חומר עזר בנושא רדוקציות והוכחת נכונות

הגדרת היחס "ניתן לרדוקציה":

תהיינה A ו-B זוג בעיות נתונות. נאמר כי בעיה A ניתנת לרדוקציה לבעיה B אם קיימות זוג פונקציות f,g, כך ש:

- f היא פונקציה המעבירה מופע של בעיה A למופע של בעיה B.
- g היא פונקציה המעבירה פתרון של בעיה B לפתרון של בעיה A.
- עבור מופע a לבעיה A, אם $B(f(a))$ הוא פתרון עבור המופע $f(a)$ תחת בעיה B אזי $g(B(f(a)))$ הוא פתרון למופע a תחת בעיה A.

כדי להוכיח את נכונות הרדוקציה, יש להוכיח שהאלגוריתם הבא פותר את הבעיה A:

1. עבור מופע a לבעיה A, חשב את $f(a)$.
2. עבור המופע $f(a)$ לבעיה B, חשב את הפיתרון b.
3. חשב את $g(b)$ להיות הפתרון של A.

סימון: אם בעיה A ניתנת לרדוקציה לבעיה B נסמן $A \leq B$.

הערות לגבי ההגדרה לעיל:

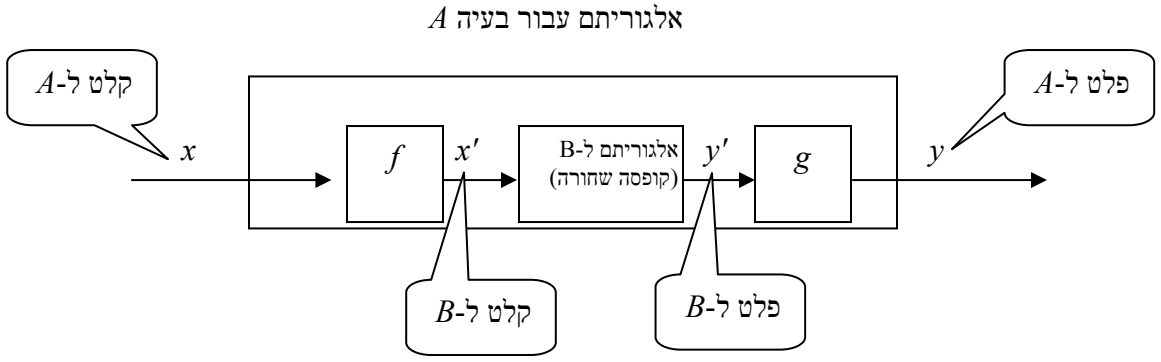
א. הפונקציה f נקראת **פונקצית המרת הקלט**.

ב. הפונקציה g נקראת **פונקצית המרת הפלט**.

ג. נתייחס לאלגוריתם שפותר את בעיה B כאל "**קופסה שחורה**", מבלי להניח דבר על אופן פעולתו מלבד העובדה שהוא אכן פותר נכון את B לכל מופע.

ד. היחס "ניתן-לרדוקציה" הוא טרנזיטיבי: כלומר אם A ניתנת לרדוקציה ל-B, ו-B ניתנת לרדוקציה ל-C, אזי A ניתנת לרדוקציה ל-C (חישבו: מדוע?).

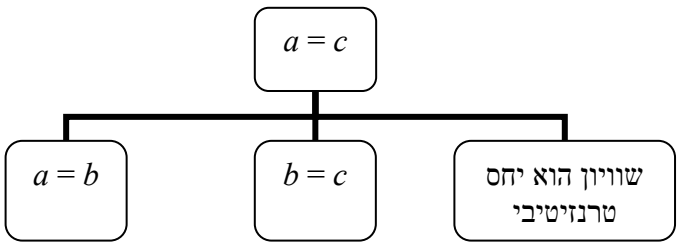
הסכימה הבאה מסכמת את כל הנאמר לעיל:



מבנה הוכחת נכונות

אחד הקשיים העיקריים בכתיבת הוכחות הוא בניסוח מדויק של הטענה העיקרית ושל טענות העזר. קושי נוסף הוא בהחלטה אילו חלקים חשובים יותר ואילו פחות, היכן לפרט והיכן לא. הוכחה "טובה" כתובה "מלמעלה למטה" (Top-Down), כאשר בתחילתה מוסברת נכונות הטענה העיקרית תוך התבססות על טענות עזר, ובהמשך מוסברת נכונות טענות העזר במידת הצורך. במובן זה, ניתן לחשוב על הוכחה כעל "עץ", בו השורש מייצג את הטענה העיקרית, ושאר הקודקודים מייצגים את טענות העזר בהן משתמשים. קבוצת ה"בנים" של קודקוד מסוים מייצגת את קבוצת הטענות הדרושות לצורך הוכחת הטענה המובעת בקודקוד. הוכחה "טובה" היא כזו בה אנו בוחרים היכן "לגזום" את העץ, כלומר בוחרים לאלו טענות להראות הוכחות ולאלו לא, באופן כזה שאנו אומרים את הדברים הכי רלוונטיים הדרושים לצורך שכנוע הקורא בנכונות הטענה מחד, ולא מכבידים בפרטים מאידך. כל הטענות שאנו טוענים **חייבות להיות נכונות,** גם אם בחרנו שלא לפרט ההוכחות עבור חלקן. עבור הטענות אותן בחרנו שלא להוכיח, נשתדל לתת הפניה למקום אחר בו הן אולי מוכחות, או לנסח אותן כ"אבחנות" ולספק הסבר אינטואיטיבי קצר לנכונותן. עבור הטענות שהחלטנו להוכיח אנו חייבים לפרט את כל הדרוש להוכחתן, והשמטה של טענות הכרחיות גורמת להוכחה להיות **לא מספקת**.

הוכחה בה אנו מראים לפרטי פרטים את רוב הענפים בעץ, אך מתעלמים מאחרים, גרועה יותר מהוכחה אשר מציינת את כל הענפים אך לא מוכיחה אף אחד מהם באופן מלא. לדוגמה, הניחו כי אנו רוצים להוכיח כי $a = c$ לפי הסכמה הבאה:



אם נוכיח עד לפרטים הכי קטנים כי $a = b$ וכי שוויון הוא יחס טרנזיטיבי, ההוכחה שלנו עדיין לא מראה מדוע $a = c$. לעומת זאת, אם נראה באופן כללי את כל שלוש טענות העזר אזי סיפקנו הסבר (משכנע יותר או פחות) לשוויון $a = c$.

אלגוריתם והוכחה לדוגמה

הקשיים העיקריים עמם נתמודד במהלך הקורס יהיו היכולת לתכנן אלגוריתם יעיל הפותר בעיה נתונה, וכן היכולת להוכיח את נכונות האלגוריתם. מטרת הטקסט הבא אינה רק להראות דוגמה נוספת לרדוקציה, אלא גם להדגים את התהליך החשיבתי שאנו עוברים בבואנו להגדיר אלגוריתם ולנסח הוכחה עבורו. ישנו פער גדול בין היכולת לקרוא תיאור של אלגוריתם והוכחה עבורו ולהבין אותם, לבין היכולת לכתוב אותם. הדרך הבטוחה ביותר לשפר את יכולת החשיבה והכתיבה שלנו היא ע"י התנסות רבה ככל האפשר הן בקריאת אלגוריתמים והוכחות (ולצורך כך יש את ההרצאות והתרגולים) והן בכתיבתם (ולצורך כך יש את עבודות הבית). ננסה בכל זאת להראות כאן מספר טכניקות ו"כללי אצבע" שיכולים לעזור במקרים רבים.

דרך החשיבה שלנו על אלגוריתמים וההוכחות יכולה להיות שונה מהאופן בו הם יהיו כתובים לבסוף, והיא בעלת אופי "מעגלי" יותר. למשל, יתכן כי נתכנן אלגוריתם אשר אנו "מרגישים" כי הוא פותר את הבעיה, וננסח טענה המוכיחה את נכונותו. יתכן גם כי נצליח להוכיח את הטענה העיקרית בעזרת קבוצת טענות עזר, אולם בבואנו להוכיח את טענות העזר נגלה שחלקן אינן נכונות עבור מקרי קצה מסוימים. בשלב זה, ננסה לתקן את טענות העזר באופן כזה שהן יהיו נכונות, אך עתה נצטרך לשוב ולתקן את הוכחת הטענה העיקרית

כך שתתבסס על טענות העזר החדשות, ואולי אפילו נאלץ לשנות את ניסוח הטענה העיקרית עצמה. שינוי הטענה העיקרית יכול בתורו לגרום לכך שהיא כבר אינה מוכיחה את האלגוריתם, ויתכן כי נצטרך עתה לתקן את האלגוריתם, וכן הלאה. לעתים דווקא לאלגוריתמים והוכחות הנכונות שנראים הכי פשוטים קדם תהליך חשיבה ארוך שכלל מספר סבבים של "ניסוי וטעייה", עד לקבלת התוצאה ה"אלגנטית".

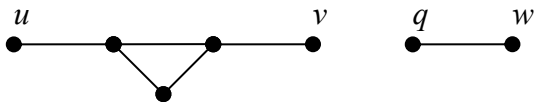
שאלה: כיצד ניגשים לתכנון אלגוריתם עבור בעיה מסוימת? תשובה: טכניקה טובה לקבלת החלטות לגבי מה עלינו לעשות היא פשוט לשאול את עצמנו בכל שלב את **השאלות הנכונות**, ואז לענות על שאלות אלו. מהן השאלות הנכונות? ובכן, זה כבר עניין של אינטואיציה, ואינטואיציה זה עניין של ניסיון. הטקס שיופיע בכחול ייצג מחשבות, והוא אינו חלק מניסוח הרדוקציה או ההוכחה.

<u>הגדרות</u>

בהינתן גרף לא מכוון $G=(V,E)$, נגדיר:

- סדרת קודקודים בגרף $P=(v_0,v_1,\ldots,v_k)$ נקראת **מסלול** אם לכל $0<i\le k$ מתקיים $(v_{i-1},v_i)\in E$. **אורך המסלול** P מוגדר להיות $|P|-1$ (כלומר, אורך סדרת הקשתות המתאימה למסלול).
- בהינתן $u,v\in V$ נסמן ב-$d_G(u,v)$ את אורך מסלול מינימאלי בין u ו-v ב-G אם קיים כזה מסלול, ואחרת נגדיר $d_G(u,v)=\infty$. נאמר כי $d_G(u,v)$ הוא **המרחק** בין u ו-v ב-G. עבור $s\in V$, נגדיר $d_G(s)=\{(v,d_G(s,v)):v\in V\}$ (כלומר $d_G(s)$ הוא אוסף של זוגות המתאר לכל קודקוד בגרף את מרחקו מ-s).
- בהינתן $u,v\in V$ נסמן ב-$ed_G(u,v)$ את אורך מסלול מינימאלי בו יש מספר זוגי של קשתות בין u ו-v ב-G אם קיים כזה מסלול, ואחרת נגדיר $ed_G(u,v)=\infty$. נאמר כי $ed_G(u,v)$ הוא **המרחק הזוגי** בין u ו-v ב-G. עבור $s\in V$, נגדיר $ed_G(s)=\{(v,ed_G(s,v)):v\in V\}$ (כלומר $ed_G(s)$ הוא אוסף של זוגות המתאר לכל קודקוד בגרף את מרחקו הזוגי מ-s).

לדוגמה, בגרף G להלן, $d_G(u,v)=3$, $ed_G(u,v)=6$, $d_G(q,w)=1$, ו-$ed_G(q,w)=\infty$.



<u>בעיית SP (Single Source Shortest Paths):</u>

בהינתן גרף לא מכוון $G=(V,E)$ וקודקוד $s\in V$, יש לחשב את $d_G(s)$.

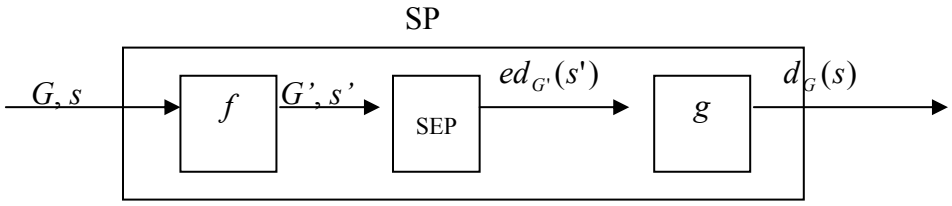
<u>בעיית SEP (Single Source Shortest Even Paths):</u>

בהינתן גרף לא מכוון $G=(V,E)$ וקודקוד $s\in V$, יש לחשב את $ed_G(s)$.

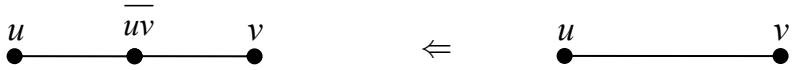
<u>משימה:</u> הראו כי SP ≤ SEP.

שאלה: מה בדיוק עלינו לעשות? תשובה: עלינו להראות שאם קיים אלגוריתם אשר יודע לחשב את המרחק <u>הזוגי</u> מקודקוד מקור כלשהו לכל שאר הקודקודים בגרף, אזי ניתן לכתוב אלגוריתם אשר מחשב את המרחק מקודקוד מקור כלשהו לכל שאר הקודקודים בגרף.

אנו יכולים לקבוע מייד את המבנה הצורני של הרדוקציה, לפני שהשקענו שניית מחשבה בפתרון, רק על סמך הגדרת הבעיות. מבנה זה יסיע לנו לבחור את הפתרון הנכון. במקרה זה, המבנה הוא המבנה להלן:



שאלה: איך עושים זאת? תשובה: ננסה לשנות את גרף הקלט שקבלנו כך שכל מסלול בגרף המקורי G יתאים למסלול בעל אורך כפול בגרף החדש G'. נעשה זאת ע"י כך ש נפצל כל קשת מקורית לשתי קשתות המחוברות ע"י קודקוד "דמה". לדוגמה, גרף המכיל שני קודקודים וקשת ביניהם, יתורגם באופן הבא:



לאחר שנחשב מרחקים זוגיים בין קודקוד המקור לכל הקודקודים בגרף החדש, נוכל לטעון כי המרחק בין s ו-v בגרף המקורי שווה למחצית המרחק הזוגי בין הקודקודים המתאימים בגרף החדש. כמובן, נצטרך לתאר רדוקציה זו באופן מדויק, ולהוכיח את נכונותה.

<u>תיאור הרדוקציה:</u>

<u>פונקצית המרת הקלט f:</u> בהינתן גרף לא מכוון $G=(V,E)$ וקודקוד $s\in V$, נחשב את הפונקציה $f(G,s)=(G',s')$, כאשר $G'=(V',E')$, ו-

- $V'=V^1\cup V^2$, כאשר V^1 הוא העתק של V, ו-$V^2=\{\overline{uv}:(u,v)\in E\}$ (כלומר, הקודקודים ב-V' מתאימים לקודקודים ולקשתות ב-G). עבור קודקוד $v\in V$, נסמן ב-v' את הקודקוד המתאים ל-v ב-V^1. נשתמש בסימון $\overline{uv}$ כדי לייצג את הקודקוד ב-V^2 המתאים לקשת (u,v) ב-E (נשים לב כי כיוון שהגרף לא מכוון, הסימונים $\overline{uv}$ ו-$\overline{vu}$ מתייחסים לאותו הקודקוד).
- $E'=\{(u',\overline{uv}):(u,v)\in E\}$ (כלומר, קשתות ב-E' הן בין קודקודים $u'\in V^1$ לבין הקודקודים $\overline{uv}\in V^2$ המתאימים לקשתות הנוגעות ב-u ב-G).
- s' הוא הקודקוד המתאים ל-s ב-V^1.

<u>פונקצית המרת הפלט g:</u> בהינתן $G'=(V',E')$ ו-$s'\in V'$ כך ש-$(G',s')=f(G,s)$ עבור $G=(V,E)$ ו-$s\in V$, וכן פתרון $ed_{G'}(s')$ לבעיית SEP על המופע G' ו-s', נחשב את קבוצת הזוגות הבאה: $\widetilde{d}_G(s)=g(ed_{G'}(s'))=\left\{(v,\widetilde{d}_G(s,v)):v\in V,\widetilde{d}_G(s,v)=\frac{1}{2}ed_{G'}(s',v')\right\}$.

<u>תיאור אלגוריתם המבוסס על הרדוקציה:</u>

1. בהינתן מופע לבעיית SP, גרף לא מכוון $G=(V,E)$ וקודקוד $s\in V$, נחשב $(G',s')=f(G,s)$.
2. נפעיל אלגוריתם הפותר את SEP על המופע G' ו-s', לקבלת קבוצת הזוגות $ed_{G'}(s')$.
3. נחזיר את קבוצת הזוגות $\widetilde{d}_G(s)=g(ed_{G'}(s'))$ כפתרון לבעיית SP על G ו-s.

הערות:

1. אנו משתמשים כאן בסימון $\widetilde{d}_G(s)$ כדי להבחין בין פלט הרדוקציה לבין התשובה המבוקשת $d_G(s)$, כל עוד לא הוכחנו שהם שווים.
2. הרדוקציה עצמה מתייחסת רק לתיאור פונקציות המרת הקלט והפלט (f ו-g). הפעלת הקופסה השחורה אינה חלק מהרדוקציה, אלא חלק מהאלגוריתם המתבסס על הרדוקציה. למעשה, כדי להראות כי SEP $\leq$ SP צריך לתאר את הרדוקציה בלבד, והאלגוריתם המבוסס על הרדוקציה מובא כאן מטעמי בהירות.

<u>הוכחת נכונות הרדוקציה:</u>

שאלה: מה עלינו להראות כדי להוכיח את נכונות הרדוקציה? תשובה: עלינו להוכיח כי הערכים $\widetilde{d}_G(s)$ שהוחזרו על ידי הרדוקציה הם אכן הפתרון המבוקש $d_G(s)$ לבעיית SP על המופע הנתון.

<u>טענה עיקרית</u>: $d_G(s,v)=\tilde{d}_G(s,v)$ לכל $v\in V$.

שאלה: מה בעצם הטענה אומרת? תשובה: כיוון ש-$\tilde{d}_G(s,v)=\frac{1}{2}ed_{G'}(s',v')$, הטענה אומרת כי לכל $v\in V$, אורך <u>מסלול קצר ביותר</u> בין s ו-v בגרף המקורי <u>שווה</u> למחצית מאורך <u>מסלול זוגי קצר ביותר</u> בין הקודקודים המתאימים בגרף החדש. נשים לב כי הטענה מתייחסת לקשר בין ארכי מסלולים קצרים ביותר. ניסיון להוכיח טענה זו באופן ישיר יכול לגרור הוכחה מסורבלת, שכן נצטרך להתייחס בהוכחה הן לקשר בין ארכי מסלולים והן לתכונות ספציפיות של המסלולים (במקרה זה – מסלולים שהם "קצרים ביותר").

טיפ: במקרים רבים, נח להוכיח טענה ספציפית ע"י הוכחת טענה כללית יותר, הגוררת אותה (באופן ישיר או עקיף).

במקרה זה ננסח טענת עזר המתייחסת לקשר בין ארכי המסלולים ללא התייחסות להיותם קצרים ביותר. ניסיון לניסוח טענה כזו יכול להוביל לטענה הבאה:

<u>טענת עזר</u>: קיים מסלול באורך l בין u ו-v ב-G אם"ם קיים מסלול באורך $2l$ בין u' ו-v' ב-G'.

הטענה לעיל מתאימה לאינטואיציה עליה ביססנו את הרדוקציה, וניתן לכתוב הוכחה עבור הטענה העיקרית אשר מתבססת עליה. הבעיה היחידה היא שטענה זו אינה נכונה, ויתכן כי נגלה זאת רק כאשר ננסה לכתוב לה הוכחה. אם אינכם רואים מיד מדוע הטענה אינה נכונה, נסו לכתוב הוכחה עבור טענה זו, והבינו את הטעות שבה.

<u>הגדרה</u>: **מסלול פשוט** הוא מסלול המבקר בכל קודקוד לכל היותר פעם אחת.

<u>טענת עזר</u>: קיים מסלול פשוט באורך l בין u ו-v ב-G אם"ם קיים מסלול פשוט באורך $2l$ בין u' ו-v' ב-G'.

<u>הוכחת הטענה העיקרית</u>: יהי $v \in V$. נוכיח בנפרד את המקרה בו אין מסלול ואת המקרה בו יש מסלול בין s ו-v ב-G.

א. אם אין מסלול בין s ו-v ב-G: במקרה זה, $d_G(s,v) = \infty$ לפי ההגדרה.

עתה אנו רוצים להראות כי גם $\tilde{d}_G(s,v) = \frac{1}{2} ed_{G'}(s',v') = \infty$. אחת הדרכים להוכיח טענות היא להניח בשלילה כי הן לא מתקיימות, ולהגיע לסתירה. נניח בשלילה כי $ed_{G'}(s',v') \neq \infty$. מכך נובע כי קיים מסלול בין s' ו-v' ב-G'. הינו רוצים בשלב זה להשתמש בטענת העזר כדי להראות שקיים מסלול בין s ו-v ב-G ולכן $d_G(s,v) \neq \infty$ ובכך להגיע לסתירה, אך נשים לב שטענת העזר לא עובדת על מסלול כלשהו בין s ו-v ב-G', אלא רק על מסלולים פשוטים באורך זוגי. כדי שנוכל בכל זאת להשתמש בטענה נצטרך להראות שקיים מסלול כזה. **שאלה: איך מראים שקיים מסלול כזה? תשובה:** חישבו לבד. לאחר מכן, קראו את אבחנות I ו-II המובאות בסוף הוכחה זו ובדקו אם הייתם בסדר. מתוך אבחנות I ו-II עולה כי קיים מסלול פשוט בין s' ו-v' ב-G' בעל מספר זוגי של קשתות, ולכן מטענת העזר נובע כי קיים מסלול בין s ו-v ב-G, כלומר $d_G(s,v) \neq \infty$ - סתירה.

לכן, $\tilde{d}_G(s,v) = d_G(s,v) = \infty$.

ב. אם קיים מסלול בין s ו-v ב-G: **שאלה: מה עושים עכשיו? תשובה:** ננסה להשתמש בטענת העזר בכדי למצוא קשר בין אורכי המסלולים המבוקשים. כיוון שהטענה מתייחסת למסלולים כלשהם ואנו רוצים להוכיח קשר בין אורכי מסלולים אופטימאליים, נשתמש באורכי המסלולים האופטימאליים כחסמים לאורכי המסלולים שאנו יודעים על קיומם (ונקווה לטוב...). במקרה זה קיים מסלול כזה P באורך מינימאלי $d_G(s,v)$. מאבחנה I, P הוא מסלול פשוט, ולכן מטענת העזר קיים ב-G' מסלול P' בין s' ו-v' אשר אורכו $2d_G(s,v)$. אורך המסלול הזוגי המינימאלי בין s' ו-v' ב-G' קטן או שווה לאורכו של P', ולכן $ed_{G'}(s',v') \leq 2d_G(s,v)$ (1). עתה, נתבונן על מסלול Q' בעל אורך זוגי מינימאלי $ed_{G'}(s',v')$ בין s' ו-v' ב-G' (בשלב זה ידוע כבר כי קיים מסלול כזה). מאבחנות I ו-II, Q' הוא מסלול פשוט (חישבו מדוע), ולכן מטענת העזר קיים ב-G מסלול Q בין s ו-v אשר אורכו $\frac{1}{2} ed_{G'}(s',v')$. אורך המסלול המינימאלי בין s ו-v ב-G קטן או שווה לאורכו של Q, ולכן

$d_G(s,v) \leq \frac{1}{2} ed_{G'}(s',v')$ (2). מ-(1) ו-(2) עולה כי $d_G(s,v) = \frac{1}{2} ed_{G'}(s',v') = \tilde{d}_G(s,v)$.

<u>מ.ש.ל</u>

<u>אבחנה I</u>: מסלול קצר ביותר בין קודקודים בגרף הוא בהכרח פשוט.

<u>אבחנה II</u>: לכל $u', v' \in V^1$, כל מסלול בין u' ו-v' ב-G' מכיל מספר זוגי של קשתות.

*הערה, בד"כ רצוי לכתוב את כל האבחנות וטענות העזר לפני שעושים בהן שימוש בתוך הוכחה כלשהי. במקרה זה האבחנות רשומות בסוף כדי לאפשר לכם לחשוב ולהבין את הצורך בהן לפני שאתם רואים אותן מנוסחות.

בכדי להשלים את ההוכחה, יש להוכיח את טענת העזר, ורצוי להסביר את האבחנות. נסו להוכיח את טענת העזר, וכן לספק אינטואיציה קצרה לכל אחת מהאבחנות. כמו כן, נתחו את זמן הריצה של הרדוקציה (זמן ריצת הרדוקציה כולל את זמן ריצת f ו-g בלבד, ללא התייחסות לזמן ריצת הקופסה השחורה).