0,9]$$

2. המדיניות האופטימלית $\pi(s)$ מתקבלת ממקס $V(s)$ המחושב כך:

$$V(s) = \max_{a \in A} \left\{ r(s,a) + \sum_{s' \in S} P(s'|s,a) \cdot V(s') \right\}$$

נוכיח באינדוקציה כי $\pi(s)$ תלוי אך ורק במספר התאים הריקים.

עבור 0 תאים ריקים, המשחק נגמר - מקרה טריוויאלי.

עבור תא ריק בודד, נמקם את הספרה החדשה d_{n+1} בתא הריק.

<u>בסיס האינדוקציה</u> - <u>עבור שני תאים ריקים:</u>

$$S = (1,1,1 \ldots 0 \ldots 1,1 \ldots 0 \ldots 1,1,d_{n+1})$$

(ה-0 הראשון: אינדקס i, ה-0 השני: אינדקס j)

נחשב את $V(s)$ לפי שני המסלולים האפשריים - $a = \{i,j\}$

$$V(s) = \max \left\{ d_{n+1} \cdot 10^i + \sum_{k=0}^{9} \frac{1}{10} \cdot V(1,1,1 \ldots \underset{j}{0} \ldots 1,1,k),\ d_{n+1} \cdot 10^j + \sum_{k=0}^{9} \frac{1}{10} \cdot V(1,1,1 \ldots \underset{i}{0} \ldots 1,1,k) \right\}$$

(מתחת ל-0: אינדקס j, אינדקס i)

$$V(1,1,1 \ldots \underset{j}{0} \ldots 1,1,d'_{n+1}) = \frac{k \cdot 10^j}{9 \cdot 10^N} = \frac{k \cdot 10^j}{t}$$

(t - פקטור נרמול)

נגדיר ונקבל:

$$V(s) = \max \left\{ d_{n+1} \cdot \frac{10^i}{t} + \sum_{k=0}^{9} \frac{1}{10} \cdot k \cdot \frac{10^j}{t},\ d_{n+1} \cdot \frac{10^j}{t} + \sum_{k=0}^{9} \frac{1}{10} \cdot k \cdot \frac{10^i}{t} \right\}$$

$$V(s) = \max \left\{ d_{n+1} \cdot \frac{10^i}{t} + \frac{10^j}{t} \cdot 4.5,\ d_{n+1} \cdot \frac{10^j}{t} + \frac{10^i}{t} \cdot 4.5 \right\}$$

נחפש decision regions :

$$d_{n+1}\cdot\frac{10^i}{t} + 4.5\cdot\frac{10^j}{t} = d_{n+1}\cdot\frac{10^j}{t} + 4.5\cdot\frac{10^i}{t}$$

$$\frac{1}{t}\cdot d_{n+1}\left[10^i - 10^j\right] = \frac{1}{t}\,4.5\left[10^i - 10^j\right]$$

$$\boxed{d_{n+1} = 4.5}$$

בשלב זה ראינו כי ה Threshold של d_{n+1} לא תלוי ב i,j עבור שני תאים ריקים.

המיקום האופטימלי לשני תאים ריקים הינו

$$q = \begin{cases} \max\{i,j\} & d_{n+1} \ge 5 \\ \min\{i,j\} & d_{n+1} < 5 \end{cases}$$

צעד האינדוקציה – נניח כי עבור k תאים פנויים המיקום האופטימלי לא תלוי בדבר פרט למספר התאים הפנויים.

נראה כי הדבר מתקיים גם עבור $k+1$ תאים פנויים:

נסמן את רשימת הפעולות החוקיות כרגע כ q' ונסמן ש $q' = \{q_0, q_1, q_2, \dots, q_k\} \cup \{q_{k+1}\}$

$$V(s) = \max_{q \in q'} \left\{ r(s,q) + \sum_{s' \in S_{t+1}} P(s' \mid s,q)\cdot V(s') \right\} =$$

$$\max\left\{\max_{a\in\{a_0,a_1,a_2,\dots,a_k\}}\left\{r(s,a)+\sum_{s'\in S_{k+1}} P(s'|s,a)\cdot V(s')\right\},\; r(s,a_{k+1})+\sum_{s'\in S_{k+1}} P(s'|s,a_{k+1})\cdot V(s')\right\}$$

(האיבר הראשון:) תלוי רק במספר הגומות הריקות, מהנחת האינדוקציה.

(כל הביטוי:) תלוי רק במספר הגומות הריקות מבסיס האינדוקציה.

המעבר חוקי משום שאם $A\cup B=C$ אז $\max\{C\}=\max\{\max\{A\},\max\{B\}\}$

ובפרט קיבלנו כי עבור רצף גומות ריקות, $V(s)$ ובהתאמה $\pi(s)$ תלוי אך ורק במספר הגומות הריקות, כנדרש.

3. נחשב את המדיניות האופטימלית עבור $N=2$ בעזרת
בעזרת אלגוריתם התכנות הדינאמי הבא –

- Initialize: $V(s) = \max_{a\in A} r(s,a)$ for all $s \in S_T$.
- For $t = T-1, \dots, 1$
  - For all $s \in S_t$, set
    - $V(s) = \max_{a\in A} r(s,a) + \sum_{s' \in S_{t+1}} P(s' \mid s,a)V(s')$.
    - $\pi(s) = \arg\max_{a\in A} r(s,a) + \sum_{s' \in S_{t+1}} P(s' \mid s,a)V(s')$.
- Return π. — Deterministic Markov policy

עבור $t=3$

$$V_3((d_3,1,1,1),a) = 0 \quad \forall\, a \in (0,\dots,9)$$ ← מחרוזת מלאה ולכן לא ניתן לבצע שום פעולה ע״פ [illegible] לה רווח.

עבור $t=2$, יש אחד פנוי.

$$V_2((d_3,0,1,1),2) = \frac{d_3}{9}10^{2-2} = \frac{d_3}{9} \Rightarrow \pi_2^*(d_3,0,1,1) = 2$$

$$V_2((d_3,1,0,1),1) = \frac{d_3}{9}\cdot 10^{-1} \Rightarrow \pi_2^*(d_3,1,0,1) = 1$$

$$V_3((d_3,1,1,0),0) = \frac{d_3}{9}\cdot 10^{-2} \Rightarrow \pi_2^*(d_3,1,1,0) = 0$$

עבור $t=1$, שני תאים פנויים.

$$\pi_1^*(d_3,0,0,1) = \arg\max_{a\in[0,1]} \Big[r((d_3,0,0,1),2) + \sum_{s'\in S_{t+1}} P((k,1,0,1)\mid(d_3,0,0,1),2)\cdot V_2(k,1,0,1),$$

$$r((d_3,0,0,1),1) + \sum_{s'\in S_{t+1}} P((k,0,1,1)\mid(d_3,0,0,1),1)\cdot V_2(k,0,1,1) \Big] =$$

$$= \arg\max_{a\in[2,1]} \left[\frac{d_3}{9} + \sum_{k=0}^{9} \frac{1}{10}\cdot\frac{k}{9}\cdot 10^{-1},\ \frac{d_3}{9}\cdot 10^{-1} + \sum_{k=0}^{9}\frac{1}{10}\cdot\frac{k}{9} \right] =$$

$$= \underset{a\in[2,1]}{\operatorname{argmax}} \left[\frac{d_3}{9} + 0.05 \,,\, \frac{d_3}{90} + 0.5 \right] \Rightarrow \frac{d_3}{9} + 0.05 \underset{1}{\overset{2}{\gtrless}} \frac{d_3}{90} + 0.5 \Rightarrow$$

$$\Rightarrow \frac{d_3}{10} \underset{1}{\overset{2}{\gtrless}} 0.45 \Rightarrow d_3 \underset{1}{\overset{2}{\gtrless}} 4.5 \Rightarrow$$

אם המספר שהוגרל גדול
מ-4.5 נציב אותו בסיבית 2
אחרת נציב אותו בסיבית 1.

$$\pi_1^*(d_3,1,0,0) = \underset{a\in[1,0]}{\operatorname{argmax}} \Big[r((d_3,1,0,0),1) + \sum p((k,1,1,0)\,|\,(d_3,1,0,0),1)\, V_2(k,1,1,0) \,,\, r((d_3,1,0,0),0) +$$

$$+ \sum p((k,1,0,1)\,|\,(d_3,1,0,0),0)\, V_2(k,1,0,1) \Big] =$$

$$= \underset{a\in[1,0]}{\operatorname{argmax}} \left[\frac{d_3}{9} \cdot 10^{-1} + \sum_{k=0}^{9} \frac{1}{10} \cdot \frac{k}{9} \cdot 10^{-2} \,,\, \frac{d_3}{9} \cdot 10^{-2} + \sum_{k=0}^{9} \cdot \frac{1}{10} \cdot \frac{k}{9} \cdot 10^{-1} \right] =$$

$$= \underset{a\in[1,0]}{\operatorname{argmax}} \left[\frac{d_3}{90} + \frac{1}{200} \,,\, \frac{d_3}{900} + \frac{1}{20} \right] \Rightarrow \frac{d_3}{90} + \frac{1}{200} \underset{0}{\overset{1}{\gtrless}} \frac{d_3}{900} + \frac{1}{20} \Rightarrow$$

$$\Rightarrow \frac{d_3}{100} \underset{0}{\overset{1}{\gtrless}} \frac{9}{200} \Rightarrow d_3 \underset{0}{\overset{1}{\gtrless}} 4.5 \Rightarrow$$

אם המספר שהוגרל גדול מ-4.5
נציב אותו בסיבית 1 אחרת
נציב אותו בסיבית 0.

$$\pi_1^*(d_3,0,1,0) = \underset{a\in[2,0]}{\operatorname{argmax}} \Big[r((d_3,0,1,0),2) + \sum p((k,1,1,0)\,|\,(d_3,0,1,0),2)\, V_2(k,1,1,0) \,,\, r((d_3,0,1,0),0) +$$

$$+ \sum p((k,0,1,1)\,|\,(d_3,0,1,0),0)\, V_2(k,0,1,1) \Big] =$$

$$= \underset{a\in[2,0]}{\operatorname{argmax}} \left[\frac{d_3}{9} + \sum_{k=0}^{9} \frac{1}{10} \cdot \frac{k}{9} \cdot 10^{-2} \,,\, \frac{d_3}{9} \cdot 10^{-2} + \sum_{k=0}^{9} \frac{1}{10} \cdot \frac{k}{9} \right] = \underset{a\in[2,0]}{\operatorname{argmax}} \left[\frac{d_3}{9} + \frac{1}{200} \,,\, \frac{d_3}{900} + \frac{1}{2} \right] \Rightarrow$$

$$\Rightarrow \frac{d_3}{9} + \frac{1}{200} \underset{0}{\overset{2}{\gtrless}} \frac{d_3}{900} + \frac{1}{2} \Rightarrow \frac{11}{100} \cdot d_3 \underset{1}{\overset{2}{\gtrless}} \frac{99}{200} \Rightarrow d_3 \underset{1}{\overset{2}{\gtrless}} 4.5$$

אם המספר שהוגרל גדול מ-4.5
נציב אותו בסיבית 2 אחרת, נציב אותו בסיבית 0.

$$\pi_1^*(d_3,0,0,1)=\begin{cases}1, & 0\le d_3\le 4\\ 2, & 5\le d_3\le 9\end{cases}\quad V_1(d_3,0,0,1)=\begin{cases}\frac{d_3}{90}+\frac{1}{2}, & 0\le d_3\le 4\\ \frac{d_3}{9}+\frac{1}{200}, & 5\le d_3\le 9\end{cases}$$

$$\pi_1^*(d_3,1,0,0)=\begin{cases}0, & 0\le d_3\le 4\\ 1, & 6\le d_3\le 9\end{cases}\quad V_1(d_3,1,0,0)=\begin{cases}\frac{d_3}{900}+\frac{1}{20}, & 0\le d_3\le 4\\ \frac{d_3}{90}+\frac{1}{200}, & 5\le d_3\le 9\end{cases}$$

$$\pi_1^*(d_3,0,1,0)=\begin{cases}0, & 0\le d_3\le 4\\ 2, & 6\le d_3\le 9\end{cases}\quad V_1(d_3,0,1,0)=\begin{cases}\frac{d_3}{900}+\frac{1}{2}, & 0\le d_3\le 4\\ \frac{d_3}{9}+\frac{1}{200}, & 5\le d_3\le 9\end{cases}$$

$$\pi_0^*(d_3,0,0,0)=\underset{a\in(2,1,0)}{\operatorname{argmax}}\Big[r\big((d_3,0,0,0),2\big)+\sum p\big((k,1,0,0)\,\big|\,(d_3,0,0,0),2\big)V_1(k,1,0,0),$$

$$r\big((d_3,0,0,0),1\big)+\sum p\big((k,0,1,0)\,\big|\,(d_3,0,0,0),1\big)V_1(k,0,1,0),$$

$$r\big((d_3,0,0,0),0\big)+\sum p\big((k,0,0,1)\,\big|\,(d_3,0,0,0),0\big)V_1(k,0,0,1)\Big]=$$

$$=\underset{a\in(2,1,0)}{\operatorname{argmax}}\Big[\frac{d_3}{9}+\frac{1}{10}\Big(\underbrace{\sum_{k=0}^{4}\Big(\frac{k}{900}+\frac{1}{20}\Big)}_{=\frac{47}{180}}+\underbrace{\sum_{k=5}^{9}\Big(\frac{k}{90}+\frac{1}{200}\Big)}_{=\frac{149}{360}}\Big),$$

$$\frac{d_3}{90}+\frac{1}{10}\Big(\underbrace{\sum_{k=0}^{4}\Big(\frac{k}{900}+\frac{1}{2}\Big)}_{=\frac{113}{45}}+\underbrace{\sum_{k=5}^{9}\Big(\frac{k}{9}+\frac{1}{200}\Big)}_{=\frac{1409}{360}}\Big),$$

$$\frac{d_3}{900}+\frac{1}{10}\cdot\Big(\underbrace{\sum_{k=0}^{4}\Big(\frac{d_3}{90}+\frac{1}{2}\Big)}_{=\frac{47}{18}}+\underbrace{\sum_{k=5}^{9}\Big(\frac{k}{9}+\frac{1}{20}\Big)}_{=\frac{149}{36}}\Big)\Big]=$$

$$= \underset{a\in[2,1,0]}{\arg\max} \left[\frac{d_3}{9} + \frac{27}{400},\ \frac{d_3}{90} + \frac{257}{400},\ \frac{d_3}{900} + \frac{27}{40} \right]$$

$$\frac{d_3}{9} + \frac{27}{400} \underset{1}{\overset{2}{\lessgtr}} \frac{d_3}{90} + \frac{257}{400}$$

$$d_3 \cdot \frac{1}{10} \underset{1}{\overset{2}{\lessgtr}} \frac{23}{40}$$

$$d_3 \underset{1}{\overset{2}{\lessgtr}} 5.75$$

$$\frac{d_3}{90} + \frac{257}{400} \underset{0}{\overset{1}{\gtrless}} \frac{d_3}{900} + \frac{27}{40}$$

$$d_3 \cdot \frac{1}{100} \underset{0}{\overset{1}{\gtrless}} \frac{13}{400}$$

$$d_3 \underset{0}{\overset{1}{\gtrless}} 3.25$$

$$\frac{d_3}{9} + \frac{27}{400} \underset{0}{\overset{2}{\gtrless}} \frac{d_3}{900} + \frac{27}{40}$$

$$d_3 \cdot \frac{11}{100} \underset{0}{\overset{2}{\gtrless}} \frac{243}{400}$$

$$d_3 \underset{0}{\overset{2}{\gtrless}} 5.52$$

— סיום

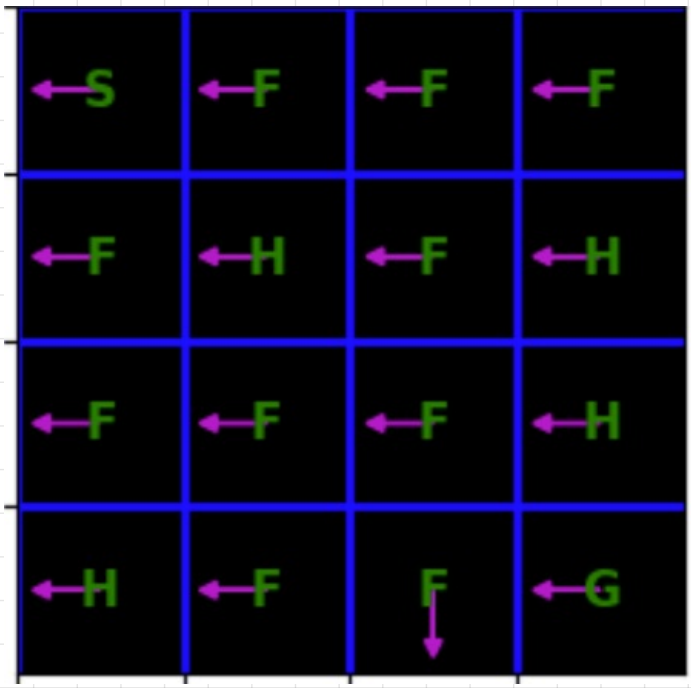
$$\pi_0^*(d_3, 0, 0, 0) = \begin{cases} 0, & 0 \le d_3 \le 3 \\ 1, & 4 \le d_3 \le 5 \\ 2, & 6 \le d_3 \le 9 \end{cases}$$

תרגיל מעשי –

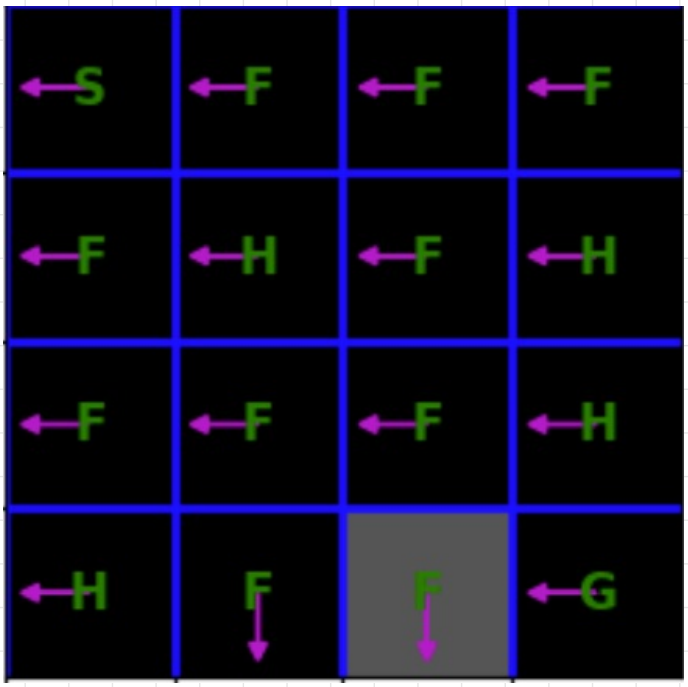
Quetion 1 – Value Iteration

מימוש קוד בסיסי אשר חישב את המדיניות האופטימלית בסביבת "Frozen Lake" עפ"י אלגוריתם של ה- Value Iteration Bellman, בקובץ מצורף.

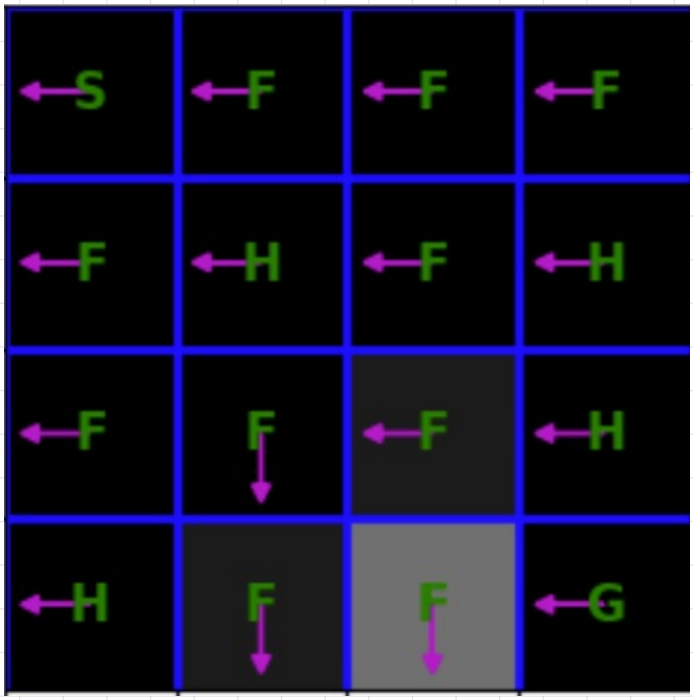
מדיניות אלגוריתמית עפ"י אלגוריתם מסוים דומות שנוצרו.



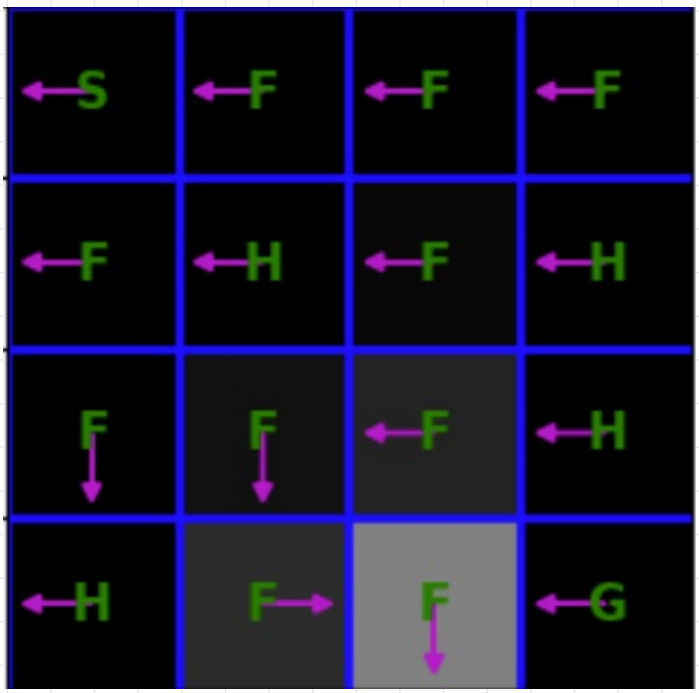
Iteration 1



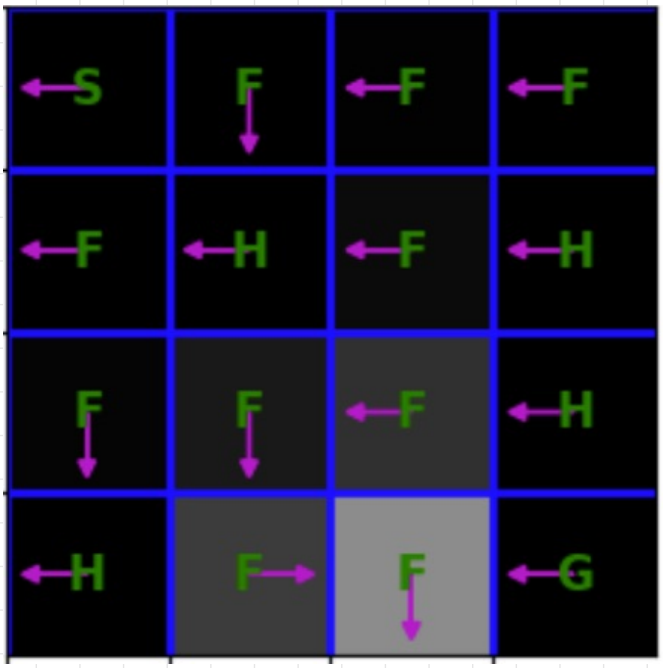
Iteration 2



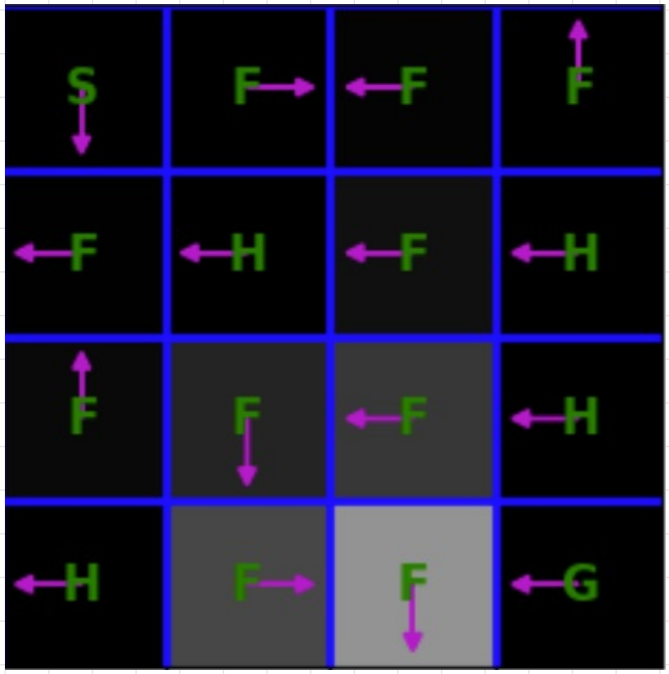
Iteration 3



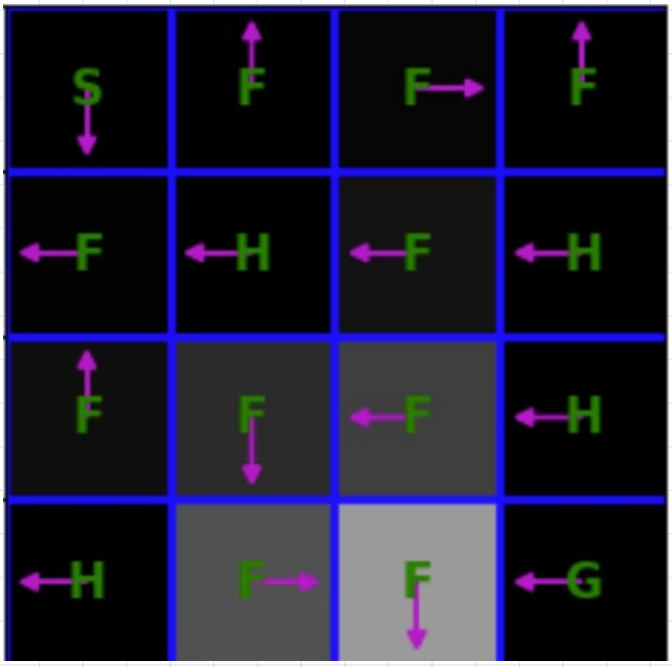
Iteration 4



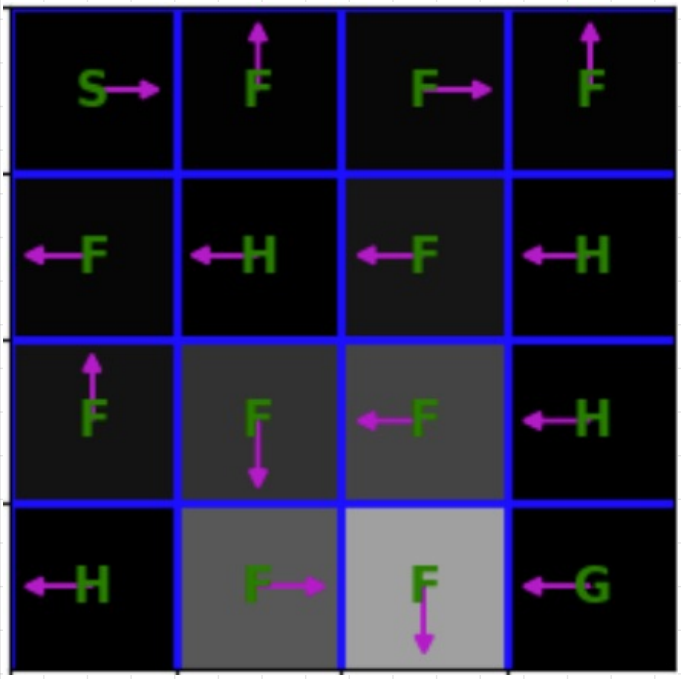
Iteration 5



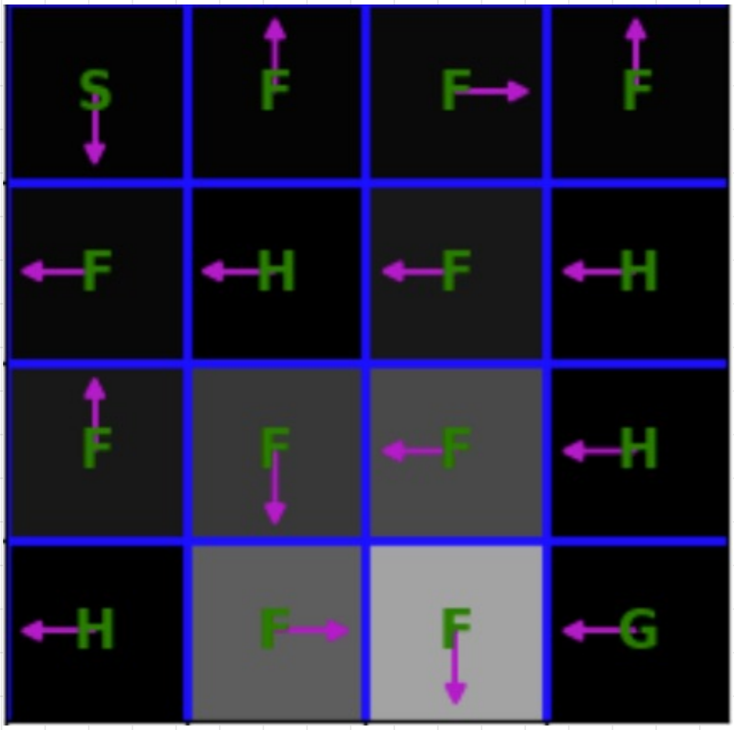
Iteration 6



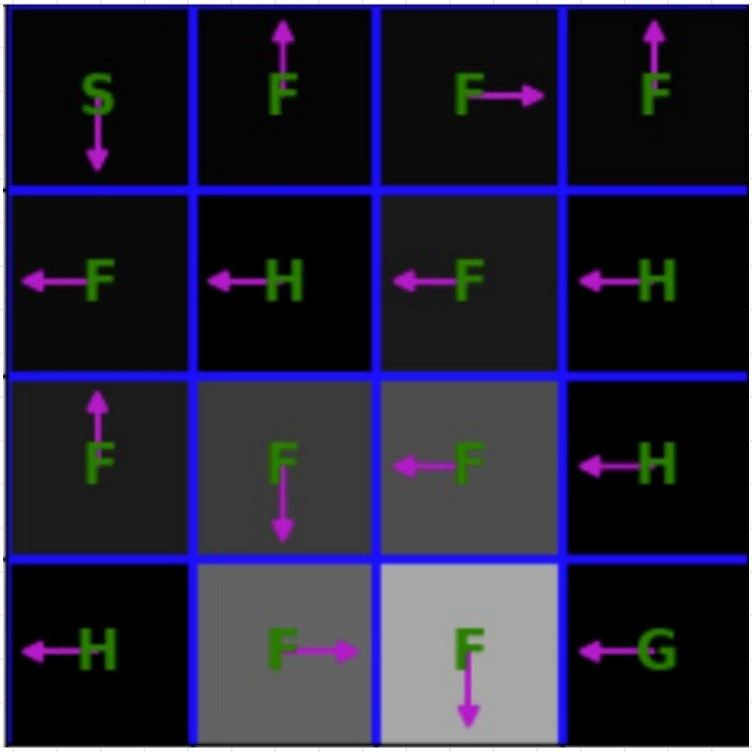
Iteration 7



Iteration 7



Iteration 12



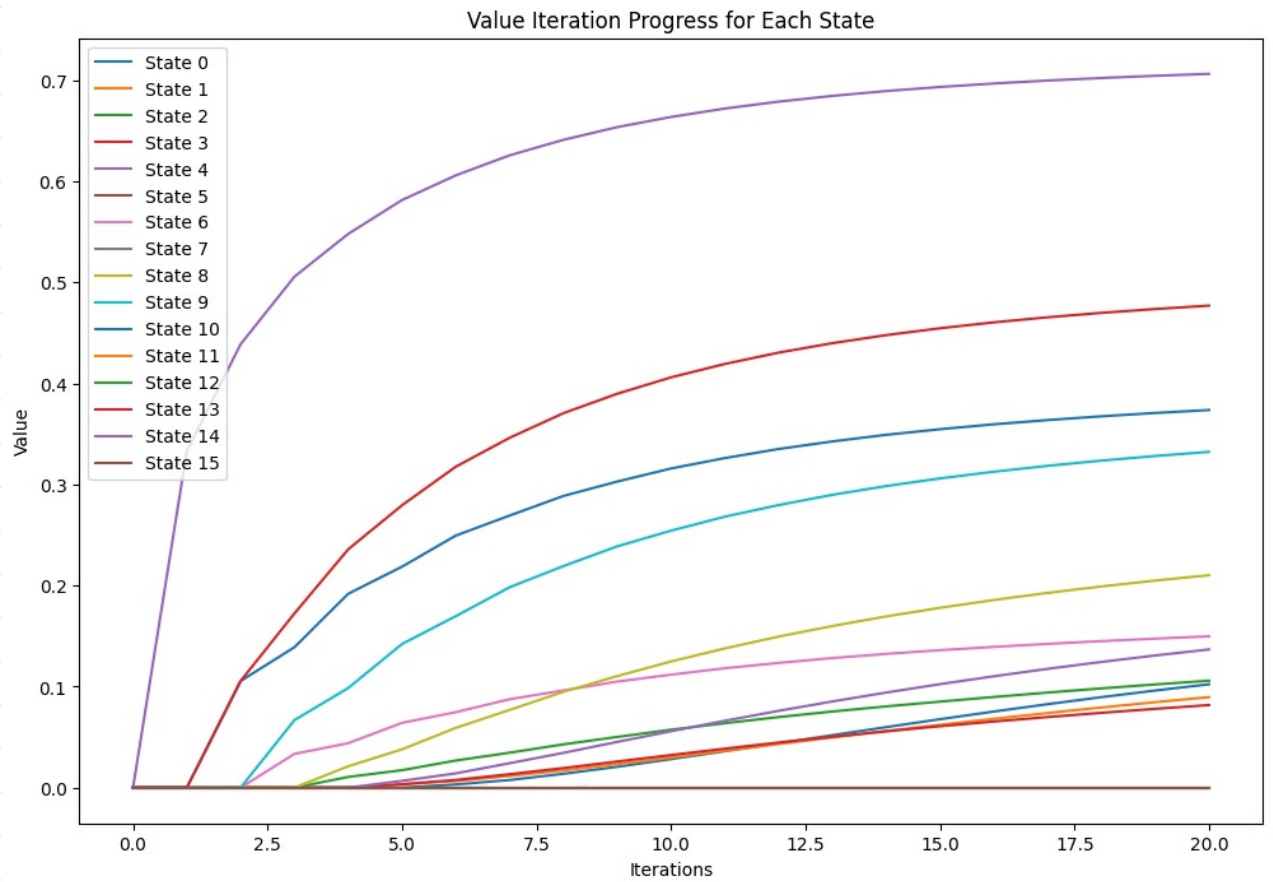
Iteration 14

המדיניות האופטימלית לאחר
התכנסות.

Iteration Table

Iteration	max|V-Vprev|	# chg actions	V[0]
0	0.33333	N/A	0.000
1	0.10556	1	0.000
2	0.06685	1	0.000
3	0.06351	2	0.000
4	0.04357	1	0.000
5	0.03821	4	0.003
6	0.02857	2	0.008
7	0.02437	1	0.014
8	0.01952	1	0.021
9	0.01624	0	0.028
10	0.01384	0	0.036
11	0.01173	0	0.044
12	0.01047	1	0.052
13	0.00948	0	0.060
14	0.00852	1	0.068
15	0.00782	0	0.075
16	0.00733	0	0.083
17	0.00694	0	0.090
18	0.00656	0	0.096
19	0.00618	0	0.102

לכלן הרצת המ"צ את
ה- Value Bellman לכל מצב
בעזרת אופרטורים של הסרק —

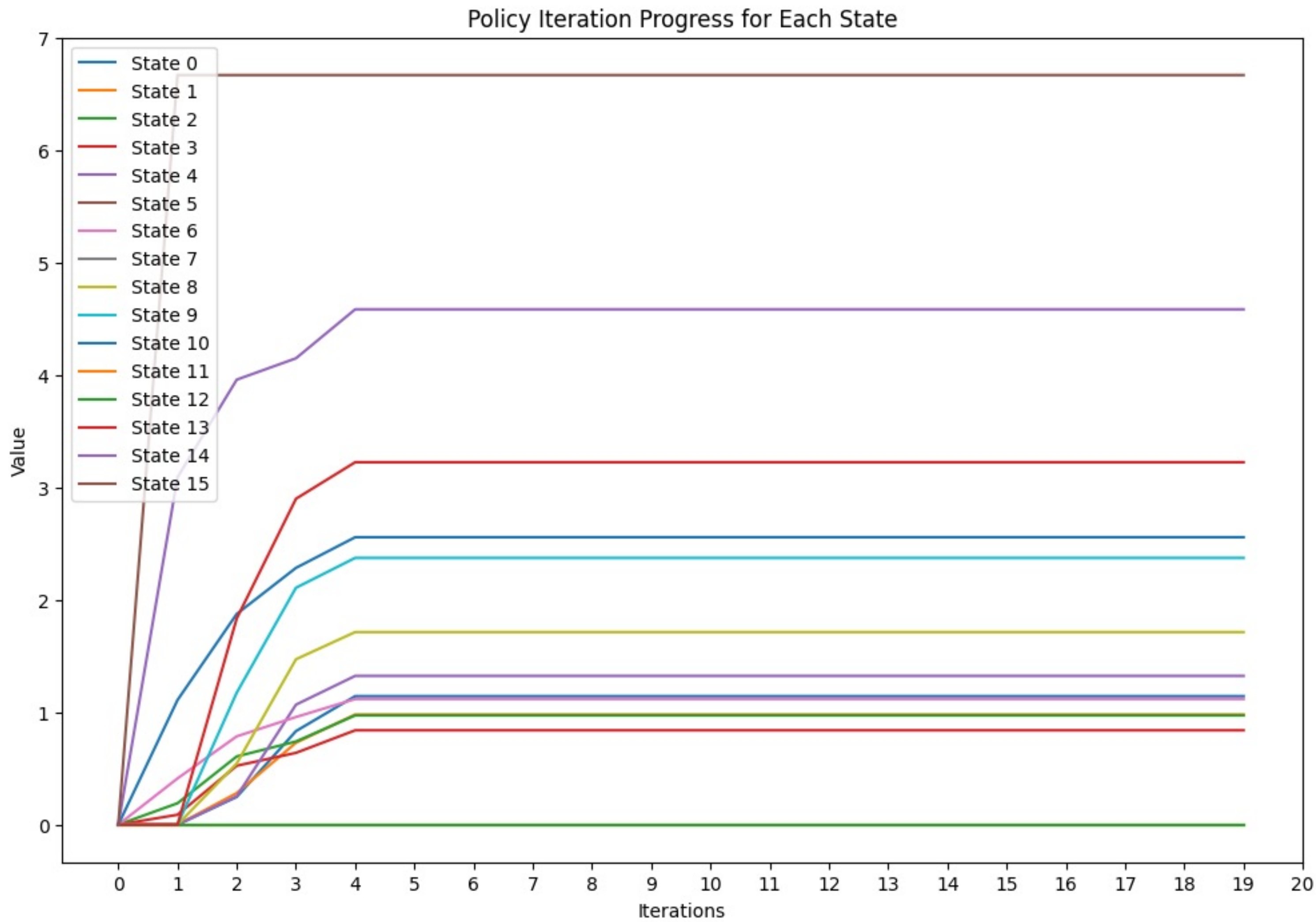


Question 2 – Policy Iteration

Iteration	# chg actions	V[0]
0	1	0.00000
1	9	0.00000
2	5	0.24722
3	3	0.83222
4	0	1.14299
5	0	1.14299
6	0	1.14299
7	0	1.14299
8	0	1.14299
9	0	1.14299
10	0	1.14299
11	0	1.14299
12	0	1.14299
13	0	1.14299
14	0	1.14299
15	0	1.14299
16	0	1.14299
17	0	1.14299
18	0	1.14299
19	0	1.14299

בשאלה זו כתבנו קוד בפייתון אשר חישב את המדיניות האופטימלית עבור "Frozen Lake" במספר איטרציות עד המדיניות התקבעה מפעל. להלן טבלת האיטרציות.

לסיכום מצורף את המצג ה-Bellman Value של כל מצב
בעזרת ויזואליזציה של המדיניות -



נשים לב כי האלגוריתם של ה-Policy Iteration
מתכנס מהר יותר מה-Value Iteration.