ע"ע ו-ו"ע: v ו"ע ל-ע"ע λ ב- A $\Leftrightarrow$ $Av = \lambda v$

רדיוס ספקטרלי: $\rho(A) = \max_{1\le i\le n} |\lambda_i|$

$\forall_K:\ \rho(A) \le ||A^K||^{\frac{1}{K}}$

ערך סינגולרי: $\sigma_i = \sqrt{\lambda_i}$ עבור λ_i ע"ע של A^TA

פירוק ספקטרלי: אם ל- A יש n ו"ע עם ע"ע שונים אז A לכסינה: $A = V\Lambda V^{-1}$ $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$
$V = [v_1|\dots|v_n]$

פירוק נורמלי: אם A נורמלית אז קיימת $U \in \mathbb{C}^{n\times n}$ אוניטרית: $A = U\Lambda U^*$, $U^*U = I$
$\lambda_i \in \mathbb{C}$; $U = [u_1|\dots|u_n]$, $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$

חיובית מוגדרת: $x^TAx > 0\ \ \forall x \in \mathbb{R}^n$
(SPD- מטריצה סימטרית חיובית מוגדרת $\lambda_i > 0$)

חיובית חצי מוגדרת: $x^TAx \ge 0\ \ \forall x \in \mathbb{R}^n$
(מתקיים לכל A^TA כאשר A מדרגה מלאה)

מטריצה הפיכה: קיים A^{-1} $\Leftrightarrow$ $\det(A) \ne 0$ $\Leftrightarrow$
A לא סינגולרית $\Leftrightarrow$ A היא full rank

טענה: AB סינגולרית $\Leftarrow$ $\det(A) = 0$ או $\det(B) = 0$

טענה: עבור A חיובית מוגדרת / הפיכה / לכסינה,
אם $Av_i = \lambda_i v_i$ אז $A^{-1}v_i = (1/\lambda_i)v_i$

טענה: עבור $A \in \mathbb{R}^{2\times 2}$ מהצורה $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ אם
$a > b\,; c > b\,; a > 0\,; c > 0$ אז A SPD

נורמלית: $A^TA = AA^T$

תלת אלכסונית: $a_{i,j} = 0\ if\ |i-j| > 1$

strictly diagonally dominant (SDD): מטריצה A
אשר מקיימת $|a_{i,i}| > \sum_j |a_{i,j}|$ או $|a_{j,j}| > \sum_i |a_{i,j}|$

דטרמיננטות:
* $\det(AB) = \det(A)\det(B)$
* אם L היא מטריצה משולשית אז $\det(L) = \prod_i l_{i,i}$
* עבור $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ נקבל $\det(A) = ad - bc$

חוקי מטריצות:
* $(AB)^{-1} = B^{-1}A^{-1}$ עבור A, B ריבועיות
* $(AB)^T = B^TA^T$
* $(A+B)^T = A^T + B^T$

סכום מטריצות ו-ע"ע:
* עבור מטריצות A, B בעלי אותו ו"ע:
$(A+B)v = Av + Bv = \lambda_A v + \lambda_B v = (\lambda_A + \lambda_B)v$

נגזרות: $x, v \in \mathbb{R}^n$ $A \in \mathbb{R}^{m\times n}$ $M \in \mathbb{R}^{n\times n}$
* גרדיאנט של $f(x_1, \dots, x_n)$ הוא $\nabla f = \left[\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right]^T$
* $f = v^Tx$ אז $\nabla f = v$
* $f = x^Tx$ אז $\nabla f = 2x$
* $f = v^TA^TAx$ אז $\nabla f = A^TAv$
* $f = x^TA^TAx$ אז $\nabla f = 2A^TAx$
* $f = x^TMx$ אז $\nabla f = (M + M^T)x$

פתרון מערכת משוואות: $Ax = b$

טענה: קיים פתרון יחיד אם A $full\ rank$, אחרת יש אינסוף פתרונות או שאין פתרון כלל.

pivoting , LU , Gaussian Elemination: קיים פירוק $PA = LU$ ומתקיים: $Ax = P^TLUx = b$,
$y = L^{-1}(Pb)$, $x = U^{-1}y$

הפיכת מטריצה: $Y = L^{-1}I$, $A^{-1} = U^{-1}Y$
* אם A הפיכה אז קיים לה פירוק יחיד
* אם A SPD אז קיים פתרון ללא Pivot
* $\det(A) = \det(LU) = \det(U)$

Cholesky למטריצות SPD: אם A היא SPD קיימת L משולשית תחתונה עם $l_{i,i} > 0$ כך ש-$A = LL^T$.
* $\det(A) = \det(LL^T) = \det(L)^2$

Thomas למערכת תלת אלכסונית: האלגוריתם מצליח כאשר המטריצה A היא SPD או SDD.

פירוק QR: עבור $A = QR$ נקבל $x = R^{-1}Q^Tb$.

נורמה וקטורית: $||\cdot||$ תקרא נורמה אם מתקיים:
* אי שליליות: $||\vec{x}|| \ge 0$ או $||\vec{x}|| = 0 \Leftrightarrow \vec{x} = 0$
* הומוגניות: $||\lambda\vec{x}|| = |\lambda|||\vec{x}||$:$\forall\vec{x} \in \mathbb{R}^n, \lambda \in \mathbb{R}$
* אי שוויון המשולש: $||\vec{x} + \vec{y}|| \le ||\vec{x}|| + ||\vec{y}||$

קבוצת נורמות Lp: $||\vec{v}||_p = (\sum_{i=1}^n |v_i|^p)^{\frac{1}{p}}$
למשל: $||\vec{v}||_1 = \sum_{i=1}^n |v_i|$, $||\vec{v}||_\infty = \max_i\{|v_i|\}$
$||\vec{v}||_2 = (\sum_{i=1}^n v_i^2)^{\frac{1}{2}} = \sqrt{v^Tv} = \sqrt{\langle v, v\rangle}$
נשים לב כי $||\vec{v}||_\infty \le ||\vec{v}||_2 \le ||\vec{v}||_1$

נורמת אנרגיה: לכל מטריצה M מסוג SPD תוגדר הנורמה $||\vec{v}||_M = \sqrt{<\vec{v}, M\vec{v}>}$

טענה: מטריצה אורתוגונלית משמרת נורמה וקטורית l_2 ומתקיים: $||Q\vec{v}||_2 = ||\vec{v}||_2$

טענה: כל נורמה $||\cdot||$ היא פונקציה רציפה (לא בהכרח גזירה).

טענה: לכל נורמות וקטוריות ϕ_1, ϕ_2 קיימים קבועים $0 < a \le b$ ומתקיים $a\phi_2(\vec{v}) \le \phi_1(\vec{v}) \le b\phi_2(\vec{v})$ (כלומר כל הנורמות שקולות, אם באחת נתכנס ל-0 אז גם בשניה).

מכפלה פנימית וקטורית:
$\langle\cdot,\cdot\rangle$ תקרא מכפלה פנימית לכל $u, v \in \mathbb{R}^n$ אם:
* סימטריות: $\langle u, v\rangle = \langle v, u\rangle$
* לינאריות: $\langle u, v_1 + v_2\rangle = \langle u, v_1\rangle + \langle u, v_2\rangle$ וגם $\langle \alpha u, \beta v\rangle = \alpha\beta\langle u, v\rangle$
* אי שליליות: $\langle u, u\rangle \ge 0$ או $\langle u, u\rangle = 0 \Leftrightarrow u = 0$

מכפלות מוכרות: $\langle u, v\rangle = \sum_i u_iv_i = u^Tv$, $\langle u, v\rangle_M = u^TMv$ למטריצה M מסוג SPD

נורמת מטריצה: (מקיימות את אותן תכונות של נורמות וקטוריות)

נורמות מטריצות אנאלוגיות לנורמות וקטוריות:
* Forbenius Norm - $||A||_F = \left(\sum_i\sum_j |a_{i,j}|^2\right)^{\frac{1}{2}}$
$||A||_F = \sqrt{trace(AA^T)}$
* $||A||_1 = \sum_{ij} |a_{i,j}|$
* $||A||_\infty = \max |a_{i,j}|$

נורמות מטריצות מושרות:
$||A||_{a,b} = \max_{\vec{x}\in\mathbb{R}^n} \frac{||A\vec{x}||_a}{||\vec{x}||_b}$ יסומן $||A||_a = ||A||_{a,a}$

דוגמאות: $||A||_1 = \max_j\left(\sum_{i=1}^m |a_{i,j}|\right)$
$||A||_\infty = \max_i\left(\sum_{j=1}^n |a_{i,j}|\right)$
$||A||_2 = \max_i \sqrt{\mu_i}$ כאשר μ_i ע"ע של A^TA

טענה: $||A\vec{x}||_a \le ||A||_{a,b}||\vec{x}||_b$

טענה: $||AB||_{b,a} \le ||A||_{c,a}||B||_{b,c}$

טענה: עבור A נורמלית מתקיים $||A||_2 = \rho(A)$

מספר מצב (Condition Number): מסומן $\kappa(A)$
מתקיים $\frac{||\delta x||}{||x||} \le ||A^{-1}||||A||\left(\frac{||\delta b||}{||b||} + \frac{||\delta A||}{||A||}\right)$ כאשר
$\kappa(A) = ||A^{-1}||||A|| \overset{2-norm}{\triangleq} \frac{\sigma_{\max}(A)}{\sigma_{min}(A)}$
ו-$\sigma_{\max}(A) = \max_i \sqrt{\lambda_i}$ עבור λ_i ע"ע של A^TA.

מטריצה עם מספר מצב גדול נקראת מטריצה "חולנית" ועלולה ליצור שגיאות גדולות.

Power method: למציאת ו"ע עם ע"ע גדול ביותר.
בהינתן ניחוש התחלתי $x^{(0)}$: $x^{(k+1)} = \frac{(Ax^{(k)})}{||Ax^{(k)}||}$
לכל A לכסינה עם ע"ע $|\lambda_1| > |\lambda_2| \ge \dots \ge |\lambda_n|$:
$\lim_{k\to\infty}\{x^{(k)}\} = v_1\ ;\ \lim_{k\to\infty} \frac{\left((x^{(k)})^TAx^{(k)}\right)}{(x^{(k)})^Tx^{(k)}} = \lambda_1$

הערה: עבור A^{-1} השיטה תחזיר את v_n, λ_n , ו"ע עם ע"ע קטן ביותר

Inverse Power method: מציאת ו"ע של ע"ע נתון.
בהינתן ניחוש לע"ע μ וניחוש התחלתי לו"ע $x^{(0)}$:
$x^{(k+1)} = (A - \mu I)^{-1}x^{(k)}\ ;\ x^{(k+1)} \leftarrow \frac{x^{(k+1)}}{||x^{(k+1)}||}$

פקטור ההתכנסות ל-I.P.M הינו $\frac{|\mu - \lambda_{closest\ to\ \mu}|}{|\mu - \lambda_{second\ closest\ to\ \mu}|}$

בעיית הריבועים הפחותים: $A \in \mathbb{R}^{m\times n}$ full rank, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$
אנו מחפשים וקטור x כך ש-Ax יהיה כמה שיותר קרוב ל-b.
* $r(x) = Ax - b$ וקטור השארית.
* $\hat{x} = \arg\min_{x\in\mathbb{R}^n} ||Ax - b||_2^2$ הוא הוקטור האופטימלי או בכתיב מטריציוני $\hat{x} = (A^TA)^{-1}A^Tb$ (כאשר A היא full rank)
$A^TAx = A^Tb$ הן המשוואות הנורמליות, אם A אינה full rank אז יש להן אינסוף פתרונות.
* עבור קבוצות נק' (x_i, y_i) נרצה למצוא ישר $ax + b$ העובר כמה שיותר קרוב לנק'.
נפתור זאת כך: ונקבל:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} m & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i y_i \\ \sum_i x_iy_i \end{bmatrix}$$

$$\arg\min_{a,b} \left\| \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|_2$$

* ניתן להוסיף משקולות לבעיה ע"י מטריצה W שהיא SPD ולקבל $\hat{x} = \arg\min_{x\in\mathbb{R}^n} ||Ax - b||_W^2$ או $\hat{x} = (A^TWA)^{-1}A^TWb$
* אם A סינגולרית ניתן להוסיף רגולריזציה לבעיה באמצעות $\lambda > 0$ ומטריצה C (בדר"כ $C = I$) ולקבל:
$\hat{x} = \arg\min_{x\in\mathbb{R}^n} ||Ax - b||_2^2 + \lambda||Cx||_2^2$
או $\hat{x} = (A^TA + \lambda C^TC)^{-1}A^Tb$

טענה: כל משוואה ריבועית ניתן לכתוב כך: $x^TAx + x^Tb$

מינימום ריבועים ע"י SVD: $\hat{x} = \arg\min_{x\in\mathbb{R}^n} ||U\Sigma V^Tx - b||_2^2$
כאשר A היא full rank אז $\hat{x} = (A^TA)^{-1}A^Tb = V(\Sigma)^{-1}U^Tb$
כאשר A לא full rank אז $\hat{x} = V(\Sigma)^\dagger U^Tb$
עבור $(\Sigma)^\dagger = \begin{cases} (\Sigma_{i,i})^{-1} & \Sigma_{i,i} \ne 0 \\ 0 & \Sigma_{i,i} = 0 \end{cases}$

תהליך Gram Schmidt: $A = [a_1|\cdots|a_n]$ ניצור מטריצה אורתוגונלית Q בהתבסס על עמודות A כך: $q_i = a_i - \sum_{j=1}^{(i-1)} \frac{\langle a_i, a_j\rangle}{\langle q_j, q_j\rangle} q_j$

פירוק QR: באמצעות גרהאם שמידט עם נירמול נפרק מטריצה $A = QR$ למטריצה Q אורתוגונלית ו-R משולשית עליונה.

Algorithm: Gram Schmidt QR
$\#\ A = [\mathbf{a}_1|\mathbf{a}_2|\dots|\mathbf{a}_n] \in \mathbb{R}^{m\times n}$
Initialize: $R = 0^{n\times n}$, $R_{1,1} = ||\mathbf{a}_1||_2$, $\mathbf{q}_1 = \frac{\mathbf{a}_1}{R_{1,1}}$
for $i = 2, \dots, n$ **do**
  $\mathbf{q}_i \leftarrow \mathbf{a}_i$
  for $j = 1 : i - 1$ **do**
    $R_{j,i} = \mathbf{q}_j^\top \mathbf{a}_i$
    $\mathbf{q}_i \leftarrow \mathbf{q}_i - R_{j,i}\mathbf{q}_j$
  end
  $R_{i,i} = ||\mathbf{q}_i||_2$
  $\mathbf{q}_i \leftarrow \frac{\mathbf{q}_i}{R_{i,i}}$
end

פירוק SVD: לכל $A \in \mathbb{R}^{m\times n}$, $p = \min(m, n)$, קיימות מטריצות אורתוגונליות $U \in \mathbb{R}^{m\times p}, V \in \mathbb{R}^{n\times p}$ ומטריצה אלכסונית $\Sigma \in \mathbb{R}^{p\times p} = diag(\sigma_1, \dots, \sigma_p)$ (σ_i הם ערכים סינגולריים של A ומתקיים $\sigma_1 \ge \dots \ge \sigma_p$) כך שמתקיים: $A = U\Sigma V^T$
- σ_i הם שורשי הע"ע של A^TA
- v_i הם הו"ע של A^TA
- u_i הם הו"ע של AA^T

קירוב מטריצה עם מטריצה מדרגה קטנה ופירוק SVD: נחפש B הממזערת את $||A - B||_F = ||\Sigma - U^TBV||_F$ כאשר $A = U\Sigma V^T$.
נבחר את $B = U\Sigma_BV^T$ כאשר ל-Σ_B נבחר את $rank(B)$ הערכים הסינגולריים הגדולים ביותר של A.

מטריצת Householder: מטריצת Householder היא מטריצה אורתוגונלית מהצורה $Q = I - 2ww^T$
לכל וקטור v מתקיים $Qv = ||v||_2e_1$
עבור $\beta = ||v||_2$ הוקטור w מוגדר כך: $w = \frac{v - \beta e_1}{||v - \beta e_1||_2}$

שיטות איטרטיביות לפתרון מערכת משוואות: $Ax=b$

* עבור ניחוש התחלתי $x^{(0)}$ השיטה ϕ תוגדר כך $x^{(k+1)}=\phi(x^{(k)})$
* **קצב ההתכנסות** $\lim_{k\to\infty}\frac{||x^{(k+1)}-x^*||}{||x^{(k)}-x^*||^p}=C\le\rho(I-M^{-1}A)$

p מגדיר את קצב ההתכנסות ו-C הוא ה-Convergence factor.

* ההתכנסות היא לינארית כאשר $p=1$ ו-$|C|<1$.
* **וקטור השגיאה:** $e^{(k)}=x^*-x^{(k)}$ (לא ניתן לחישוב כי x^* נעלם)
* **וקטור השארית:** $r^{(k)}=b-Ax^{(k)}=Ae^{(k)}$
* השארית אנכית לעמודות המטריצה, כלומר $A^T(b-Ax)=0$
* נרצה שיתקיים $\lim_{k\to\infty}\{e^{(k)}\}=\lim_{k\to\infty}\{r^{(k)}\}=0$
* תנאי עצירה $\frac{||x^{(k)}-x^{(k-1)}||}{||x^{(k)}||}<\epsilon$ או $\frac{||Ax^{(k)}-b||}{||b||}<\epsilon$
* **כתיב מטריציוני:** $A=M+N$ ולכן $Mx+Nx=b$ ונפתור $x^{(k+1)}=M^{-1}(b-Nx^{(k)})=x^{(k)}+M^{-1}(b-Ax^{(k)})$

שיטת ריצ'רדסון (Richardson): $M=cI$ עבור $c>0$, $x^{(k+1)}=x^{(k)}+\frac{1}{c}(b-Ax^{(k)})$

שיטת ג'ייקובי הממושקלת (Jacobi): $1\le\omega<0$, אם $\omega=1$ נקבל את שיטת ג'ייקובי הרגילה, $M_J=D=diag(A)$: $D_{i,i}=A_{i,i}$.
$A=L+D+U$ ונקבל $x^{(k+1)}=x^{(k)}+\omega D^{-1}(b-Ax^{(k)})$ או $x^{(k+1)}=(1-\omega)x^{(k)}+\omega D^{-1}(b-(L+U)x^{(k)})$ או
$x_i^{(k+1)}=(1-\omega)x_i^{(k)}+\omega\frac{1}{a_{i,i}}(b_i-\sum_{j\ne i}a_{i,j}x_j^{(k)})$

שיטת גאוס זיידל ממושקלת (Gauss-Seidel): $1\le\omega<0$, אם $\omega=1$ נקבל את שיטת גאוס זיידל הרגילה, $M=(L+D)^{-1}$.
$x^{(k+1)}=(\omega L+D)^{-1}\left((1-\omega)Dx^{(k)}+\omega(b-Ux^{(k)})\right)$ או
$x^{(k+1)}=x^{(k)}+\omega(\omega L+D)^{-1}(b-Ax^{(k)})$ או
$x_i^{(k+1)}=\frac{1}{a_{i,i}}(b_i-\sum_{j<i}a_{i,j}x_j^{(k+1)}-\sum_{j>i}a_{i,j}x_j^{(k)})$

התכנסות שיטות איטרטיביות:
* וקטור השגיאה הינו $e^{(k+1)}=(I-M^{-1}A)^ke^{(0)}$.
* אם $\rho(I-M^{-1}A)<1$ אז השיטה מתכנסת לכל $x^{(0)}$.

טענה: בגאוס זיידל וג'ייקובי, אם A הינה SDD אז השיטות מתכנסות.
טענה: אם A היא לא סינגולרית אז גאוס זיידל מתכנסת.
טענה: אם A היא חיובית מוגדרת אז גאוס זיידל מתכנסת.
טענה: גאוס זיידל שקולה למינימיזציה של $f(x)=\frac12||x-x^*||_A^2=\frac12x^TAx-x^Tb+\frac12(x^*)^Tb$.

שיטת Steepes Descent: אם A היא SPD אז $\nabla f(x)=Ax-b$ עבור $f(x)=\frac12||x^*-x||_A^2=\frac12x^TAx-x^Tb+\frac12(x^*)^Tb$
השיטה תוגדר כך:
$$x^{(k+1)}=x^{(k)}-\alpha\nabla f(x^{(k)})=x^{(k)}+\alpha(b-Ax^{(k)})$$
עבור $\alpha_{opt}=\frac{(r^{(k)})^TAe^{(k)}}{(r^{(k)})^TAr^{(k)}}=\frac{\langle r^{(k)},r^{(k)}\rangle}{\langle r^{(k)},Ar^{(k)}\rangle}$ שנבחר בכל איטרציה מחדש.

התכנסות לשיטת Steepes Descent:
עבור $f(x)$ הנ"ל $\rho(I-\alpha A)=\max\{|1-\alpha\lambda_{max}|,|1-\alpha\lambda_{min}|\}$
ו- $\alpha_{opt}=\frac{2}{\lambda_{max}+\lambda_{min}}$, פקטור ההתכנסות הוא $\rho_{min}=1-\frac{2}{\kappa(A)}$

שיטת Conjugate gradients (CG):
טענה: אם A היא SPD ו- full rank אז CG מתכנסת אחרי n איטרציות לכל היותר. עבור כל ניחוש התחלתי.

אופטימיזציה עם אילוצים: צורה כללית לאופטימיזציה עם אילוצים:
$$(*)=\min_{x\in\mathbb{R}^n} f(x)\quad subject\ to \begin{cases} c_j^{eq}(x)=0 & j=1,..,m_{eq}\\ c_l^{ieq}(x)\le 0 & l=1,...,m_{ieq}\end{cases}$$
Feasible set: קבוצת הפתרונות שמקיימים את האילוצים.
$\Omega=\{x|c_j^{eq}(x)=0,\ j=[1,m_{eq}]\ ;c_l^{ieq}(x)\le0,\ l=[1,m_{ieq}]\}$
ניתן להגדיר את הבעיה כך: $(*)=\min_{x\in\Omega}f(x)$
x^* הוא פתרון locally ל-$(*)$ אם $x^*\in\Omega$ וגם קיים $\mathcal{N}$ המקיימת $f(x)\ge f(x^*)\ ;\forall x\in\mathcal{N}\cap\Omega$

אופטימיזציה עם אילוצי שוויון:
$$(**)=\min_{x\in\mathbb{R}^n}f(x)\quad subject\ to\ c_j^{eq}(x)=0\quad j=1,..,m_{eq}$$
$$c^{eq}(x)=\overbrace{\begin{bmatrix}-c_1^{eq}(x)-\\\vdots\\-c_{m_{eq}}^{eq}(x)-\end{bmatrix}}^{\text{מטריצת אילוצי השוויון}}\qquad J^{eq}(x)=\overbrace{\begin{bmatrix}-\nabla c_1^{eq}(x)-\\\vdots\\-\nabla c_{m_{eq}}^{eq}(x)-\end{bmatrix}}^{\text{מטריצת הגרדיאנטים של אילוצי השוויון}}$$
וקטור כופלי לגראנז': $(\lambda^{eq})^T=[\lambda_1\cdots\lambda_{m_{eq}}]$
פונקציית לגרז'יאן: $\mathcal{L}(x,\lambda^{eq})=f(x)+(\lambda^{eq})^Tc^{eq}(x)$
נחפש x^* ו- λ^{eq*} כך שמתקיים: $\nabla\mathcal{L}_{\lambda^{eq}}(x^*,\lambda^{eq*})=0$ וגם $\nabla_x\mathcal{L}(x^*,\lambda^{eq*})=\nabla f(x^*)+(J^{eq})^T\lambda^{eq*}=0$
נקודה x^* עם כופלי לגראנז' מתאימים λ^{eq*} היא מינימום אם לכל $y\in\mathbb{R}^n\ ;J^{eq}y=0$ מתקיים $y^T\nabla_x^2\mathcal{L}(x^*,\lambda^{eq*})y\ge0$

אופטימיזציה עם אילוצי שוויון ואי שוויון: הצורה הכללית היא $(*)$.
אילוץ אי שוויון פעיל: קיים פתרון x^* כך שמתקיים $c_l^{ieq}(x^*)=0$
$\mathcal{A}(x)$ היא קבוצת האינדקסים של אילוצי אי שוויון פעילים
אילוץ אי שוויון לא פעיל: קיים פתרון x^* כך שמתקיים $c_l^{ieq}(x^*)<0$

$$c^{ieq}(x)=\overbrace{\begin{bmatrix}-c_1^{ieq}(x)-\\\vdots\\-c_{m_{ieq}}^{ieq}(x)-\end{bmatrix}}^{\text{מטריצת אילוצי האי שוויון}}\qquad J^{ieq}(x)=\overbrace{\begin{bmatrix}-\nabla c_1^{ieq}(x)-\\\vdots\\-\nabla c_{m_{ieq}}^{ieq}(x)-\end{bmatrix}}^{\text{מטריצת הגרדיאנטים של אילוצי האי שוויון}}$$
וקטור כופלי לגראנז': $(\lambda^{ieq})^T=[\lambda_1\cdots\lambda_{m_{ieq}}]$
פונקציית לגרז'יאן:
$\mathcal{L}(x,\lambda^{eq},\lambda^{ieq})=f(x)+(\lambda^{eq})^Tc^{eq}(x)+(\lambda^{ieq})^Tc^{ieq}(x)$
נחפש x^*, λ^{eq*}, λ^{ieq*} כך שמתקיים
$\nabla_x\mathcal{L}(x^*,\lambda^{eq*},\lambda^{ieq*})=\nabla f(x^*)+(J^{eq})^T\lambda^{eq*}=0$,
$\nabla\mathcal{L}_{\lambda^{eq}}(x^*,\lambda^{eq*},\lambda^{ieq*})=0$ ו- $\nabla\mathcal{L}_{\lambda^{ieq}}(x^*,\lambda^{eq*},\lambda^{ieq*})=0$
הערה: נחפש λ^{ieq*} שכל איבריו ≤0 ויש בו כמה שיותר אפסים.
נקודה x^* עם λ^{eq*} , λ^{ieq*} מתאימים היא מינימום אם לכל $y\in\mathbb{R}^n$ המקיים $J^{eq}y=0\ ;\ \forall l\in\mathcal{A}(x)\ \left(\nabla c_l^{ieq}(x^*)\right)^Ty=0$ מתקיים $y^T\nabla_x^2\mathcal{L}(x^*,\lambda^{eq*},\lambda^{ieq*})y\ge0$

אופטימיזציה ללא אילוצים: 1) נתונה פונקציה $f(x):\mathbb{R}^n\to\mathbb{R}^n$ ונרצה למצוא $x^*=argmin_{x\in\mathbb{R}^n}f(x)$
טור טיילור חד מימדי: $c\in[x,x+\epsilon]$
$$f(x+\epsilon)=f(x)+f'(x)\epsilon+\frac12f''(x)\epsilon^2+\overbrace{\frac{1}{3!}f'''(c)\epsilon^3}^{O(\epsilon^3)}$$
טור טיילור דו מימדי:
$$f(x_1+\epsilon_1,x_2+\epsilon_2)=f(x_1,x_2)+\frac{\partial f}{\partial x_1}\epsilon_1+\frac{\partial f}{\partial x_2}\epsilon_2+\frac12\frac{\partial^2f}{\partial x_1^2}\epsilon_1^2+\frac{\partial^2 f}{\partial x_1\partial x_2}\epsilon_1\epsilon_2+\frac12\frac{\partial^2f}{\partial x_2^2}\epsilon_2^2$$
עבור $\epsilon=[\epsilon_1,\epsilon_2]^T$, $\nabla f=\left[\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2}\right]^T$, $H=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\partial x_2}\\ \frac{\partial^2 f}{\partial x_1\partial x_2} & \frac{\partial^2 f}{\partial x_2^2}\end{bmatrix}$
(ב-n מימדים $H_{i,j}=\frac{\partial^2 f}{\partial x_i\partial x_j}$) נקבל:
$$f(x+\epsilon)=f(x)+\langle\nabla f,\epsilon\rangle+\frac12\langle\epsilon,H\epsilon\rangle+O\left(||\epsilon||^3\right)$$
2) אם $f(x):\mathbb{R}^n\to\mathbb{R}^m$ היא וקטור של פונקציות (ניתן לחשוב על מטריצה בגודל $n\times m$) , $x\in\mathbb{R}^n$, $f(x)\in\mathbb{R}^m$
* $f_i(x+\epsilon)\approx f_i(x)+\langle\nabla f_i(x),\epsilon\rangle$
* $\delta f=f(x+\epsilon)-f_i(x)\approx J\epsilon$ (עבור $\epsilon\to\infty$ נקבל שוויון)
* $J\in\mathbb{R}^{m\times n}$ היא מטריצת יעקוביאן, $J_{i,j}=\frac{\partial f_i}{\partial x_j}$
* אם $f=Ax$ עבור $A\in\mathbb{R}^{m\times n}$ אז $\delta f=f(x+\epsilon)-f_i(x)=A(x+\epsilon)-Ax=A\epsilon$, כלומר $J=A$
* אם $\phi:\mathbb{R}\to\mathbb{R}$ ו- $f_i(x)=\phi(x_i)$ (כלומר אם f מיוצגת באמצעות מטריצה אז רק האיבר ה-i שונה מ-0) אז:

$\delta f=\phi(x+\epsilon)-\phi_i(x)\approx diag(\phi'(x))\overbrace{\epsilon}^{\delta x}=diag(\phi'(x))\delta x$
ו- $J_{i,i}=\phi'(x_i)$, $J=diag(\phi'(x))$

פתרון בעיית אופטימיזציה ללא אילוצים:
* נחפש x^* המקיימת $\nabla f(x^*)=0$ והיא תוכל להיות נקודת מינימום.
* x^* מינימום locally: $\exists r>0:f(x)\ge f(x^*)\ ;\forall x:||x-x^*||<r$
* x^* מינימום globally: $f(x)\ge f(x^*)\ ;\forall x\in\mathbb{R}^n$
* אם H חיובית מוגדרת, x^* מינימום locally אם $\nabla f(x^*)=0$

שיטות איטרטיביות לבעיות אופטימיזציה ללא אילוצים:
שיטת steepest descend: נבחר בכל איטרציה $x^{(k+1)}=x^{(k)}-\alpha M\nabla f(x^{(k)})$ (ניתן לבחור $M=I$)
d הוא <u>כיוון ירידה</u> כאשר מתקיים: $\langle\nabla f,d^{(k)}\rangle<0$.
כיוון הירידה הוא מינוס הגרדיאנט: $-\nabla f(x^{(k)})$, שהוא כיוון הירידה המירבי של הפונקציה f.
כל כיוון $d=-M\nabla f(x^{(k)})$ כך ש- M חיובית מוגדרת הוא כיוון ירידה.
שיטת Newton: אם H הפיכה אז בכל איטרציה נבחר $x^{(k+1)}=x^{(k)}+d$ עבור $d=-H^{-1}\nabla f(x^{(k)})$
שיטת Quasi Neqton: נחליף את H ב-$M(x^{(k)})$ ואז בכל איטרציה נבחר $x^{(k+1)}=x^{(k)}+d$ עבור $d=-\left(M(x^{(k)})\right)^{-1}\nabla f(x^{(k)})$
(למשל M שווה לאלכסון של H וכך היא הפיכה)
שיטות Line Search: בכל איטרציה נבחר $x^{(k+1)}=x^{(k)}+\alpha^{(k)}d^{(k)}$ (כאן ההתייחסות היא לבחירת α, בהנחה שאת d אנו יודעים לבחור).
נרצה לבחור $\alpha^{(k)}$ מינימלית, כלומר עבור $\phi(\alpha)=f(x^{(k)}+\alpha d^{(k)})$ נרצה למזער $\alpha^{(k)}=argmin_\alpha\phi(\alpha)$
Backtracking line search using Armijo condition: נבחר $\alpha^{(0)}$ די גדול, $0<\beta<1$ (בדר"כ $\beta=\frac12$) ו- $0<c<1$ (בדר"כ $c=10^{-4}$).
בכל איטרציה נבחן $\alpha_j^{(k)}=\beta^j\alpha^{(0)}$ עבור ... $j=0,1,2,$ עד שמתקיים:
$$f(x^{(k)}+\alpha_j^{(k)}d^{(k)})\le f(x^{(k)})+c\alpha_j^{(k)}\langle\nabla f,d^{(k)}\rangle$$
שיטות Coordinate descent: בכל איטרציה נעדכן את הקואורדינטה ה-i של x. נניח כי x_j קבועים לכל $j\ne i$ ונבחר x_i המקיים $\frac{\partial f}{\partial x_i}=0$.
ניתן לעבור על ה-i סדרתית או אקראית (העיקר לעדכן את כולם).

Penalty and Barrier methods:
את בעיות האופטימיזציה עם אילוצי אי שוויון ושוויון שראינו ב-$(*)$ נכתוב כך:
$$(***)=\min_{x\in\mathbb{R}^n}f(x)+\mu\left(\sum_{j=1}^{m_{eq}}\rho_j\left(c_j^{eq}(x)\right)+\sum_{l=1}^{m_{ieq}}\rho_l\left(\max\{0,c_l^{ieq}(x)\}\right)\right)$$
עבור $\mu>0$ ו- $\{\rho_{j(x)}\},\{\rho_l(x)\}$ פונקציות סקלריות החסומות מלמטה ב-0 (לרוב $\rho(x)=|x|,\rho(x)=x^2$).
עבור $\mu\to\infty$ הפתרון יהיה זהה לפתרון של $(*)$.
בפועל נפתור בצורה איטרטיבית עם ניחוש התחלתי $x^{(0)}$ את $(***)$ עם $\mu_0<\mu_1<\cdots<\infty$ ובכל איטרציה הקלט יהיה $x^{(k)}$ מהאיטרציה הקודמת.

The projected Steepest Descent method: נדאג שבכל איטרציה נהיה בתוך התחום שמקיים את האילוצים (feasible). עבור ניחוש התחלתי $x^{(0)}$ בכל איטרציה נבצע:
$$x^{(k+1)}=\prod_\Omega\left(x^{(k)}-\alpha\nabla f(x^{(k)})\right)$$
עבור פונקציית הטלה:
$$\prod_\Omega(y)=argmin_{x\in\Omega}||x-y||$$

משפטים שאני המצאתי:
משפט: עבור $x=\sum_{i=1}^n\alpha_iv_i$ כאשר (v_i,λ_i) הם ו"ע וע"ע מתקיים $\max\left(\frac{\sum_{i=1}^n\alpha_i^2\lambda_i}{\sum_{i=1}^n\alpha_i^2}\right)=\lambda_{max}$
משפט: עבור A SPD מתקיים $\max\frac{x^TAx}{x^Tx}=\lambda_{max}$ וגם $\max\frac{x^TA^{-1}x}{x^Tx}=\lambda_{min}$
משפט: בשיטות איטרטיביות לפתרון מערכת משוואות α_{opt} הוא זה שממזער את $\rho(I-M^{-1}A)$
משפט: אם $A\in\mathbb{R}^{m\times n}$ מדרגה מלאה אז $rank(A)=\min(m,n)$
משפט: אם A SPD אז כל המינורים שלה הם גם SPD. בנוסף למינורים יש דטרמיננטות שונות מ-0.
משפט: אם B חצי מוגדרת ו-M חיובית מוגדרת אז $(M+B)$ חיובית מוגדרת.
בפרט אם A full rank אז $(I+A^TA)$ חיובית מוגדרת.
משפט: אם מכפלת מטריצות מדרגה מלאה נותנת מטריצה מסדר $(k\times k)$ כאשר k הוא המינימלי מבין כל הדרגות אז המטריצה מדרגה מלאה.

קמירות:
1) $S\in\mathbb{R}^n$ קבוצה קמורה אם לכל $x,y\in S,\alpha\in[0,1]$ מתקיים $\alpha x+(1-\alpha)y\in S$
2) f פונקציה קמורה אם ה-domain שלה הוא קבוצה קמורה ולכל $x,y,\alpha\in[0,1]$ מתקיים $f(\alpha x+(1-\alpha)y)\le\alpha f(x)+(1-\alpha)f(y)$
3) f פונקציה קמורה אם היא גזירה ולכל x_1,x_2 מתקיים $f(x_1)\ge f(x_2)+\langle\nabla f(x_2),x_1-x_2\rangle$
4) f פונקציה קמורה אם היא גזירה פעמיים ומתקיים $H=\nabla^2f(x)\ge0$, כלומר H מטריצה חיובית חצי מוגדרת
5) אם f קמורה אז כל מינימום locally x^* או כל נקודה המקיימת $\nabla f(x^*)=0$ היא מינימום globally