כמה מההיבטים של אלגברה ליניארית

בפרק זה נעמוד על כמה מהבעיות האלגוריתמיות המתעוררות בלימוד האלגברה הליניארית. הרבה מהנושאים הללו הופיעו בלימודיכם בקורסים הרלוונטיים אולם כאן נעמיק את הדיון האלגוריתמי. נעיר תחילה על היבט חשוב נוסף לבעיות אלה, המדובר ביציבות נומרית. כאשר יש לבצע חישובים במספרים ממשיים אנו נתקלים בקושי הנובע מכך שבמחשב מיוצגים המספרים בדרגת דיוק חסומה הנובעת מאורך המילה (מספר הביטים) שבו מיוצגים המספרים במחשב שלנו. כתוצאה מזה החישובים שאנו מבצעים במחשב הם רק בגדר קירוב לתוצאה התיאורטית. שיקולים אלה מובילים מצד אחד לתורה של אנליזה נומרית. במסגרת זו מחפשים שיטות חישוב שיושפעו מעט ככל האפשר מבעיות הדיוק המוגבל במחשב. זהו תחום שלם שלא נדון בו כאן. גם על מבנה החומרה של מחשבים משפיעים השיקולים האלה באופן מהותי. ייצוגם של מספרים ממשיים בשיטת נקודה צפה ואחרות ואופן המימוש של הפעולות האריתמ-טיות היסודיות הם בין השיקולים המרכזיים בתכנון מחשבים. אלו נושאים המכוסים בקורס "מבנה מחשב".

עוד הסתייגות אחת לפני שניכנס בעובי הקורה:

הדיון בהיבטים החישוביים של אלגברה ליניארית הוא רחב ועמוק. נושא זה בפני עצמו יכול לספק בקלות חומר לקורס אוניברסיטאי שלם ואנו נאלץ רק לטעום מקצת ממנו.

להלן סיכום העניינים העיקריים שבהם נדון בפרק זה:

א. נזכיר כמה בעיות אלגוריתמיות/אלגבריות בסיסיות שבהן האלגוריתם כהגדרתו אופטימלי (בזה לא נדון לעומק).

ב. מאידך גיסא, נראה שתי בעיות מרכזיות שבהן יכול אלגוריתם מתוחכם חסכון של ממש: כפל מטריצות (ע"פ Strassen) וכפל פולינומים (בעזרת טרנספורטים פורייה הבדיד DFT)

ג. גם כאשר איננו יודעים מהי הסיבוכיות של בעיות אלגבריות מסויימות עדיין ניתן לומר דבר מה על קושיין היחסי. נערוך כמה השוואות מסוג זה.

ד. נכניס מערכת מושגים המאפשרת לדון בחישוב אלגברי מקורב ונדגים כמה משימושיה.

ה. נציין את החשיבות שבמציאת פירוקים למטריצות: בפרט PLU ו- SVD.

ו. נסביר את הרקע הרחב יותר שמאחורי טרנספורים פורייה הבדיד (DFT) וכמה משימושיו.

ז. נראה איך במגוון של הקשרים מועיל להציג (במדוייק או בקירוב) אובייקט מתימטי מסובך כצירוף של "אבני בניין" פשוטות יותר.

יש כמה מהפעולות הבסיסיות של אלגברה ליניארית שבהן האלגוריתם הנאיבי הנובע מן ההגדרה הוא אופטימלי. כך למשל חישוב המכפלה הסקלרית ("מכפלה פנימית") של שני ווקטורים n- מימדיים $\langle x,y\rangle=\sum x_iy_i$ אכן מצריך n פעולות כפל. זו עובדה לא טריביאלית שלא נוכיח כאן ומכוסה בקורסים בסיבוכיות החישוב. כיוצא בו גם כפל של מטריצה ממשית $A_{n\times n}$ בווקטור n- מימדי x, כלומר, $(Ax)_i=\sum_i a_{ij}x_j$ אכן מצריך n^2 פעולות כפל, כלומר אין לבעיה זו אלגוריתם יעיל יותר מאשר האלגוריתם הנאיבי הנובע מן ההגדרה. זוהי שוב תוצאה מתקדמת למדי בתורה של סיבוכיות החישוב והיא אינה מוכחת בקורס זה. לעומת זאת החישוב ע"פ ההגדרה של כפל מטריצות ממשיות $AB=C\ nxn$ מצריך n^3 פעולות כפל. בהמשך נכיר אלגוריתם יעיל יותר (של Strassen) המצריך רק $o\left(n^{2.81}\right)$ פעולות אריתמטיות. מרגע שמתבררת נקודת המבט שלנו. מתעוררת באחת שאלות אלגוריתמיות רבות: מהי הסיבוכיות של חישוב הדטרמיננט של מטריצה nxn? של פתרון מערכת משוואות ליניאריות $Ax=b$?

מהו המחיר החישובי של הצבת ערך בפולינום? של כפל פולינומים? ועוד ועוד.

נפתח בדיון קצר בפתרון מערכות של משוואות ליניאריות, נזכיר ראשית את שיטת האלימינציה של גאוס. בשיטה זו פותרים את מערכת המשוואות הליניאריות $Ax=b$ ע"י כך שמציגים את A בצורה $A=PLU$. כאן P היא מטריצת תמורה, L היא מטריצה

משולשת תחתונה ו- U מטריצה משולשת עליונה. יוצא שלמערכת הליניארית $Ax = b$ יש הצורה $LUx = P^{-1}b = u$ מערכת מהצורה $Ly = u$ ניתן לפתור באופן סידרתי כך: $y_i = \frac{u_1}{\ell_{11}}$ אז מציבים ופותרים את המשוואה ל- y_2 וכן הלאה. אז עלינו לפתור מערכת מהסוג $Ux = y$ ואותה ניתן לפתור באופן דומה, הפעם כשמתחילים ב- $x_n = \frac{y_n}{u_{nn}}$.

אנו נרצה להרחיב את היריעה ולדון במצבים שבהם מערכת המשוואות $Ax = b$ אינה אלא קירוב לבעייה האמיתית. כך למשל אנו יכולים לתאר מצבים רבים שבהם איברי המטריצה A כמו גם איברי הווקטור b מתקבלים ע"י מדידה של מערכת טבע כלשהי ולכן נגועים ברעש. מתברר גם כי אותו הדיוק עוזר לטפל בבעיות הנובעות משגיאות עיגול במשתנים שבווקטור x, שגיאות אלה מקורן, כאמור, במגבלות הדיון של המחשבים שלנו.

על מנת להמחיש את הרעיון נתאר מצב כזה: המשתנים הנעלמים שהם הקואור-דינטות של הווקטור x הם n גדלים פיזיקליים שאותם ברצוננו לקבוע. אם כל ניסוי שביכולתנו מגדיר משוואה ליניארית אחת, אז באופן תיאורטי n ניסויים בלתי תלויים הכרחיים ומספיקים על מנת לקבוע את הווקטור ה- n מימדי x ביחידות.

אולם, בהיותנו מודעים לבעיית הרעש בניסויים, אנחנו מקפידים ומבצעים מספר גדול באופן ניכר $n \gg m$ של ניסויים. אנו מנסים, אם כן לפתור מערכת משוואות ליניאריות $Ax = b$ שבה A מטריצה ממשית mxn, x ווקטור n מימדי ו- b ווקטור m מימדי. אילו היו המדידות שלנו נטולות רעש, היתה המערכת הנ"ל קובעת את x באופן חד ערכי. דא עקא שהמדידות שלנו מושפעות מרעש והמערכת הנ"ל כרגיל כלל אינה פתירה (באופן כללי, אם במערכת משוואות ליניאריות יש יותר משוואות מנעלמים, אנו איננו מצפים שתהיה פתירה כלל).

יוצא שהפרשנות הסבירה היחידה למערכת המשוואות $Ax = b$ היא "מצא ווקטור x כך שהווקטור Axיהיה קרוב ככל האפשר לווקטור הנתון b". על מנת למלא את השאלה הזו בתוכן עלינו לפתח תחילה מושג של "קירבה בין ווקטורים", מה שמוביל אותנו לדיוננו הבא.

נורמות ווקטוריות: יהיה V מרחב ווקטורי מעל הממשיים. אנו רוצים לייחס לוו-קטורים $v \in V$ "אורך". את האורך של v נסמן ב- $\|v\|$ ונקרא לו הנורמה של v. נצפה ממושג הנורמה לשתי תכונות:

א. אי שיוון המשולש:$\forall_{u,u \in V} \|u + v\| \le \|u\| + \|v\|$.

ב. הומוגניות: לכל $u \in V$ ולכל $\alpha \in R$ $\|\alpha v\| = |\alpha| \, \|v\|$

האם יש נורמות כנ"ל?

מסתבר שיש אינסוף נורמות שונות ויש גם נורמות רבות ושונות שהן מעניינות. שלוש הנורמות החשובות ביותר נקראות נורמות ℓ_2, נורמת ℓ_1 ונורמת ℓ_∞. להלן הגדר-ותיהן והסימון המקובל להן. נניח ש- V הוא מרחב n מימדי

$$x = (x_1, \_, x_n)$$

הוא ווקטור אופייני ב- V .

את נורמת ℓ_2 אתם מכירים בעצם מקדמת דנא $\|x\|_2 := \sqrt{\sum_i^n x_i^2}$

נורמת ℓ_1 מוגדרת כך: $\|x\|_1 := \sum_i^n |x_i|$

ונורמת ℓ^∞: $\|x\|_\infty := max\,|x_i|$

נעיר ששלושתן הן מקרים פרטיים של משפחה אינסופית של נורמות הקרויות נורמות ℓ_p (כאן $1 \le p < \infty$) $\|x\|_p := (\sum |x_i|^p)^{\frac{1}{p}}$

שימו לב ש- $lim_{p\to\infty} \|x\|_p = \|x\|_\infty$

(הוכיחו)

עתה נוכל לנסח את שאלתנו הקודמת בצורה מלאה: נקבע איזושהי נורמה ווקטור-ית $\|\bullet\|$ על ווקטורים m מימדים.

קלט: מטריצה ממשית A_{mxn} כש $m > n$ ודרגתה $rank\,(A) = n$ ווקטור m מימדי $.b$

פלט: ווקטור n- מימדי x כך ש $||Ax - b||$מזערי (הווקטור Axקרוב ל- b ככל הא-פשר).

כפי שכבר רמזנו הנורמה החשובה ביותר וכרגיל (אך לא תמיד) גם הנוחה ביותר לחישוב היא נורמת ℓ_2 זהו אכן המקרה שבו נטפל. כאשר הנורמה שבשימוש היא נורמת ℓ_2 נקראת הבעייה הנ"ל בעיית הקירוב בריבועים פחותים (least squares approximation).

בהינתן A, b כנ"ל מחפשים אם כן x כך ש: $min \sum_i \left(\sum_j a_{ij}x_j - b_i\right)^2$

נגזור ע"פ x_k ונקבל כי לכל k:

$\sum_i a_{ik}\left(\sum_j a_{ij}x_j - b_i\right) = 0$

נסדר מחדש את המחוברים: $\sum_j x_j \sum_i a_{ik}a_{ij} = \sum_i a_{ik}b_i$ נרשום זאת כך: $\sum_j x_j \sum \left(a^T_{ki}\right) a_{ij} =$ $\sum_i a_{ik}b_i$ כלומר $\sum_j \left(A^T A\right)_{kj} x_j = (bA)_k$ כלומר $A^T Ax = A^T b$ נעיר ש- $A^T A$ לא סינגולרית, מפני שע"פ ההנחה $rank\,(A) = n$ אם $A^T Au = 0$ אז גם $u^T A^T Au = 0$ כלומר $||Au||^2 = 0$ כלומר $u = 0$ קיבלנו אם כן $x = \left(A^T A\right)^{-1} A^T b$.

הערה: כאן ראינו איך לפתור במקורב מערכת משוואות ליניאריות כשהקירוב נעשה ב- ℓ_2. באמצעות תכנון ליניארי ניתן לפתור את המקרה של פתרון מקורב ב- l_∞ אולם אין עוד מקרים מעניינים רבים בהם יש לנו אלגוריתמים יעילים לפתרון הבעייה.

דרך חליפית לביצוע החשבון הנ"ל:

$||Ax - b||_2^2 = (x^T A^T - b^T, Ax - b) = x^T A^T Ax - 2b^T Ax + ||b||^2$

נסמן $Z = A^T A$ אז הנ"ל שווה ל:

$$\sum_{k,l} x_k x_l z_{kl} - 2\sum_k (A^T b)_k x_k + ||b||^2$$

נגזור ע"פ x_k ונאפס אז נקבל $2\sum_l z_{kl}x_l = 2(A^T b)_k$ כלומר $Zx = A^T b$ $x = Z^{-1}A^T b = \left(A^T A\right)^{-1} A^T b$. למטריצה $\left(A^T A\right)^{-1} A^T$ קוראים הפסאודו הופכית של A או הופכית $Moore - Penrose$ של A.

תיאור חלופי לבעיית הפתרון המקורב של מערכת משוואות:

אוסף הווקטורים $L = \{Ax | x \in \Re^m\}$ הוא תת מרחב n מימדי של $\Re^m$. הבעייה היא למצוא את הנקודה ב-L הקרובה ביותר ל- b וזהו כמובן ההיטל של b על L. הנוסחה שפיתחנו לפסאודו הופכי נותנת ביטוי מפורש להיטל.

להלן שימוש יפה וחשוב לבעייה שפתרנו: התאמת עקומות (Curve fitting). נפתח בעובדה מתימטית בסיסית שאולי כבר מוכרת לכם מהקורסים באנליזה מתימטית:

יהיו (x_i, y_i) $i = 1, ..., n$ n נקודות במישור כשה - x_i שונים זה מזה. אז יש פולינום אחד ויחיד P ממעלה $n >$ כך שלכל i מתקיים $P\,(x_i) = y_i$. תזכורת מהירה על ההוכחה: יחידות נובעת מכך שאם P, Q שני פתרונות אז $P - Q = 0$ ב- n הנקודות $x_1, ...x_n$ אבל לפולינום P-Q יש לכל היותר n-1 שורשים אלא אם $P \equiv Q$.

קיום: ניתן להוכיח א' ע"י תרגום הבעייה למערכת משוואות ליניאריות במקד-מי הפולינום P. מטריצת המערכת היא מטריצת ונדרמונדה שאינה סינגולרית ולכן המערכת פתירה ופתרונה מספק את ה- P המבוקש. לחילופין: ניתן לרשום את P במ-פורש ע"י מה שנקרא נוסחת האינטרפולציה של לגרנז': $P\,(x) = \sum_i y_i \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}$.

הסבר: כאשר $x \equiv x_r$, כל המחוברים שבהם $i \neq r$ מתאפסים כי הגורם $x_r - x_r = 0$ מופיע במונה. מאידך המחובר ה-r נותן לנו y_r כי הגורם הכופל את y_r הוא 1.

שוב תהיינה -$(a_1, b_1) .... (a_p, b_p)$ונרצה למצוא פולינום ממעלה $p > q \geq$ המקרב את הנקודות האלה היטב ככל האפשר במובן של l_2. כלומר מחפשים $P\,(x) = \sum_{i=0}^q \propto_i x^i$

כך ש $min\sum_1 \left(P\left(a_i\right)-b_i\right)^2$ אם נגדיר $A=\left(\begin{array}{cccc} 1 & a_1 & .. & a_1^q \\ 1 & a_2 & ... & a_2^q \\ & & \vdots & \\ 1 & a_p & .. & a_p^q \end{array}\right)$ $b=\left(\begin{array}{c} b_1 \\ \vdots \\ b_p \end{array}\right)$ אז

רצוננו למצוא ווקטור $z=\left(\alpha_o,\ldots,\alpha_q\right)$ כך ש $min_z \left\|Az-b\right\|_2$ הביטוי שפיתחנו לעיל מספק אם כן תשובה לבעיית האינטרפולציה המקורבת.