דיפרנציאביליות - טענות בלבד

אנליזה אלמנטרית רב-ממדית - 80116

מרצה: נועה ניצן

סוכם ע"י: שריה אנסבכר

סמסטר ב' תשפ"ג, האונ' העברית

תוכן העניינים

אשמח לקבל הערות והארות על הסיכומים על מנת לשפרם בעתיד,
כל הערה ולו הפעוטה ביותר (אפילו פסיק שאינו במקום או רווח מיותר) תתקבל בברכה;
אתם מוזמנים לכתוב לי לתיבת הדוא"ל: sraya.ansbacher@mail.huji.ac.il.

לסיכומים נוספים היכנסו לאתר:
אקסיומת השלמות - סיכומי הרצאות במתמטיקה
https://srayaa.wixsite.com/math

1 התחלה

גם בפרק זה אנחנו נעסוק רק בפונקציות מהצורה $f:D\to\mathbb{R}$ כאשר $D\subseteq\mathbb{R}^{n}$ ולכן לא אטרח לציין זאת בכל פעם.

טענה 1.1. תהא f פונקציה דיפרנציאבילית בנקודה $P\in\mathbb{R}^{n}$, f גם רציפה ב-P.

משפט 1.2. תהא f פונקציה מוגדרת בסביבה מלאה של נקודה $P\in\mathbb{R}^{n}$, אם כל הנגזרות החלקיות של f רציפות ב-P[1] אז f דיפרנציאבילית ב-P.

♣ הכיוון ההפוך אינו נכון, נביא דוגמה נגדית.
תהא $f:\mathbb{R}^{2}\to\mathbb{R}$ המוגדרת ע"י (לכל $(x,y)\in\mathbb{R}^{2}$):

$$f\begin{pmatrix}x\\y\end{pmatrix}:=\begin{cases}\left(x^{2}+y^{2}\right)\cdot\sin\left(\frac{1}{\sqrt{x^{2}+y^{2}}}\right) & (x,y)\neq(0,0)\\0 & (x,y)=(0,0)\end{cases}$$

הגרף של הפונקציה הזו הוא סיבוב של גרף הפונקציה 'סינוס "משוגעת" - הכפלה ב-x^{2}' (ראו בקובץ "מאגר פונקציות פתולוגיות") ולכן נדע שהנגזרות החלקיות שלה אינן רציפות בראשית הצירים אבל למרות זאת היא דיפרנציאבילית.

משפט 1.3. תהא f פונקציה דיפרנציאבילית בנקודה $P\in\mathbb{R}^{n}$ ויהי $\vec{u}$ וקטור יחידה, מתקיים $D_{\vec{u}}f\left(P\right)=\left(D_{f}\left(P\right)\right)\left(\vec{u}\right)$, כלומר הדיפרנציאל השלם של $\vec{u}$ הוא הנגזרת הכיוונית בכיוון $\vec{u}$.

מסקנה 1.4. לכל וקטור יחידה $\vec{u}\in\mathbb{R}^{n}$ מתקיים (תהא $\theta\in\mathbb{R}$ זווית שבין $\vec{u}$ ל-$\vec{\nabla}f\left(P\right)$):

$$\begin{aligned}D_{\vec{u}}f\left(P\right)&=\vec{\nabla}f\left(P\right)\cdot\vec{u}=\left\Vert \vec{\nabla}f\left(P\right)\right\Vert \cdot\left\Vert \vec{u}\right\Vert \cdot\cos\theta\\&=\left\Vert \vec{\nabla}f\left(P\right)\right\Vert \cdot1\cdot\cos\theta=\left\Vert \vec{\nabla}f\left(P\right)\right\Vert \cdot\cos\alpha\end{aligned}$$

♣ נניח בהג"כ ש-$\theta\in\left[0,\pi\right]$, המסקנה הזו מספרת לנו שהכיוון שבו העליה מהנקודה היא התלולה ביותר הוא הכיוון עליו מצביע הגרדיאנט (כאשר $\theta=0$), הכיוון שבו הירידה היא המהירה ביותר הוא בכיוון הנגדי (כאשר $\theta=\pi$) והכיוונים הניצבים להם (כאשר $\theta=\frac{\pi}{2}$) הם אלו שבהם לא נעלה ולא נרד. ובכלל: בכיוונים שעבורם $\theta<\frac{\pi}{2}$ הנגזרת הכיוונית תהיה חיובית (עליה) כאשר השיא הוא עבור $\theta=0$, ובכיוונים שעבורם $\theta>\frac{\pi}{2}$ הנגזרת הכיוונית תהיה שלילית (ירידה) כאשר השיא הוא עבור $\theta=\pi$.

[1] ובפרט מוגדרות בסביבה מלאה שלה.

2 דיפרנציאביליות ונקודות קיצון

גם בפרק זה אנחנו נעסוק רק בפונקציות מהצורה $f:D\to\mathbb{R}$ כאשר $D\subseteq\mathbb{R}^{n}$ ולכן לא אטרח לציין זאת בכל פעם.

טענה 2.1. תהא f פונקציה כך ש-$P\in\mathbb{R}^{n}$ היא נקודת קיצון מקומית של f, ערכה של כל נגזרת כיוונית של f ב-P (אם קיימת כזו) הוא 0.

משפט 2.2. משפט פרמה

תהא f פונקציה כך ש-$P\in\mathbb{R}^{n}$ היא נקודת קיצון מקומית של f, אם f דיפרנציאבילית ב-P אז הדיפרנציאל השלם של f ב-P הוא העתקת האפס ($D_{f}\left(P\right)\equiv0$).

משפט 2.3. משפט שוורץ[2] [3]

תהא f פונקציה ותהא $P\in\mathbb{R}^{n}$ נקודה בתחום ההגדרה, לכל $i,j\in\mathbb{N}\geq n$, אם $D_{ij}f\left(P\right)$ ו-$D_{ji}f\left(P\right)$ רציפות ב-P[4] אז הן שוות.

- ♣ משפט שוורץ נכון לכל שתי נגזרות כיווניות ולאו דווקא כאלה המקבילות לצירים.
- ♣ דוגמה לכך שהמשפט אינו נכון אם הנגזרות המעורבות אינן רציפות בנקודה היא הפונקציה $f:\mathbb{R}^{2}\to\mathbb{R}$ המוגדרת ע"י:

$$f\begin{pmatrix}x\\y\end{pmatrix}:=\begin{cases}\frac{x^{3}y-xy^{3}}{x^{2}+y^{2}} & \left(x,y\right)\neq\left(0,0\right)\\0 & \left(x,y\right)=\left(0,0\right)\end{cases}$$

אם נחשב את הנגזרות המעורבות בראשית הצירים נקבל $D_{12}f\left(0,0\right)=1\neq-1=D_{21}f\left(0,0\right)$.

2.1 מיון נקודות קריטיות

משפט 2.4. תהא $U\subseteq\mathbb{R}^{2}$ ותהא $f:U\to\mathbb{R}$ פונקציה המקיימת שכל הנגזרות החלקיות שלה מסדר שני בנקודה $P_{0}\in U$ רציפות ב-P_{0}, $T_{2,f,P_{0}}\left(P\right)$ הוא הפולינום היחיד מדרגה קטנה או שווה ל-2 המקיים:

$$\lim_{P\to P_{0}}\frac{f\left(P\right)-T_{2,f,P_{0}}\left(P\right)}{\left\Vert P-P_{0}\right\Vert ^{2}}=0$$

- ♣ אם נסמן ב-$\phi\left(P\right)$ את הביטוי שבתוך הגבול ונגדיר $\phi\left(P_{0}\right)=0$ נקבל שמתקיים $f\left(P\right)=T_{2,f,P_{0}}\left(P\right)+\phi\left(P\right)\cdot\left\Vert P-P_{0}\right\Vert ^{2}$ לכל $P\in U$, כעת נזכור שאם P_{0} היא נקודה קריטית אז מתקיים[5] (לכל $\left(x,y\right)\in U$):

$$\begin{aligned}T_{2,f,P_{0}}\begin{pmatrix}x\\y\end{pmatrix} & =f\left(P_{0}\right)+\frac{1}{2}\cdot\left(A\left(x-x_{0}\right)^{2}+2B\left(x-x_{0}\right)\left(y-y_{0}\right)+C\left(y-y_{0}\right)^{2}\right)\\\Rightarrow f\begin{pmatrix}x\\y\end{pmatrix} & =f\left(P_{0}\right)+\frac{1}{2}\cdot\left(A\left(x-x_{0}\right)^{2}+2B\left(x-x_{0}\right)\left(y-y_{0}\right)+C\left(y-y_{0}\right)^{2}\right)+\phi\left(P\right)\cdot\left\Vert P-P_{0}\right\Vert ^{2}\end{aligned}$$

המחובר השלישי שואף ל-0 כשמתקרבים ל-P_{0} ולכן אם נצליח להוכיח שהמחובר השני (יסומן מעתה ב-G) מקיים $0<M<G$ או $0>-M>G$ (עבור $0<M\in\mathbb{R}$ כלשהו) בסביבה קטנה מספיק של P_{0} נוכל לקבוע ש-P_{0} היא נקודת מינימום/מקסימום מקומית של f בהתאמה; אחרת, לכל סביבה של P_{0} המחובר השני מקבל סימנים מנוגדים בנקודות שונות ו-P_{0} היא נקודת אוכף של f.

לכל $\left(x_{0},y_{0}\right)\neq\left(x,y\right)\in U$ כך ש-$y=y_{0}$ הסימן של המחובר השני שווה לסימן של A (עוד מעט נראה למה זה אומר

---

[2] ערך בוויקיפדיה הרמן שוורץ.

[3] בשיעור קראנו למשפט זה בשם "הלמה של שוורץ" אך לא מצאתי לכך שום מקור ברשת (בעברית לא מצאתי מקור המזכיר את שם המשפט כלל), את השם "משפט שוורץ" מצאתי בוויקיפדיה האנגלית יחד עם שמות נוספים (ראו כאן).

[4] ובפרט מוגדרות בסביבה מלאה של P.

[5] נשתמש באותם סימונים שבקובץ ההגדרות.

שאין צורך להתייחס למקרה זה באופן מיוחד).
לכל $(x_0,y_0)\neq(x,y)\in U$ כך ש-$y\neq y_0$ מתקיים[6]:

$$\begin{aligned}&A\left(x-x_0\right)^2+2B\left(x-x_0\right)\left(y-y_0\right)+C\left(y-y_0\right)^2\\&=\left(y-y_0\right)^2\cdot\left(A\left(\frac{x-x_0}{y-y_0}\right)^2+2B\left(\frac{x-x_0}{y-y_0}\right)+C\right)\end{aligned}$$

אנחנו יודעים איך מתנהגות פרבולות ולכן נוכל לומר בביטחון שאם הדיסקרימיננטה - $4B^2-4AC=4\left(B^2-AC\right)$ - שלילית אז קיים $0<M\in\mathbb{R}$ כנ"ל שהוא או הנגדי שלו הם הערך של הפרבולה בקודקוד ($-\frac{2B}{2A}=-\frac{B}{A}$), ואז אם A חיובי[7] נוכל לומר שהפרבולה "מחייכת" (פורמלית: קמורה) ולכן קיימת סביבה קטנה מספיק של P_0 שבה $0<M<G$ ומכאן ש-P_0 היא נקודת מינימום, ואם A שלילי נוכל לומר שהפרבולה "עצובה" (פורמלית: קעורה) ולכן קיימת סביבה כנ"ל שבה מתקיים $0>-M>G$ ומכאן ש-P_0 היא נקודת מקסימום. אחרת, אם הדיסקרימיננטה חיובית אז G מקבל ערכים חיוביים ושליליים בכל סביבה של P_0[8] ולכן היא נקודת אוכף. כאשר הדיסקרימיננטה שווה ל-0 כל האפשרויות פתוחות, דוגמאות:

- x^4+y^4, ראשית הצירים היא נקודת מינימום.
- $-\left(x^4+y^4\right)$, ראשית הצירים היא נקודת מקסימום.
- x^4-y^4, ראשית הצירים היא נקודת אוכף.

לסיכום נראה את המשפט הבא.

משפט 2.5. תהא $U\subseteq\mathbb{R}^2$ ותהא $f:U\to\mathbb{R}$ פונקציה כך שהנגזרות החלקיות מסדר ראשון של f מתאפסות בנקודה $P_0\in U$ וכל הנגזרות החלקיות מסדר שני ב-P_0 רציפות ב-P_0, נסמן[9]:

$$\begin{aligned}\Delta&:=D_{11}f\left(P_0\right)\cdot D_{22}f\left(P_0\right)-D_{12}f\left(P_0\right)\cdot D_{21}f\left(P_0\right)\\&=D_{11}f\left(P_0\right)\cdot D_{22}f\left(P_0\right)-\left(D_{12}f\left(P_0\right)\right)^2\end{aligned}$$

ואז מתקיימים שלושת הפסוקים הבאים[10]:

- אם $\Delta>0$ ו-$D_{11}f\left(P_0\right)>0$ אז P_0 היא נקודת מינימום מקומית של f.
- אם $\Delta>0$ ו-$D_{11}f\left(P_0\right)<0$ אז P_0 היא נקודת מקסימום מקומית של f.
- אם $\Delta<0$ אז P_0 היא נקודת אוכף של f.

[6] באותה מידה היה ניתן לדרוש ש-$x\neq x_0$ ולחלק ב-$x-x_0$ ואז התפקידים של A ו-C בהמשך יתהפכו.
[7] כשהדיסקרימיננטה שלילית הסימנים של A ושל C שווים אבל ברור למה בחרנו ב-A.
[8] בכל סביבה של P_0 קיימות נקודות (x,y) בסביבה כך ש-$\frac{x-x_0}{y-y_0}$ מקבלות כל מספר ממשי שנרצה.
[9] שימו לב שהסימון הוא הנגדי של הדיסקרימיננטה מההסבר שלעיל, את ההסבר להיפוך נראה בהמשך.
[10] מהגדרה לא ייתכן ש-$\Delta>0$ ו-$D_{11}f\left(P_0\right)=0$

מסקנה 2.6. תהא $U\subseteq\mathbb{R}^2$ ותהא $f:U\to\mathbb{R}$ פונקציה כך שהנגזרות החלקיות מסדר ראשון של f מתאפסות בנקודה $P_0\in U$ וכל הנגזרות החלקיות מסדר שני ב-P_0 רציפות ב-P_0 ונסמן ב-H את מטריצת הסיאן של f ב-P_0; מהגדרה H סימטרית ולכן לכסינה ומכאן שיש לה שני ערכים עצמיים, מתקיימים שלושת הפסוקים הבאים:

- אם שניהם חיוביים אז P_0 היא נקודת מינימום מקומית של f.
- אם שניהם שליליים אז P_0 היא נקודת מקסימום מקומית של f.
- אם אחד מהם חיובי והאחר שלילי אז P_0 היא נקודת אוכף של f.

♣ זה אולי נראה כמו קסם אבל יש לזה הסבר אלגברי פשוט מאד.
הפולינום האופייני של H הוא (נשתמש בסימונים שבהערה הקודמת):

$$(T-A)(T-C)-B^2=T^2-(A+C)T+\left(AC-B^2\right)$$

ולכן השורשים שלו הם (שהם הערכים העצמיים של H):

$$\frac{(A+C)\pm\sqrt{(A+C)^2-4\left(AC-B^2\right)}}{2}$$

ואז:

- אם $\Delta>0$ אז $|A+C|>-\sqrt{(A+C)^2-4\left(AC-B^2\right)}\geq 0$ ולכן הערכים העצמיים בעלי אותו סימן וזהו הסימן של A ו-C[11].
- אם $\Delta<0$ אז $\sqrt{(A+C)^2-4\left(AC-B^2\right)}>|A+C|$ ולכן אחד הערכים העצמיים חיובי והאחר שלילי.

♣ ההסבר האמיתי לקשר בין הערכים העצמיים של ההסיאן לבין אפיון נקודה קריטית נעוץ בביטוי:

$$A\left(x-x_0\right)^2+2B\left(x-x_0\right)\left(y-y_0\right)+C\left(y-y_0\right)^2$$

לו היינו מקבלים $B=0$ היה ברור שהנקודה היא נקודת קיצון אם"ם $\text{sgn}A=\text{sgn}C$ ושהיא מינימום או מקסימום ע"פ האפיון הנ"ל, אבל מה קורה כאשר $B\neq 0$?
כאן צריך לזכור שאין שום דבר מיוחד בבסיס הסטנדרטי, אנחנו פשוט עובדים עם בסיס אחר שייתן לנו $B=0$, איך אנחנו יודעים שניתן לעשות זאת? מפני שההסיאן סימטרית ולכן לכסינה, לכסון הוא פשוט מעבר לבסיס של וקטורים עצמיים וכאשר נעבור לבסיס כזה נקבל מהגדרה ש-A ו-C הם הערכים העצמיים.

---

[11] כפי שראינו סימניהם זהים.

3 כלל השרשרת

♣ תזכורת: בקובץ ההגדרות ציינתי שהגדרת הדיפרנציאביליות עובדת באותה צורה גם עבור פונקציות מהצורה $f:\mathbb{R}^n\to\mathbb{R}^m$ כאשר $m>1$ (כמובן שגם הטווח של הדיפרנציאל צריך להשתנות בהתאם); ואז הגדרת הדיפרנציאביליות של פונקציות מהצורה $f:\mathbb{R}^n\to\mathbb{R}^m$ כוללת בתוכה גם את הגדרת הדיפרנציאביליות שראינו עבור פונקציות מהצורה $f:\mathbb{R}^n\to\mathbb{R}$ וגם את הגדרת הגזירות של מסילות, ומוסיפה על אלו את כל הפונקציות שעבורן $n,m>1$.
כעת אני רוצה להשתמש בידע הזה כדי לתת כאן את כלל השרשרת בצורתו המלאה (עבור פונקציות מ-$\mathbb{R}^n$ ל-$\mathbb{R}^m$) וזאת למרות ש**לא למדנו אותו בכיתה**, לאחר מכן נראה ששתי הגרסאות של כלל השרשרת שלמדנו בכיתה הן בעצם מקרה פרטי של כלל השרשרת המלא.

משפט. כלל השרשרת המלא
תהיינה $f:A\to B$ ו-$g:B\to\mathbb{R}^k$ (כאשר $A\subseteq\mathbb{R}^n$ ו-$B\subseteq\mathbb{R}^m$) כך ש-f דיפרנציאבילית בנקודה $a\in A$ ו-g דיפרנציאבילית ב-$f(a)$, $g\circ f$ דיפרנציאבילית ב-a ומתקיים:

$$D_{g\circ f}(a)=D_g(f(a))\circ D_f(a)$$

♣ נזכור שישנו איזומורפיזם בין מרחב ההעתקות הליניאריות ממרחב וקטורי אחד למשנהו (במקרה זה $\mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m)$ ו-$\mathrm{Hom}\left(\mathbb{R}^m,\mathbb{R}^k\right)$) לבין מרחב המטריצות בגודל המתאים (במקרה זה $M_{m\times n}(\mathbb{R})$ ו-$M_{k\times m}(\mathbb{R})$ בהתאמה) ושהרכבת העתקות ליניאריות שקולה לכפל המטריצות המייצגות שלהן, א"כ כלל השרשרת אומר שמתקיים (E הוא הבסיס הסטנדרטי):

$$[D_{g\circ f}(a)]_E=[D_g(f(a))]_E\cdot[D_f(a)]_E$$

כאן כבר קל לראות שכלל השרשרת הנ"ל מכליל את כלל השרשרת שלמדנו באינפי' 1 שהרי באינפי' 1 ההעתקות הליניאריות המתאימות[12] מיוצגות ע"י מטריצה מגודל 1×1 ולכן כפל המטריצות הוא בעצם כפל בשדה $\mathbb{R}$, כלומר שני הביטויים הבאים שקולים:

$$[D_{g\circ f}(a)]_E=[D_g(f(a))]_E\cdot[D_f(a)]_E$$
$$(g\circ f)'(a)=D_{g\circ f}(a)=D_g(f(a))\cdot D_f(a)=g'(f(a))\cdot f'(a)$$

בעמודים הבאים נראה כיצד גם שתי הגרסאות של כלל השרשרת שלמדנו בכיתה הן בעצם מקרה פרטי של כלל השרשרת הנ"ל.

[12] ראו את האינטואיציה שלפני הגדרת הדיפרנציאביליות (בקובץ ההגדרות כמובן).

משפט 3.1. כלל השרשרת - גרסה ראשונה

תהיינה $x,y:\mathbb{R}\to\mathbb{R}$ כך ש-x ו-y גזירות בנקודה $t_0\in\mathbb{R}$ ותהא $f:\mathbb{R}^2\to\mathbb{R}$ פונקציה דיפרנציאבילית בנקודה $\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}$, נגדיר $g:\mathbb{R}\to\mathbb{R}$ ע"י $g(t):=f\begin{pmatrix}x(t)\\y(t)\end{pmatrix}$ (לכל $t\in\mathbb{R}$); g גזירה ב-t_0 ומתקיים:

$$g'(t_0)=f'_x\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\cdot x'(t_0)+f'_y\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\cdot y'(t_0)$$

♣ נשים לב לכך ש-x ו-y מגדירות יחדיו פונקציה $h:\mathbb{R}\to\mathbb{R}^2$ ע"י:

$$h(t)=\begin{pmatrix}x(t)\\y(t)\end{pmatrix}$$

כלומר g היא בעצם $f\circ h$. נזכר שראינו בפרק שעסק במסילות שעובדת היותן של x ו-y גזירות ב-t_0 גוררת את גזירותה של h ב-t_0 וכפי שהזכרנו לעיל הגזירות שראינו כשעסקנו במסילות היא בעצם דיפרנציאביליות של פונקציות מ-$\mathbb{R}$ ל-$\mathbb{R}^n$, א"כ h דיפרנציאבילית ב-t_0 ולכן מכלל השרשרת המלא נובע שמתקיים:

$$\left[D_{f\circ h}(t_0)\right]_E=\left[D_f(h(t_0))\right]_E\cdot\left[D_h(t_0)\right]_E$$

והרי מהגדרה מתקיים[13]:

$$\begin{aligned}\left[D_f(h(t_0))\right]_E&=\begin{bmatrix}f'_x(h(t_0)) & f'_y(h(t_0))\end{bmatrix}\\&=\begin{bmatrix}f'_x\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix} & f'_y\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\end{bmatrix}\\\left[D_h(t_0)\right]_E&=\begin{bmatrix}x'(t_0)\\y'(t_0)\end{bmatrix}\end{aligned}$$

$$\Rightarrow\left[D_{f\circ h}(t_0)\right]_E=\begin{bmatrix}f'_x\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix} & f'_y\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\end{bmatrix}\cdot\begin{bmatrix}x'(t_0)\\y'(t_0)\end{bmatrix}=\begin{bmatrix}f'_x\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\cdot x'(t_0)+f'_y\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\cdot y'(t_0)\end{bmatrix}$$

ומכיוון שהאגפים הקיצוניים הם בעצם מטריצות מגודל 1×1 נקבל:

$$g'(t_0)=(f\circ h)'(t_0)=D_{f\circ h}(t_0)=f'_x\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\cdot x'(t_0)+f'_y\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\cdot y'(t_0)$$

[13] נזכור שהגרדיאנט הוא בעצם המשוחלפת של המטריצה המייצגת.

משפט 3.2. כלל השרשרת - גרסה שנייה

תהיינה $x,y:\mathbb{R}^2\to\mathbb{R}$ כך שהנגזרות החלקיות שלהן קיימות בנקודה $P_0=\begin{pmatrix}s_0\\t_0\end{pmatrix}$ ותהא $f:\mathbb{R}^2\to\mathbb{R}$ פונקציה דיפרנציאבילית בנקודה $\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}$, נגדיר פונקציה $g:\mathbb{R}^2\to\mathbb{R}$ ע"י $g\begin{pmatrix}s\\t\end{pmatrix}:=f\begin{pmatrix}x\binom{s}{t}\\y\binom{s}{t}\end{pmatrix}$ (לכל $\begin{pmatrix}s\\t\end{pmatrix}\in\mathbb{R}^2$); הנגזרות החלקיות של g ב-P_0 קיימות ומתקיים:

$$g'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}=f'_x\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}\cdot x'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}\cdot y'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}$$
$$g'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}=f'_x\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}\cdot x'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}\cdot y'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}$$

♣ הגרסה הזו היא בעצם הגרסה הקודמת בתחפושת, להלן ההסבר לכך.
תהיינה $x_1,x_2,y_1,y_2:\mathbb{R}\to\mathbb{R}$ המוגדרות ע"י (לכל $s,t\in\mathbb{R}$):

$$y_1(s):=y\begin{pmatrix}s\\t_0\end{pmatrix}\qquad x_1(s):=x\begin{pmatrix}s\\t_0\end{pmatrix}$$
$$y_2(t):=y\begin{pmatrix}s_0\\t\end{pmatrix}\qquad x_2(t):=x\begin{pmatrix}s_0\\t\end{pmatrix}$$

מקיומן של הנגזרות החלקיות של x ו-y ב-P_0 נובע ש-x_1,x_2,y_1,y_2 גזירות ב-$s=s_0$ וב-$t=t_0$ בהתאמה ומתקיים:

$$y'_1(s_0):=y'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}\qquad x'_1(s_0):=x'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}$$
$$y'_2(t_0):=y'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}\qquad x'_2(t_0):=x'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}$$

נגדיר שתי פונקציות חדשות $g_1,g_2:\mathbb{R}\to\mathbb{R}$ ע"י (לכל $s,t\in\mathbb{R}$):

$$g_1(s):=g\begin{pmatrix}s\\t_0\end{pmatrix}=f\begin{pmatrix}x_1(s)\\y_1(s)\end{pmatrix}$$
$$g_2(t):=g\begin{pmatrix}s_0\\t\end{pmatrix}=f\begin{pmatrix}x_2(t)\\y_2(t)\end{pmatrix}$$

ומהגרסה הקודמת נובע שמתקיים:

$$\begin{aligned}g'_1(s_0)&=f'_x\begin{pmatrix}x_1(s_0)\\y_1(s_0)\end{pmatrix}\cdot x'_1(s_0)+f'_y\begin{pmatrix}x_1(s_0)\\y_1(s_0)\end{pmatrix}\cdot y'_1(s_0)\\&=f'_x\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}\cdot x'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}\cdot y'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}\\g'_2(t_0)&=f'_x\begin{pmatrix}x_2(t_0)\\y_2(t_0)\end{pmatrix}\cdot x'_2(t_0)+f'_y\begin{pmatrix}x(t_0)\\y(t_0)\end{pmatrix}\cdot y'_2(t_0)\\&=f'_x\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}\cdot x'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x(P_0)\\y(P_0)\end{pmatrix}\cdot y'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}\end{aligned}$$

מהגדרת g_1 ו-g_2 נקבל שאם g_1 גזירה ב-s_0 אז הנגזרת החלקית $g'_s(P_0)$ קיימת ושווה לה וכמו כן אם g_2 גזירה ב-t_0

אז הנגזרת החלקית $g'_t\left(P_0\right)$ קיימת ושווה לה, כלומר:

$$g'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}=g'_1\left(s_0\right)=f'_x\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot x'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot y'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}$$
$$g'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}=g'_2\left(t_0\right)=f'_x\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot x'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot y'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}$$

♣ אם x ו-y הן פונקציות המקבלות שלושה משתנים (וממילא גם g תהיה כזו) נקבל באותה דרך שמתקיים:

$$g'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}=f'_x\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot x'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot y'_s\begin{pmatrix}s_0\\t_0\end{pmatrix}$$
$$g'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}=f'_x\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot x'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot y'_t\begin{pmatrix}s_0\\t_0\end{pmatrix}$$
$$g'_u\begin{pmatrix}s_0\\t_0\end{pmatrix}=f'_x\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot x'_u\begin{pmatrix}s_0\\t_0\end{pmatrix}+f'_y\begin{pmatrix}x\left(P_0\right)\\y\left(P_0\right)\end{pmatrix}\cdot y'_u\begin{pmatrix}s_0\\t_0\end{pmatrix}$$

כמובן שאפשר להמשיך לארבעה משתנים ויותר וכמו כן גם ניתן להגדיל את מספר המשתנים שמקבלת f אלא שאז נצטרך להוסיף מחוברים ולא שוויונות.

3.1 פונקציה סתומה

דוגמה. הקבוצה $C:=\left\{(x,y)\in\mathbb{R}^2\mid x^2+y^2=1\right\}$ המהווה את מעגל היחידה אינה יכולה להיות גרף של פונקציה אך עדיין היינו רוצים לדבר על המשיק לקבוצה בנקודה על המעגל; הרעיון הוא להגדיר פונקציה $F:\mathbb{R}^2\to\mathbb{R}$ ע"י $f\left(x,y\right)=x^2+y^2-1$ ואז לכל $(x,y)\in C$ מתקיים $f\left(x,y\right)=0$, כלומר C היא קו הגובה 0 של f, כעת לכל נקודה $c:=\left(x_0,y_0\right)\in C$ קיימת סביבה מלאה שלה כך שקיימת פונקציה $y:B_r\left(c\right)\to\mathbb{R}$ המוגדרת ע"י $y\left(x\right):=\operatorname{sgn}\left(y_0\right)\cdot\sqrt{1-x^2}$. א"כ בסביבה זו מתקיים $f\left(x,y\right)=f\left(x,y\left(x\right)\right)$ ואחרי שנגדיר $g\left(x\right):=f\left(x,y\left(x\right)\right)$ בסביבה זו (נשים לב שמהגדרה g היא פונקציית האפס) נקבל מכלל השרשרת שמתקיים:

$$0=g'\left(x\right)=f'_x\begin{pmatrix}x_0\\y\left(x_0\right)\end{pmatrix}\cdot 1+f'_y\begin{pmatrix}x_0\\y\left(x_0\right)\end{pmatrix}\cdot y'\left(x_0\right)$$

וסוף סוף מופיע הביטוי המיוחל $y'\left(x_0\right)$ במשוואה שממנה ניתן לחלץ אותו שכן אם $f'_y\left(x_0,y\left(x_0\right)\right)\neq 0$ אז מתקיים:

$$y'\left(x_0\right)=-\frac{f'_x\left(x_0,y\left(x_0\right)\right)}{f'_y\left(x_0,y\left(x_0\right)\right)}$$

♣ הרעיון הוא כמובן שאנו מצטמצמים לסביבה מספיק קטנה של הנקודה כך שבה $C\cap B_r\left(c\right)$ כן מהווה גרף של פונקציה.

משפט 3.3. משפט הפונקציה הסתומה

תהא $f:\mathbb{R}^n\to\mathbb{R}$ פונקציה המוגדרת בסביבה U_1 נקודה $(x_1,x_2,\ldots,x_n)\in\mathbb{R}^n$ המקיימת את התנאים הבאים:

1. $f(P)=0$ לכל $P\in U_1$.
2. כל הנגזרות החלקיות של f מוגדרות ורציפות ב-U_1.
3. $D_nf(x_1,x_2,\ldots,x_n)\neq 0$.

קיימת סביבה U_2 של $b:=(x_1,x_2,\ldots,x_{n-1})\in\mathbb{R}^{n-1}$ וקיימת פונקציה יחידה $g:U_2\to\mathbb{R}$ המקיימת את המשוואה $f(y_1,y_2,\ldots,y_{n-1},g(y_1,y_2,\ldots,y_{n-1}))=0$ ובעלת התכונות הבאות:

1. $x_n=g(x_1,x_2,\ldots,x_{n-1})$
2. g רציפה ב-U_2.
3. כל הנגזרות החלקיות של g קיימות ב-$(x_1,x_2,\ldots,x_{n-1})$ ולכל $n-1\geq i\in\mathbb{N}$ מתקיים:
$$D_ig(x_1,x_2,\ldots,x_{n-1})=-\frac{D_if(x_1,x_2,\ldots,x_n)}{D_nf(x_1,x_2,\ldots,x_n)}$$

♣ אנו מגדירים את f ולכן נגדיר אותה כך שיתקיים את התנאי $f(P)=0$ לכל $P\in U_1$, הדרך לעשות זאת היא לבודד את 0 באחד האגפים של משוואת הפונקציה הסתומה ואז להגדיר את f ע"פ האגף השני[14].

♣ בדוגמה שלעיל ובכלל בפונקציות בשתי משתנים ישנה אינטואיציה חזקה מאד לנוסחה הנ"ל, שימו לב שהגרדיאנט של f ב-c[15] הוא:
$$\vec{\nabla}f\begin{pmatrix}x_0\\y_0\end{pmatrix}=\begin{bmatrix}f'_x(x_0,y(x_0))\\f'_y(x_0,y(x_0))\end{bmatrix}$$
ולכן אחד הווקטורים המאונכים לו[16] הוא:
$$\begin{bmatrix}f'_y(x_0,y(x_0))\\-f'_x(x_0,y(x_0))\end{bmatrix}$$
והשיפוע של ישר הנפרש ע"י וקטור זה הוא בדיוק:
$$-\frac{f'_x(x_0,y(x_0))}{f'_y(x_0,y(x_0))}$$
אם נזכור שהמשיק לקו הגובה מאונך לגרדיאנט נבין למה זה קרה.

[14] זה מה שעשינו בדוגמה שלעיל.
[15] f אכן דיפרנציאבילית ב-c.
[16] כל השאר הם כפל בסקלר כמובן.

4 חזרה לנקודות קיצון

♣ עד כה ראינו כיצד ניתן למצוא ולמיין נקודות קריטיות של פונקציות כשהן מוגדרות על קבוצות פתוחות (או קומפקטיות ששפתן אינה הקבוצה כולה) אך מה נעשה כשהפונקציה מוגדרת על קבוצה קומפקטית שכולה שפה (עקומה חד-ממדית במרחב)?

משפט 4.1. משפט כופלי לגראנז'[17]

תהיינה $f,g:D\to\mathbb{R}$ (כאשר $D\subseteq\mathbb{R}^{n}$ קבוצה פתוחה) פונקציות בעלות נגזרות חלקיות רציפות ב-D, אם $P_{0}\in D$ היא נקודת קיצון כללית של f כשהיא מצומצמת לקבוצה $\left\{ P\in D\mid g\left(P\right)=0\right\}$, וגם $\overrightarrow{\nabla}g\left(P_{0}\right)\neq0$ אז קיימת $\lambda\in\mathbb{R}$ כך שמתקיים:

$$\overrightarrow{\nabla}f\left(P_{0}\right)=\lambda\cdot\overrightarrow{\nabla}g\left(P_{0}\right)$$

אותה λ היא:

$$\lambda=\frac{D_{i}f\left(P_{0}\right)}{D_{i}g\left(P_{0}\right)}$$

עבור כל אחת מהקואורדינטות שעבורן $D_{i}g\left(P_{0}\right)\neq0$.

♣ הפתרון לשאלה שבהערה הקודמת הוא להרחיב את תחום הפונקציה לקבוצה פתוחה ואז כל נקודת קיצון שלה תצטרך לקיים את משפט כופלי לגראנז', לאחר מכן נמיין את הנקודות כרגיל.

♣ כמו במשפט הפונקציה הסתומה גם כאן נגדיר את g ע"פ האילוץ שקיבלנו כך שתקיים שקבוצת האילוץ היא בדיוק $\left\{ P\in D\mid g\left(P\right)=0\right\}$.

משפט 4.2. דרך שקולה לבטא את משפט כופלי לגראנז'

תהיינה $f,g:D\to\mathbb{R}$ (כאשר $D\subseteq\mathbb{R}^{n}$ קבוצה פתוחה) פונקציות בעלות נגזרות חלקיות רציפות ב-D, אם $P_{0}\in D$ היא נקודת קיצון כללית של f כשהיא מצומצמת לקבוצה $\left\{ P\in D\mid g\left(P\right)=0\right\}$וגם $\overrightarrow{\nabla}g\left(P_{0}\right)\neq0$ אז P_{0} היא נקודה קריטית של הפונקציה $h:D\times\mathbb{R}\to\mathbb{R}$ המוגדרת ע"י $h\left(x_{1},x_{2},\ldots,x_{n},\lambda\right):=f\left(x_{1},x_{2},\ldots,x_{n}\right)+\lambda\cdot g\left(x_{1},x_{2},\ldots,x_{n}\right)$.

♣ העובדה ש-P_{0} היא נקודה קריטית של h אומרת שלכל $n\geq i\in\mathbb{N}$ מתקיים[18]:

$$\begin{aligned}0 & =D_{i}h\left(P_{0}\right)=D_{i}f\left(P_{0}\right)+\lambda\cdot D_{i}g\left(P_{0}\right)\\ \Longleftrightarrow & D_{i}f\left(P_{0}\right)=-\lambda\cdot D_{i}g\left(P_{0}\right)\end{aligned}$$

נשאר רק לחלק ב-$-D_{i}g\left(P_{0}\right)$.

---

[17] ערך בוויקיפדיה: ז'וזף-לואי לגראנז'.

[18] ה-λ פה היא הנגדית של ה-λ מהניסוח הקודם.

משפט 4.3. הכללה של משפט כופלי לגראנז'

תהיינה $f, g_1, g_2, \ldots, g_k : D \to \mathbb{R}$ (כאשר $D \subseteq \mathbb{R}^n$ קבוצה פתוחה) פונקציות בעלות נגזרות חלקיות רציפות ב-D, אם ש-$P_0 \in D$ היא נקודת קיצון כללית של f כשהיא מצומצמת לקבוצה

$$\left\{ P \in D \;\middle|\; \begin{cases} g_1(P) = 0 \\ g_2(P) = 0 \\ \vdots \\ g_1(P) = 0 \end{cases} \right\}$$

אז P_0 היא נקודה קריטית של הפונקציה $h : D \times \mathbb{R}^k \to \mathbb{R}$ המוגדרת ע"י:

$$h(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots \lambda_k) := f(x_1, x_2, \ldots, x_n) + \sum_{j=1}^{k} \lambda_j \cdot g_j(x_1, x_2, \ldots, x_n)$$

ובתנאי שלמערכת המשוואות הליניארית הבאה יש פתרון[19]:

$$\begin{bmatrix} D_1 f(P_0) \\ D_2 f(P_0) \\ \vdots \\ D_n f(P_0) \end{bmatrix} + \begin{bmatrix} D_1 g_1(P_0) & D_1 g_2(P_0) & \ldots & D_1 g_k(P_0) \\ D_2 g_1(P_0) & D_2 g_2(P_0) & \ldots & D_2 g_k(P_0) \\ \vdots & \vdots & \ddots & \vdots \\ D_n g_1(P_0) & D_n g_2(P_0) & \ldots & D_n g_k(P_0) \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\iff \begin{bmatrix} D_1 g_1(P_0) & D_1 g_2(P_0) & \ldots & D_1 g_k(P_0) \\ D_2 g_1(P_0) & D_2 g_2(P_0) & \ldots & D_2 g_k(P_0) \\ \vdots & \vdots & \ddots & \vdots \\ D_n g_1(P_0) & D_n g_2(P_0) & \ldots & D_n g_k(P_0) \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix} = \begin{bmatrix} -D_1 f(P_0) \\ -D_2 f(P_0) \\ \vdots \\ -D_n f(P_0) \end{bmatrix}$$

[19] צריך לבדוק אם זה באמת התנאי הדרוש.