שיטות איטרטיביות:

נזכור מערכת דינמית: המתחילה ב-u_0 יוצרת ממנו את u_1 וממנו את u_2 ... על ידי הכפלה סידרתי במט' ריבועית קבועה T מגודל $n\times n$:

$$u_k = T u_{k-1} = T\cdot(T\cdot u_{k-2}) = \dots = T^k u_0$$

במקרה ש T לכסינה: $T = V\Lambda V^{-1}$

$$u^k = T^k u_0 = [V\Lambda V^{-1}]^k u_0 = V\Lambda^k \underbrace{V^{-1}u_0}_{\pi_0} = V\Lambda^k \pi_0 \qquad (V\pi_0 = u_0)$$

$$\begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}\begin{bmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{bmatrix}\begin{bmatrix} \pi_0(1) \\ \vdots \\ \pi_0(n) \end{bmatrix} = \sum_{j=1}^{n} \pi_0(j)\lambda_j^k v_j$$

אם $\lambda<1$ הוקטורים u_k ישאפו ל-0!

הגדרה – מערכת דינמית מהצורה $u_k = Tu_{k-1} = T^k u_0$ תקרא <u>יציבה אסימפטוטית</u> אם מתקיים כי $u_k \xrightarrow[k\to\infty]{} 0$. באופן דומה, נאמר כי המטריצה T יציבה אם מתקיים כי $T^k \xrightarrow[k\to\infty]{} 0$

<u>הערה:</u> ההתכנסויות המתוארות, לוקטור האפס ולמטריצת האפס מתייחסות בכל איבר.

<u>הגדרה:</u> בהינתן T כלשהי, נגדיר את הרדיוס הספקטרלי שלה כ:

$$\rho\{T\} = \max_{1\le k\le n}|\lambda_k| = |\lambda_1|$$

כל ערכיה העצמיים בעלי גודל קטן ממש מ-1

<u>משפט:</u> מערכת דינמית מהצורה: $u_k = Tu_{k-1} = T^k u_0$ בה המטריצה T לכסינה תהיה יציבה אסימפטוטית אם מתקיים כי המטריצה T לכסינה ובנוסף, כלומר, כל ע"ע שלה קטנים מ-1.

$$\rho\{T\}<1 \longrightarrow |\lambda_i|<1 \quad \forall j\in[1,n]$$

במקרה הזה צריך T לכסינה למצוא ע"ע שלה, לקחת את הכי גדול ולוודא שהוא קטן מ-1 ואז המטריצה יציבה.

תכונות והגדרה שלהן:

• לינאריות:

$$\forall v,u,w\in V,\ \forall \alpha,\beta\in F:\ \langle \alpha v+\beta u, w\rangle = \alpha\langle v,w\rangle + \beta\langle u,w\rangle$$

• סימטריות (הרמיטיות):

$$\forall v,u\in V:\ \langle v,u\rangle = \overline{\langle u,v\rangle}$$ כלומר צמוד מרוכב

• חיוביות בהחלט:

$$\forall v\in V:\ \langle v,v\rangle \ge 0,\quad v=0 \Rightarrow \langle v,v\rangle = 0$$

• הומוגניות:

$$\forall v\in V,\ \forall \alpha\in F:\ \|\alpha v\| = |\alpha|\cdot\|v\|$$

• אי שוויון המשולש:

$$\forall u,v\in V:\ \|v+u\| \le \|v\| + \|u\|$$

תכונה עבור כל שני נורמות A,B כלשהן מתקיים:

$$\forall v,\ 0<c_1\le c_2<\infty:$$

$$c_1\cdot\|v\|_A \le \|v\|_B \le c_2\|v\|_A$$

דוגמא:

$$\forall v:\ \|v\|_2 \le \|v_1\| \le \sqrt{n}\,\|v_A\|$$

מרחב מכפלה פנימית:

הגדרה: יהי V מעל F נגדיר פונקציה $\langle\cdot,\cdot\rangle: V\times V\to F$ אשר תקרא מכפלה פנימית אם היא מקיימת 3 תכונות:

• לינאריות • סימטריות • חיוביות בהחלט

ניצבות:

הגדרה: שני וקטורים במרחב המכפלה הפנימית, $u,v\in V$ "יחשבו" כניצבים אם מתקיים כי $\langle v,u\rangle = 0$. סימון: $v\perp u$

נורמה:

הגדרה: למרחב וקטורי נורמי V מעל שדה V. נגדיר פונקציה $\|\cdot\|: V\to\mathbb{R}$ שתקרא נורמה ותקיים את התכונות הבאות:

• חיוביות בהחלט • הומוגניות • אי שוויון המשולש

נורמה מושרית:

הגדרה: דרך אחת לבניית נורמה היא על ידי מכפלה פנימית, לפי:

$$\|v\| \triangleq \sqrt{\langle v,v\rangle}$$

נורמות חשובות שצריך לדעת:

$$x = [x_1, x_2, \dots, x_n]$$

כלומר x וקטור קורדינטות

$$l^1 = \|x\|_1 = \sum_{k=1}^{n} |x_k|$$

$$WL^2 = \|x\|_2 = \left(\sum_{k=1}^{n} \lambda_k |x_k|^2\right)^{\frac{1}{2}} \quad , \forall k, \lambda_k > 0$$

$$L^p = \|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p} \qquad p \geq 1$$

$$L^\infty = \|x\|_\infty = \max_{k=1,2,3,\dots,n} |x_k|$$

אי שוויונות קלאסיים:

• **אי שוויון קושי - שוורץ:**

$$|\langle v,u\rangle| \leq \|v\| \cdot \|u\|$$

• **שוויון המקבילית:** נכון לנורמה מושרית

$$\|v+u\|^2 + \|v-u\|^2 = 2\|v\|^2 + 2\|u\|^2$$

• **תכונת פיתגורס:** אם $v \perp u$ אזי:

$$\|v+u\|^2 = \|v\|^2 + \|u\|^2$$

מרחב מטרי:

הגדרה: הרעיון הוא להוסיף להגדרות המרחב הבסיסיות הגדרת שמקבילה מרחק בין שני וקטורים (זה: מטריקה) עם תכונות מסוימות:

$$d(v,u) = \|v-u\|$$

וכל זו תהי ה מטריקה מושרית מהגדרת נורמה.

הגדרה: מהינתן הנורמה הוקטורית $\|\cdot\|_A$ לוקטורים מממד n, נגדיר נורמה של מט' T באופן הבא ע"י:

$$|||T|||_A = \max_{v \,|\, \|v\|_A = 1} \|Tv\|_A = \max_{v \neq 0} \frac{\|Tv\|_A}{\|v\|_A}$$

תכונות של $|||T|||_A = \max \|Tv\|_A$

T לא חייבת להיות לכסינה.

• $|||T_1 + T_2|||_A \leq |||T_1|||_A + |||T_2|||_A$

• $|||\alpha T|||_A = |\alpha|\, |||T|||_A$

• $\forall v \in \mathbb{R}^n, T \in \mathbb{R}^{n \times n}: \|Tv\|_A \leq |||T|||_A \cdot \|v\|_A$

• $|||T_1 \cdot T_2||| \leq |||T_1|||_A \cdot |||T_2|||_A$

• $|||T \cdot T \cdot T \cdots T|||_A = |||T^k|||_A \leq |||T|||_A^k$

• $|||T|||_A \geq \rho(T)$

ערך העצמי הגדול ביותר של ים.

• $\|Tv\|_A > 0 \quad \forall v \neq 0$

$\|Tv\|_A = 0 \quad \forall v = 0$

$$! \quad |||T^k v||| \leq |||T^k|||_A \cdot \|v\|_A \leq |||T|||_A^k \cdot \|v\|_A$$

אם נמצא נורמה אופרטורית כלשהי למטריצה T שערכה קטן מ-1, אזי תתקבל התכנסות אסימפטוטית של האיטרציות

$$u_k = T_1^k u_0$$

<u>משפט:</u> עבור המערכת הדינמית $u_k = T^k u_0$ אם מתקיים כי $|||A|||<1$ אז המערכת מובטחת להתכנס אסימפטוטית

<u>נורמות אופרטוריות שנוחות לשימוש:</u>

- נורמת ה-1 המושרית (היא) סכום העמודה הגדול ביותר בערך מוחלט.

$$|||T|||_1 = \max_j \sum_{i=1}^{m} |T_{i,j}|$$

- נורמת האינסוף המושרית (היא) סכום השורה הגדול ביותר בערך מוחלט.

$$|||T|||_\infty = \max_i \sum_{j=1}^{n} |T_{i,j}|$$

- נורמת ה-2 המושרית הינה הערך הסינגולרי הגדול ביותר של המטריצה

$$|||T|||_2 = \sigma_1$$

$$|\lambda_1| \le \sigma_1$$

- נורמת הפרובניוס קשורה לנורמות המשרות על ידי הקשר הבא:

$$\|T\|_F = \sqrt{\sum_{k=1}^{n} \sigma_k^2} \ge \sigma_1 = |||T|||_2$$

- אם T מטריצה סימטרית אז...

$$|||T|||_2 = \sigma_1 = \dots = \rho\{T\}$$

! עבור שתי מטריצות סימטריות A,B מתקיים: $\rho\{AB\} \le \rho\{A\}\rho\{B\}$

<u>משפט גרשגורין:</u>

<u>משפט:</u> למטריצה T ריבועית נעשה. בגודל n על n, נגדיר את n הדיסקאות הבאות:

$$O_k = \left\{ z \in \mathbb{C} \;\middle|\; |z - t_{kk}| \le \sum_{j \ne k} |t_{kj}| \right\}$$

כל דיסקה נוצרת משורה אחת ב-T, כשמרכזה הוא איבר האלכסון וברדיוס הוא סכום יתר האיברים בערך מוחלט. אז כל הערכים העצמיים של T מוכלים באיחוד של דיסקאות אלו.

מטריצה T נקראת דומיננטית אלכסונית אם מתקיים כי כל איבר אלכסון גדול ממש (בערכו) מסכום יתר האיברים בשורתו: $\forall 1 \le k \le n,\ |t_{k,k}| > \sum_{k \ne j} |t_{kj}|$

כלומר מטריצה תהיה DD אם כל הדסקאות גרשגורין שלה לא מכילות את הראשית ומכאן שהמטריצה הפיכה (לא סינגולרית) כי אין לה ע"ע=0.

~~פתרון איטרטיבי של מערכת משוואות:~~

עבור $Ax=b$ כאשר A ריבועית והפיכה:

- בחר פתרון אתחול מסוים x_0 (למשל 0)
- עדכון פתרון זה באופן איטרטיבי על ידי הנוסחה: $x_{k+1}=Tx_k+c$

הקשר בין $\{A,b\}$ ל- $\{T,c\}$: בהנחה שהתהליך מתכנס עבור $k\to\infty$ מ"ס לפתרון $x^*=Tx^*+c$ וכדי שהפתרון גם יקיים $x^*=A^{-1}b$

יש: $(I-T)A^{-1}b=c$

חייב לדרוש את זה!

לדוגמא (נדיר):

$T=I-A$

$c=b$

איך נבחר $\{T,c\}$: נבנה מ- $Ax=b$ ונבצע מניפולציות על שני האגפים באופן שלא יפגע את המשוואה או ישנה אותה. לבסוף נסמן אינדקסים $k+1$, k לצד ה-x-ים שרצינו

אם נחסיר את המשוואות:

$$
\begin{aligned}
&x_{k+1}=Tx_k+c\\
-\ &x^*=Tx^*+c\\
\hline
&(x_{k+1}-x^*)=T\underbrace{(x_k-x^*)}_{e_k} \longrightarrow e_{k+1}=Te_k
\end{aligned}
$$

e זו השגיאה שלנו ונרצה שישאף ל-0. לכן אם T יציבה באלגוריתם יתכנס.

- ~~אלגוריתם jacobi ו gs:~~

נתונה מערכת משוואות $Ax=b$ נפרק $A=L+D+U$

משולשת תחתונה ממש (בלי אלכסון)

אלכסון של A

משולשת עליונה בלי האלכסון

jacobi: $x_{k+1}=D^{-1}[b-(L+U)x_k]$

gauss: $x_{k+1}=(D+L)^{-1}[b-Ux_k]$

משפט: בהינתן מערכת משוואות $Ax=b$ אם A הינה דומיננטית האלכסון אזי האלגוריתם jacobi מתכנס. אם A דומיננטית האלכסון או חיובית מוגדרת אזי gauss מתכנס.

אם A אינה PD אפשר להכפיל ב- A^t לקבל משוואות נורמליות

האיברים חיוביים באלכסון

נשים לב ש- T בכל מקרה:

jacobi: $T=D^{-1}(L+U)$

gauss: $(D+L)^{-1}U$

אלגוריתם SD:

$$x_{k+1} = x_k - \mu A^t (A x_k - b)$$

זמנו מתכנסים של מ"ע A^tA סם של
עד הכי קטן λ_n והכי גדול λ_1.

משפט: אלגוריתם ה-SD מבטיח להתכנס לפתרון בע"ת LS (ללא תלות בוקטור ההתחלה x_0)
אם $0<\mu<2/\lambda_1$

משפט: עבור הבחירה המיטבית: $\mu_{opt} = 2/(\lambda_1+\lambda_n)$ שואף הפתרון מהאלגוריתם SD
מקיימת:

$$\| x_k - x^* \|_2 \le \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^k \| x_0 - x^* \|_2 = \left(\frac{\kappa-1}{\kappa+1}\right)^k \| x_0 - x^* \|_2$$

$\kappa = \frac{\lambda_1}{\lambda_n}$ מספר מצב של מ"ע A^tA

MSD: בחירת μ שונה כל איטרציה. ראינו שהאלגוריתם מבצע פס צויר של 90°
מעלות בין k לשני צעדים עוקבים

$$e_k = A^t(Ax_k - b) \Longrightarrow f_k = A\cdot e_k \Longrightarrow \mu = \frac{e_k^t e_k}{f_k^t f_k} = \frac{\|e_k\|_2^2}{\|f_k\|_2^2}$$

$$\Longrightarrow x_{k+1} = x_k - \mu_k e_k$$

משמעות הבאות לא צריך לחשב ע"ע של A^tA!

1) שיטת המומנטום:

אנחנו מבצעים ב-K:

$$x_{temp} = x_k - \mu A^t (A x_k - b)$$

$$x_{k+1} = x_{temp} + \alpha(x_{temp} - x_{k-1}) = (1+\alpha)x_{temp} - \alpha x_{k-1} =$$

$$= (1+\alpha)[x_k - \mu A^t(Ax_k - b)] - \alpha x_{k-1}$$

$\alpha=0$ האלגוריתם SD המקורי

$\alpha>0$ יוצר את האלגוריתם המומנטום.

לא ניקח את x_k בחזור הבא היעדה עצמו כלמין

פה יש קו ישר בין x_{k-1} ל-x_{k+1} כב. שלא עשה זיגזג

אפשר לרשום את המטריצה הבאה שתלויה בקצב ההתכנסות:

$$\begin{bmatrix} x_{k+1} \\ x_k \end{bmatrix} = \left[\begin{array}{c:c} (1+\alpha)(I-\mu A^tA) & -\alpha I \\ I & 0 \end{array}\right] \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix} + \begin{bmatrix} (1+\alpha)\cdot\mu A^t b \\ 0 \end{bmatrix}$$

2) אלגוריתם SGD:

עבור: $\min_{\underline{x}} \frac{1}{2}\|A\underline{x}-\underline{b}\|_2^2 \xRightarrow{\text{משיטה הרגילה}} x_{k+1} = x_k - \mu A^t(Ax_k - b)$

$= x_k - \mu \sum_{t=1}^{m} a_t(a_t^t x_k - b_t)$

כתיבה של סכום וקטורים

m מספר שורות

(אלגוריתם הקצב

$$\begin{cases} x_1 = x_0 - \mu a_1(a_1^t x_0 - b_1) \\ x_2 = x_1 - \mu a_2(a_2^t x_1 - b_2) \\ \vdots \\ x_m = x_{m-1} - \mu a_m(a_m^t x_{m-1} - b_m) \end{cases}$$

לא בדיוק עדכון ב- sgd
הוא עדיין הליכה בכיוון
הגרדיאנט של משוואה
אחת במערכת $Ax=b$

אפשר לכתוב גם

$$x_{t+1} = x_t - \mu a_{t \bmod M}(a_{t \bmod M}^t \cdot x_t - b_{t \bmod M})$$

!בעצם עובדים
מקבוצות של m וקטורים