הגדרה: קבוצה קמורה:

קבוצה S תיקרא קמורה אם לכל $x,y \in S$, $0 \leq \alpha \leq 1$

מתקיים: $\alpha x + (1-\alpha)y \in S$

הגדרה: פונקציה קמורה:

פונקציה $f(x)$ תיקרא קמורה על קבוצה קמורה S אם לכל $x,y \in S$

ועבור $0 \leq \alpha \leq 1$ מתקיים:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

משפט: על פונקציה קמורה:

A continuosly differentiable function $f(x)$ is convex over an interval I if and only if it is above its linear approximation.

$$f(x) \geq f(y) + (x-y)f'(y), \quad \forall x,y \in S.$$

הוכחה:

$\Leftarrow$ נניח ש- f קמורה:

$$f(\alpha x + (1-\alpha)y) = f(\alpha x + y - \alpha y) = f(y + \alpha(x-y)) \leq$$

$$\leq \alpha f(x) + (1-\alpha) f(y)$$

עפ"י פיסוי הראשון.

$$f(y+\alpha(x-y)) \leq \alpha f(x) + (1-\alpha) f(y) \quad / : \alpha$$

$$f(x) \geq \frac{f(y+\alpha(x-y))}{\alpha} - \frac{(1-\alpha) f(y)}{\alpha}$$

$$f(x) \geq \frac{f(y+\alpha(x-y)) - (1-\alpha) f(y)}{\alpha(x-y)} (x-y)$$

$$f(x) \geq \left[\frac{f(y+\alpha(x-y)) - f(y)}{\alpha(x-y)} + \frac{\alpha f(y)}{\alpha(x-y)}\right](x-y)$$

$$f(x) \geq f'(y)(x-y) + f(y) \quad \blacksquare$$

$\Rightarrow$ נניח שמתקיים: $f(x) \ge f(y) + (x-y)f'(y)$

אזי נבחר:

$z = \lambda x + (1-\lambda)y$ (נקודה בין x ל-y).

Apply the condition twice $\Rightarrow$ $f(x) \ge f(z) + (x-z)f'(z) \quad / \cdot (\lambda)$

$$f(y) \ge f(z) + (y-z)f'(z) \quad / \cdot (1-\lambda)$$

→ Функция выше своего линеарного кирува.

$$\lambda f(x) + (1-\lambda)f(y) \ge f'(z)$$

■

משפט: גם על פונקציה קמורה

פונקצייה $f(x)$ גזירה ברציפות פעמיים הינה קמורה על קבוצה קמורה אמ"מ $\forall x \in S \; f''(x) \ge 0$.

הוכחה:

$\Rightarrow$ נניח כי $f''(x) \ge 0$:

Using a Taylor expansion:

$$f(y) = f(x) + (y-x)f'(x) + \frac{(y-x)^2}{2} f''(z)$$

Due to $f'' > 0$ we get

один эвар убрали.

$$f(y) \ge f(x) + (y-x)f'(x)$$

$$f(y) - f(x) \ge (y-x)f'(x)$$

(лидер -'y'). $$f(y) \ge f(x) + (y-x)f'(x)$$

ועפ"י משפט הקודם זה מוכיח שזה קמור. ■

$\Leftarrow$ נניח ש- f קמורה:

1. נניח בשלילה כי קיים S ו-x כך שמתקיים $f''(x) \le S$.
2. ניתן לבחור y קרוב מספיק כך ש: $f''(z) < 0$

$$f(y) - f(x) = (y-x)f'(x) + \frac{1}{2}(y-x)^2 f''(z) < (y-x)f'(x)$$

סתירה למשפט הקודם ■

ע"י שימוש המשפטים שהראנו, ובמיוחד משפט האחרון, ניתן להוכיח בקלות כי הפונקציות הבאות הן גם קמורות:

(1) $f(x) = e^x$ (3) $f(x) = -\log x$

(2) $f(x) = x^\alpha$ (4) $f(x) = x\log x$

משפט: מינימום לוקאלי ← מינימום גלובאלי

תהיה $f(x)$ פונקציה קמורה על S (שהיא קמורה) אזי כל מינימום לוקאלי הוא גלובאלי.

הוכחה:

(1) נניח ש-x מינימום לוקאלי אך לא גלובאלי.

(2) לכן קיים y כך שמתקיים: $f(y) < f(x)$.

(3) f קמורה ולכן:

$$f(\alpha x + (1-\alpha)y) \underset{f\text{ קמורה}}{\leq} \alpha f(x) + (1-\alpha)f(y) <$$

$$\underset{f(y)<f(x)}{<} \alpha f(x) + (1-\alpha)f(x) = f(x)$$

$f(x)$

$f(\alpha x+(1-\alpha)y)$

$f(y)$

← קיבלנו סתירה לכך ש-x מינימום לוקאלי.

чтобы дойти с $f(x)$ к средней точке, нужно спуститься когда она ∪. Стира. ■

Transpose Properties:

① $(A^T)^T = A$

② $(A+B)^T = A^T + B^T$

③ $(AB)^T = B^T A^T$

④ $A^T B = B^T A$

① אם $AB=BA$ הן 2 מטריצות מתחלפות.

אם $A, B > 0$ אזי $Tr(AB) \geq 0$.

② אם $B > A \geq 0$.

אם הן מתחלפות אזי כן מתקיים: $B^{-1} < A^{-1}$.

$$\|X\|^2 = X^T X$$

$$\|AX\|^2 = X^T A^T A X$$

Chain Rule:

$$\nabla(X^T Q X) = (Q + Q^T)X = 2QX.$$

(*) $g(x) = f(Ax+b)$

(*) $g'(x) = A^T \nabla f(Ax+b)$

(*) $g''(x) = A^T \nabla^2 f(Ax+b) A$

Jensens' inequality:

F convex:

- $F(\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)w) \leq$
$\leq \lambda F(x,y) + (1-\lambda)F(z,w).$

- $F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y).$

(*) $\nabla^2 f(x) > 0$ is equivalent to $b^T f(x) b \geq 0$.

(*) אם A סימטרית ו-B סימטרית, אזי
AB לא סימטרית.

(*) Trace – סכום איברים על האלכסון
– סכום ערכים עצמיים

(*) הרכבה של 2 פונקציות קמורות גם קמורה.

(*) Cauchy Shwartz: $\sum a_i^2 \sum b_i^2 - (\sum a_i b_i)^2 \geq 0$

$\hookrightarrow + \quad \underline{a^T b \leq \|a\| \cdot \|b\|}$

הגדרה: קבוצה אפינית:

אפיני = ליניארי + קבוע. $\alpha x + \beta$.

קבוצה C תיקרא אפינית, אם הישר שעובר דרך כל שתי נקודות $x_1, x_2 \in C$ שייך ל-C:

$$x_1, x_2 \in C,\ \alpha \in \mathbb{R} \Rightarrow \alpha x_1 + (1-\alpha) x_2 \in C$$

דוגמא לקבוצה אפינית: פתרון של משוואות ליניאריות

$$C = \{x : \underline{A}\underline{x} = \underline{b}\}$$

קבוצה C, היא קבוצה אפינית.

הוכחה:

(1) $\underline{A} x_1 = b$

(2) $\underline{A} x_2 = b$

(3) ניקח את 2 הנקודות x_1 וx_2 ששייכים לקבוצה C.

נרכיב את שוויון: $\alpha x_1 + (1-\alpha) x_2,\ \alpha \in \mathbb{R}$

נציב את הפסוק לתוך A. אם נקבל שהוא שווה ל-b, אזי C אפינית.

$$\Rightarrow A(\alpha x_1 + (1-\alpha) x_2) = \alpha A x_1 + (1-\alpha) A x_2 = \alpha b + (1-\alpha) b = b$$

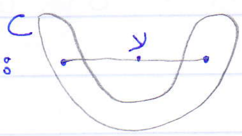
לכן, קבוצה C, היא קבוצה אפינית ■

הגדרה: קבוצה קמורה:

קבוצה C תיקרא קמורה אם לכל $x_1, x_2 \in C$ ו-α מתקיימות: $0 \le \alpha \le 1$

מתקיים: $\alpha x_1 + (1-\alpha) x_2 \in C$

Да:


Нет:


точка y вышла из группы.

← כל קבוצה אפינית היא גם קבוצה קמורה.
זה נובע מ-2 ההגדרות באופן די אוטומטי.

← מה ההבדל בין אפיני לקמור?
הם שניהם קומבינציות לינאריות שסכום של המקדמים שלהם מסתכם ל-1. אבל המקדמים בקבוצות קמורות הם בנוסף אי שליליים

דוגמאות לקבוצות קמורות:

1) HyperPlane: $\{x: a^Tx = b\}$, $a \neq 0$ is affine & convex.

נוכיח:

① ניקח 2 נקודות x ו y באופן הבא:

$$a^Tx = b$$
$$a^Ty = b$$

② ניצור מ-2 הנקודות קומבינציה לינארית:

$$\lambda x + (1-\lambda) y$$

③ נציב את הקומבינציה לתוך המשוואה:

$$a^T(\lambda x + (1-\lambda) y) = \lambda a^Tx + (1-\lambda) a^Ty =$$
$$= \lambda b + (1-\lambda) b = b.$$

זה תואם להגדרה של פונקציה קמורה, ולכן סיימנו ■

2) Halfspace: $\{x: a^Tx \leq b\}$, $a \neq 0$ is convex but not affine

$$a^Tx \leq b$$
$$a^Ty \leq b$$
$$\rightarrow a^T(\lambda x + (1-\lambda) y) = \lambda a^Tx + (1-\lambda) a^Ty \leq$$
$$\leq \lambda b + (1-\lambda) b = b. \ ■$$

$\lambda \leftarrow$ מקדם אי שלילי ולכן בטוח לא ישנה את הסימן.

↓

- **Euclidean ball** $\{x : \|x - c\|_2 \le r\}$, $r > 0$ is convex.

נוכיח:

① ניקח 2 נקודות x_1 ו x_2, נייצר בינהם קומבינציה לינארית.

נציב את הקומבינציה בתוך הנורמה ונקבל:

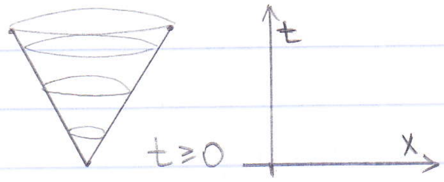
$$\|\alpha x_1 + (1-\alpha)x_2 - c\| \underset{\text{נוסיף ונוריד } \alpha c}{=} \|\alpha x_1 + x_2 - \alpha x_2 - c + \alpha c - \alpha c\| =$$

$$= \|\alpha(x_1 - c) + (1-\alpha)(x_2 - c)\|_2 \underset{\text{אי שוויון המשולש}}{\le} \alpha\|x_1 - c\| + (1-\alpha)\|x_2 - c\|_2 \le$$

$$\le \alpha r + (1-\alpha) r = r \quad \blacksquare$$

- **Second order** (ice-cream, Lorentz, quadratic) **cone**:

$\{(x,t) : \|x\|_2 < t\}$ is convex



Similar way to write: $\|x\|_2 - t \le 0$

① ניקח 2 נקודות x_1 ו x_2.

ונייצר בינהם קומבינציה לינארית. $\|x_1\|_2 \le t_1, \|x_2\|_2 \le t_2$

② ניקח 2 נקודות t_2 ו-t_1.

ונייצר בינהם קומבינציה לינארית.

$$\|\alpha x_1 + (1-\alpha)x_2\|_2 - \alpha t_1 - (1-\alpha)t_2 \underset{\text{אי שוויון משולש}}{\le}$$

$$\le \alpha\|x_1\|_2 + (1-\alpha)\|x_2\|_2 - \alpha t_1 - (1-\alpha)t_2 =$$

$$= \alpha\underbrace{(\|x_1\|_2 - t_1)}_{\le 0} + (1-\alpha)\underbrace{(\|x_2\|_2 - t_2)}_{\le 0} \le 0 \quad \blacksquare$$

• <u>Positive semidefinite cone is convex.</u>

$$S_+^n = \{ X \in S^n : X \succeq 0 \} =$$

$$= \{ X \in S^n : z^T X z \geq 0 , \forall z \}.$$

<u>נוכיח:</u>

① נבחר X_1 ו X_2 כך שמתקיים:

$$z^T X_1 z \geq 0$$
$$z^T X_2 z \geq 0$$

② נרכיב קומבינציה לינארית:

$$\alpha X_1 + (1-\alpha) X_2$$

③ נציב את הקומבינציה ונקבל:

$$z^T(\alpha X_1 + (1-\alpha) X_2) z = \alpha \underbrace{z^T X_1 z}_{\geq 0} + (1-\alpha) \underbrace{z^T X_2 z}_{\geq 0} \geq 0$$

■

פעולות ששומרות על קמירות

① חיתוך:

אם S_1 ו S_2 הן 2 קבוצות קמורות, אזי גם החיתוך $S_1 \cap S_2$ קמור.

הוכחה:

① $x \in S_1 \cap S_2 \rightarrow x \in S_1$ and $x \in S_2$

② $y \in S_1 \cap S_2 \rightarrow y \in S_1$ and $y \in S_2$

③ נייצר קומבינציה ליניארית בין x ל y:

$$\alpha x + (1-\alpha) y$$

④ $\alpha x + (1-\alpha) y \in S_1 \rightarrow$

$\alpha x + (1-\alpha) y \in S_1$ and

$\alpha x + (1-\alpha) y \in S_2$ ■

② חיתוך אינסופי:

אם S_α הוא קמור לכל $\alpha \in A$ אזי מתקיים $\bigcap_{\alpha \in A} S_\alpha$ is convex.

→ Важно обратить внимание, что A не обязана быть convex.

Example: the positive semidefinite cone can be expressed as the intersection of an infinite number of halfspaces and is therefore convex:

$$\{X \in S^n : X \succeq 0\} = \bigcap_{z \neq 0} \{X \in S^n : z^T X z \geq 0\}. \blacksquare$$

③ כפל:

אם S_1 ו S_2 הן 2 קבוצות קמורות אזי גם $S_1\times S_2$ קמור.

$$S_1\times S_2=\left\{x_1,x_2 : \begin{matrix} x_1\in S_1 \\ x_2\in S_2\end{matrix}\right\}$$

הוכחה:

① $x\in S_1\times S_2$, $x=(x_1,x_2)$

② $y\in S_1\times S_2$ $y=(y_1,y_2)$

③ $\alpha\in(0,1)$

④ $x_i y_i\in S_i$ לכל $i=1,2$.

⑤ S_1 ו S_2 קמורות ולכן:

$$\alpha x_1+(1-\alpha)y_1\in S_1$$
$$\alpha x_2+(1-\alpha)y_2\in S_2$$

⑥ ולכן:

$$\left(\alpha x_1+(1-\alpha)y_1,\ \alpha x_2+(1-\alpha)y_2\right)\in S_1\times S_2$$

או באותו אופן:

$$\alpha(x_1,y_1)+(1-\alpha)(y_1,y_2)\in S_1\times S_2$$

■

④ הרכבה עם פונקציה אפינית:

אם S קמורה, אזי התמונה שלה תחת הפונקציה האפינית $f(x)$ קמורה.

כאשר $f(x)$ מוגדרת להיות: $f(x)=Ax+b$.

והתמונה מוגדרת להיות: $f(S)=\{Ax+b : x\in S\}$

הוכחה:

① קיים $x_1\in S$ כך ש: $f_1=Ax_1+b$, $f_1\in f(S)$

② קיים $x_2\in S$, כך ש: $f_2=Ax_2+b$, $f_2\in f(S)$

אנחנו צריכים להראות שמתקיים:

$$f' = \lambda f_1 + (1-\lambda) f_2 \in f(S).$$

אם נבחר בעזרת X-ים נקבל:

$$x' = \lambda x_1 + (1-\lambda) x_2$$

ונראה שמתקיים:

$$Ax' = f' - b$$

$$\Rightarrow A(\lambda x_1 + (1-\lambda)x_2) = \lambda A x_1 + (1-\lambda) x_2 =$$

$$= \lambda(f_1 - b) + (1-\lambda)(f_2 - b) =$$

$$= \lambda f_1 + f_2 - \lambda f_2 - b =$$

$$= \lambda f_1 + f_2(1-\lambda) - b = f' - b. \blacksquare$$

5 קמורה הופכית עם הרכבה עם פונקציה אפינית:

אם S קמורה, אזי התמונה ההופכית שלה תחת פונקציה האפינית $f(x)$ הינה קמורה.

- הפונקציה $f(x)$ מוגדרת להיות: $f(x) = Ax + b$.
- והתמונה ההופכית: $f^{-1}(S) = \{x : Ax + b \in S\}$

הוכחה:

1. $x_1 \in f^{-1}(S) : Ax_1 + b \in S$
2. $x_2 \in f^{-1}(S) : Ax_2 + b \in S$
3. נבנה קומבינציה לינארית של λ- ה x_1 ו x_2:
   $\lambda x_1 + (1-\lambda) x_2$
4. נרצה להראות שקומבינציה זו מקיימת $Ax + b \in S$ ולכן $f \in f^{-1}(S)$

$$Ax + b = \lambda(Ax_1 + b) + (1-\lambda)(Ax_2 + b) \overset{\text{convex}}{\in} S \blacksquare$$

נוסיף ונוריד λb.

6 הכפלה בקבוע

אם קבוצה S היא קמורה, אזי עבור קבוע β מתקיים:

$$\beta S = \{ \beta x : x \in S \}$$

אזי βS גם קמורה.

הוכחה:

← אם $\beta = 0$, אזי $\beta S = \{0\}$ שאמור שקמור.

← אזי נוכיח עבור $\beta \neq 0$:

נגדיר x ו y כך ש: $x, y \in$.

אזי מתקיים:

$$\frac{x}{\beta} \in S \,,\ \frac{y}{\beta} \in S.$$

בגלל ש-S קמורה, מתקיים:

$$\lambda \cdot \frac{x}{\beta} + (1-\lambda)\frac{y}{\beta} \in S$$

נכפיל ב- β:

$$\lambda x + (1-\lambda) y \in \beta S$$

לכן βS הוא קמור ■

7 חיבור קבוע

אם קבוצה S קמורה, אזי עבור β קבוע מתקיים:

$$S + \beta = \{ x + \beta : x \in S \}$$

8 חיבור קבוצות קמורות:

S_1 ו S_2 קמורות אזי גם:

$$S_1 + S_2 = \{ x_1 + x_2 : x_1 \in S \,,\ x_2 \in S \}$$

הוכחה:

1. $x, y \in S_1 + S_2$
2. $x = x_1 + x_2$, $x_1 \in S_1$ $x_2 \in S_2$.
3. $y = y_1 + y_2$, $y_1 \in S_1$ $y_2 \in S_2$
4. S_1 קמור ו S_2 קמור:

$$\lambda x_1 + (1-\lambda) y_1 \in S_1$$

$$\lambda x_2 + (1-\lambda) y_2 \in S_2$$

↓

$$\Rightarrow \lambda x_1 + (1-\lambda) y_1 + \lambda x_2 + (1-\lambda) y_2 =$$

$$= \lambda x_1 + y_1 - \lambda y_1 + \lambda x_2 + y_2 - \lambda y_2 =$$

$$= \lambda(x_1 + x_2) + (1-\lambda)(y_1 + y_2) \in S_1 + S_2. \blacksquare$$

דוגמא: אליפסואיד היא קבוצה קמורה:

$$\text{Elipsoid}: \{x : (\underline{x} - \underline{c})^T P^{-1} (\underline{x} - \underline{c}) \leq 1\}$$

$\downarrow$

מתיחה | סיבוב כדי להפוך לאליפסואיד

$$P = U^{-T} U^{-1}$$
$$P^{-1} = U^T U$$

$$\{x : \underbrace{\boxed{(x-c)^T U^T}}_{A^T} \underbrace{U(x-c)}_{A} \leq 1\} = \qquad \underline{A^T B} = B^T A$$

$$= \{x : \|Ux - Uc\|_2^2 \leq 1\} =$$

$$= \{x : Ux - Uc \in \{\underbrace{z : \|z\| - 1}_{\text{כדור}}\}\}$$

ולכן קמור $\blacksquare$

Формула:

① $\|Ax\|_2^2 = x^T A^T A x$

② $(A+B)^2 = A^2 + AB + BA + B^2$

משפט:

קבוצה קמורה שנחתכת עם קו

קבוצה הינה קמורה אמ"מ חיתוך של קבוצה עם כל ישר הינו קמור.

הוכחה:

$\Leftarrow$ נניח כי הקבוצה קמורה.
אנחנו יודעים שחיתוך בין 2 קבוצות קמורות הוא קמור
מכאן נובע שחיתוך של קבוצה עם ישר הוא גם קמור.

$\Rightarrow$ נניח שהחיתוך של קו עם קבוצה הינו קמור.
ניקח 2 נקודות $x_1, x_2 \in S$.
חיתוך של S עם ישר העובר בין 2 הנקודות הינו קמור. (הנחנו)

$$\alpha x_1 + (1-\alpha) x_2 \in S \cap \text{Line}$$

$$\Downarrow$$

$$\alpha x_1 + (1-\alpha) x_2 \in S$$

$$\Downarrow$$

S קמור. ■

משפט: אי שוויון ריבועי הוא קמור

Prove that the solution set of a quadratic inequality C, is convex if $A \geq 0$.

$$C = \{ x : x^T A x + x^T b + c \leq 0 \}$$

1. ניקח ישר
2. נחפש חיתוך
3. ונחפש מתי הוא קמור יהיה קמור.

הוכחה:

① Choose an arbitrary line:

$$\{ x + tv \mid t \in \mathbb{R} \}$$

② כאשר חותכים 2 פונקציות, מציבים אחת בשנייה

$$(x+tv)^T A (x+tv) + (x+tv)^T b + c$$

③ We express ② as:

$$\alpha^2 t + \beta t + \gamma \leq 0$$

④ Where:

$$\alpha = v^T A v$$
$$\beta = b^T v + 2x^T A v$$
$$c = c + b^T x + x^T A x$$

⑤ Explanation:

$$\cdot\ (x+tv)^T = x^T + (tv)^T = x^T + v^T t^T$$

$$\square : (x^T + v^T t^T) A (x+tv) = x^T A x + x^T A t v + v^T t^T A x + v^T t^T A t v$$

המקדם של t^2 הוא יחידי פה ומקיים: $v^T A v$

המקדם של t הם שניים פה וסכומם: $2x^T A v$

$$\blacksquare\ (x+tv)^T b = b^T (x+tv) = b^T x + b^T t v$$

מתווסף למקדם של t.

2 מקדמים חופשיים

⑥ Thus, the intersection is defined as:

$$\{ x + tv \mid \alpha t^2 + \beta t + \gamma \leq 0 \}$$

קמור אם $\alpha \geq 0$.

פרבולה שקטנה מאפס עבור אינטרוול מסויים.

$$v^T A v \geq 0 \quad \forall v$$

■

משפט: פרספקטיבה של קבוצה קמורה:

הגדרה של פרספקטיבה:

$$p(x) = \frac{x_{1:n}}{x_{n+1}}$$

טענה:

The image of convex set under its perpective function is convex.

אם S קמורה, אזי $P(S)=\{P(x): x\in S\}$ גם היא קמורה.

(כלו' שגם התמונה קמורה)

הוכחה:

1. $p_1\in P(S)$, אזי קיים $x\in S$ כך ש: $p_1=p(x)$
2. $p_2\in P(S)$, אזי קיים $y\in S$ כך ש: $p_2=p(y)$
3. צריך להראות שעבור כל $0\le\mu\le 1$:

$$\mu p_1+(1-\mu)p_2 \in P(S)$$

כלומר למצוא $z\in S$ כך ש:

$$\mu p_1+(1-\mu)p_2 = p(z)$$

4. אנשים לפנינו כבר מצאו את הניחוש של z:

$$z = \theta x + (1-\theta)y$$

כאשר θ נתון באופן הבא:

$$\Rightarrow \quad \theta = \frac{\mu y_{n+1}}{\mu y_{n+1}+(1-\mu)x_{n+1}} \quad ; \quad \mu = \frac{\theta x_{n+1}}{\theta x_{n+1}+(1-\theta)y_{n+1}} =$$

$$= \frac{\mu}{x_{n+1}} = \frac{\theta}{\theta x_{n+1}+(1-\theta)y_{n+1}}$$

делим на x_{n+1}

$$\Rightarrow \quad p(z) \underset{\text{הצבת } z}{=} p(\theta x+(1-\theta)y) \underset{\text{הגדרת פרספקטיבה}}{=} \frac{\theta x_{1:n}+(1-\theta)y_{1:n}}{\theta x_{n+1}+(1-\theta)y_{n+1}} =$$

$=$

$$= \frac{\theta X_{1:n} + (1-\theta) y_{1:n}}{\theta X_{n+1} + (1-\theta) y_{n+1}} =$$

$$\frac{\theta}{\theta X_{n+1} + (1-\theta) y_{n+1}} = \frac{\mu}{X_{n+1}} \qquad \frac{1-\theta}{\theta X_{n+1} + (1-\theta) y_{n+1}} = \frac{1-\mu}{y_{n+1}}$$

$$= \frac{\mu \cdot X_{1:n}}{X_{n+1}} + \frac{(1-\mu)\, y_{1:n}}{y_{n+1}} = \mu p(x) + (1-\mu)\, p(y) =$$

$$= \mu p_1 + (1-\mu) p_2 \quad \blacksquare$$

הגדרה: פונקציה קמורה:

אם קיימים x ו y ו $0 \le \lambda \le 1$ כך שמתקיימת קומבינציה לינארית

$$\lambda x + (1-\lambda)y$$

אם מתקיים:

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$$

אזי f קמורה.

Отмич е: x и y векторы и матрицы.

משפט: קמירות של פונקציה עם הגבלה לישר

← פונקציה f הינה קמורה אמ"מ היא קמורה כאשר מגבילים אותה לישר כלשהו.

→ A function is convex if and only if it is convex when restricted to any line that intersects its domain.

$$f(x) \text{ קמורה ב-} x \iff g(t) = f(ty + (1-t)x)$$

$g(t)$ קמורה ב-t לכל x ו y.

הוכחה:

① נגדיר ישר: $f(ty + (1-t)x) = g(t) = f(x + t(y-x))$

② $\Rightarrow$ נניח כי פונקציה f קמורה

$$g(\underbrace{\theta t + (1-\theta)\bar{t}}_{t}) = f(x + \underbrace{(\theta t + (1-\theta)\bar{t})}_{t}(y-x)) =$$

$$= f(x + (y-x)(\theta t + (1-\theta)\bar{t})) = f(x + y\theta t + y(1-\theta)\bar{t} - x\theta t - x(1-\theta)\bar{t}) =$$

↲ נפתח סוגריים בפנים

$$= f(x + \theta(yt - xt) + (1-\theta)(y\bar{t} - x\bar{t})) =$$

↲ נוציא θ ו$(1-\theta)$

$$= f(x + \theta x + \theta t(y-x) + (1-\theta)\bar{t}(y-x) - \theta x) =$$

↲ נוסיף ונחסר $\theta x, -\theta x$

$$= f(\theta(x + t(y-x)) + (1-\theta)(x + \bar{t}(y-x))) \le$$

$$\leq \underbrace{\theta f(x+t(y-x))}_{g(t)} + (1-\theta)\underbrace{f(x+\bar{t}(y-x))}_{g(\bar{t})}$$

convexity

③ ⇐ נניח ש-g קמורה ב-t לכל x,y.

$$g(\theta) \leq \theta g(1) + (1-\theta) g(0)$$
$$\|$$
$$f(x+t(y-x)) \leq \theta f(y) + (1-\theta) f(x) \quad \blacksquare$$

משפט: קמירות של פונקציה:

פונקציה גזירה ברציפות $f(x)$ קמורה אמ"מ הפונקציה מעל הקירוב שלה:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

הוכחה:

① Let's restrict to a line

$$g(t) = f(ty + (1-t)x)$$

② Deriv:

$$g'(t) = \nabla f(ty + (1-t)x)^T (y-x)$$

③ ⇐ נניח כי f קמורה: ב-x לכן $g(t)$ קמורה ב-t

$$g(1) \geq g(0) + (1-0) g'(0) \qquad t=0,\ y=1.$$

Which is exactly what we need to prove.

④ ⇒ נניח כי f מעל הקירוב שלה:

ניקח 2 נקודות:

$y \in g \qquad y = ty + (1-t)x$

$x \in g \qquad x = \bar{t}y + (1-\bar{t})x$

to obtain:

$$f(ty+(1-t)x) \geq f(\bar{t}y+(1-\bar{t})x) + \nabla f(\bar{t}y+(1-\bar{t})x)^T(ty+(1-t)x - \bar{t}y-(1-\bar{t})x)$$

$$f(ty+(1-t)x) \geq f(\bar{t}y+(1-\bar{t})x) + \nabla f(\bar{t}y+(1-\bar{t})x)^T(y-x)(t-\bar{t}).$$

which means: $g(t) \geq g(\bar{t}) + g'(\bar{t})(t-\bar{t})$

and therefore $g(t)$ is convex and so is $f(x)$. ■

משפט: קמירות של פונקציה:

פונקציה גזירה ברציפות פעמיים $f(x)$ - נתון.

$f(x)$ הינה קמורה אמ"מ $\nabla^2 f(x)>0$.

הוכחה:

① נגביל על ישר $g(t) = f(a+tb)$

ונגזור:

$$g'(t) = b^T f(tb+a)$$

$$g''(t) = b^T \nabla^2 f(tb+a) b \geq 0$$

which is equivalent to $\nabla^2 f(x)>0$. ■

קשר בין קבוצות לפונקציות קמורות:

משפט:

אם $f(x)$ קמורה, אזי S קמורה.

$$S = \{x : f(x) \leq 0\}$$

הוכחה:

① $x,y \in S$ אזי: $f(x) \leq 0$, $f(y) \leq 0$.

② כדי להוכיח ש-S קמורה צריך להראות שקומבינציה לינארית בתוך קטנה מאפס.

$$f(\lambda x + (1-\lambda)y) \underset{f \text{ קמורה}}{\leq} \lambda \underset{\leq 0}{f(x)} + (1-\lambda)\underset{\leq 0}{f(y)} \leq 0$$

■

הגדרה: אפיגרף, egigraph:

The epigraph is the set above the function

i.e $\{[x\ t] : f(x) > t\}$

משפט: מרכזי

קשר בין קבוצות לפונקציות

A function $f(x)$ is convex in x iff its epigraph $\{[x\ t] : f(x) \leq t\}$ is convex in $[x\ t]$.

הוכחה:

$\Leftarrow$ כיוון אחד: נניח כי $f(x)$ קמורה

$f(x) - t$ גם קמורה.

$f(x) - t \leq 0$ קבוצה קמורה.

$\downarrow$

$\Rightarrow$ כיוון שני: נניח אפיגרף של f קמורה:

$$\left\{ \begin{bmatrix} \hat{x} \\ t \end{bmatrix} : f(x) \leq t \right\} \text{ - convex}$$

① Consider x_1 and x_2 and let:

$$t_1 = f(x_1)$$
$$t_2 = f(x_2)$$

② Then: $(x_1, t_1) \in epif$

$(x_2, t_2) \in epif$

③ So: $\lambda x_1 + (1-\lambda)x_2 \in epif$ → מכיוון שהאפיגרף

$\lambda t_1 + (1-\lambda)t_2 \in epif$ → קמור

④ And: $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda t_1 + (1-\lambda)t_2 =$

הגדרה של epi

$$= \lambda f(x_1) + (1-\lambda) f(x_2).$$

$\boxed{\text{דוגמאות לפונקציות קמורות:}}$

Norms: (גם קמורה)

כל נורמה $\|x\|$ היא פונקציה קמורה.

הוכחה:

$$\|\lambda x + (1-\lambda)y\| \underset{\text{אי שוויון משולש}}{\leq} \|\lambda x\| + \|(1-\lambda)y\| = \lambda\|x\| + (1-\lambda)\|y\| \quad ■$$

Max: (גם קמורה)

פונקציה $f(x)$ מקסימום גם קמורה $f(x) = \max\{x_1 \ldots x_n\}$

הוכחה:

$$\max\{\lambda x_i + (1-\lambda)y_i\} \leq \max_i\{\lambda x_i\} + \max_i\{(1-\lambda)y_i\} =$$

$$= \lambda \max_i\{x_i\} + (1-\lambda)\max_i\{y_i\} \quad ■$$

Quadratic - over - linear (פנ' קמורה)

$$f(x,y) = \frac{x^2}{y} \quad \text{is convex on} \{(x,y) : y > 0\}$$

הוכחה:

$$\nabla f = \begin{bmatrix} \frac{2x}{y} \\ -x^2/y^2 \end{bmatrix}$$

Не забывай что вектор!

$$\nabla^2 f = \begin{array}{cc} \left[\begin{matrix} 2/y & -2x/y^2 \\ -2x/y^2 & 2x^2/y^3 \end{matrix}\right] \\ \text{по } x \qquad\qquad \text{по } y \end{array} = \frac{2}{y^3} \cdot \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} =$$

$$= \frac{2}{y^3} \cdot \begin{bmatrix} y \\ -x \end{bmatrix} [y \; -x] \geqslant 0 \quad \blacksquare$$

если можно найти корень.

הוכחה בעזרת epi:

$$\left\{x, y, t : \frac{x^2}{y} \leq t \; , \; y \geq 0\right\}$$

$$\Downarrow$$

$$\left\{x, y, t : \begin{bmatrix} y & x \\ x & t \end{bmatrix} \geq 0\right\} \quad yt \geq x^2$$

↳ determinanta.

למדנו כי קבוצת המטריצות positive definite היא קמורה.

$\blacksquare$

המשך לדוגמאות

Log Sum Exp:

הפונקציה $f(x)$ המוגדרת באופן הבא הינה גם קמורה.

$$f(x) = \log \sum_{i=1}^{n} e^{x_i}$$

הוכחה:

← נוכיח את העובדה ע"י כך שנראה כי הנגזרת השנייה גדולה מ-0.

$$\rightarrow \nabla f(x) = \frac{\begin{bmatrix} e^{x_1} \\ \vdots \\ e^{x_n} \end{bmatrix}}{\sum e^{x_i}} \underset{e^{x_i}=z}{\overset{\uparrow}{=}} \frac{z}{1^T z}$$

$$\rightarrow \nabla^2 f(x) = \left[\frac{I}{1^T z} - \frac{z \cdot 1^T}{(1^T z)^2} \right] \underbrace{\begin{bmatrix} e^{x_i} & & 0 \\ & \ddots & \\ 0 & & \end{bmatrix}}_{\text{вынесли наружу} \;=\; = \operatorname{diag}(z)} =$$

$$\left(\frac{u}{v}\right)' = \frac{u'}{v} - \frac{v'u}{v^2}$$

$$= \left[\frac{I}{1^T z} - \frac{z 1^T}{(1^T z)^2} \right] \cdot \operatorname{diag}(z)$$

כעת צריך להראות: $u^T \cdot \nabla^2 f(x) \cdot u \geq 0$

מכיון שזה מקיים 8: $\nabla^2 f(x) \geq 0$.

$$\rightarrow u^T \nabla^2 f(x) u = \frac{u^T \operatorname{diag}(z) u}{1^T z} - \frac{u^T z 1^T \operatorname{diag}(z) u}{(1^T z)^2} =$$

$$= \frac{\sum u_i^2 z_i}{\sum z_i} - \frac{(\sum u_i z_i)^2}{(\sum z_i)^2} =$$

$$= \frac{1}{(\sum z_i)^2} \cdot \left[\sum z_i \cdot \sum u_i^2 z_i - (\sum u_i z_i)^2 \right] \geq 0$$

$$\downarrow$$

הסבר לקיום האי שוויון:

Cauchy - Shwartz:

$$\sum a_i^2 \sum b_i^2 - \left(\sum a_i b_i\right)^2 \geq 0$$

נבחר: $a_i = \sqrt{z_i}$

$b_i = \sqrt{z_i}\, u_i$

$\blacksquare$

$-\text{Logdet}(X)$:

$f(x) = -\text{logdet}(x)$ היא פונקציה קמורה.

– We will find the gradient matrix of logdet(x) by completing:

$$f(X+D) = f(x) + Tr\{D \nabla f\}$$

$$X = X^{\frac{1}{2}} X^{\frac{1}{2}}$$

λ_i ערכים עצמיים.

$$X^{\frac{1}{2}} = U D^{1/2} U^T$$

$$\Rightarrow \log|X+D| = \log|X| + \log\left|I + X^{-\frac{1}{2}} D X^{-\frac{1}{2}}\right| =$$

$$= \log|X| + \sum_i \log \lambda_i \left(I + X^{-\frac{1}{2}} D X^{-\frac{1}{2}}\right) =$$

$$= \log|X| + \sum_i \log\left(1 + \lambda_i\left(X^{-\frac{1}{2}} D X^{-\frac{1}{2}}\right)\right) =$$

$\lambda_i \cdot I = 1$

$\log(1+\varepsilon) = \varepsilon$ $\quad = \log|X| + \sum_i \lambda_i\left(X^{-\frac{1}{2}} D X^{-\frac{1}{2}}\right) =$

$\sum \lambda_i = Tr$ $\quad = \log|X| + Tr\left(X^{-\frac{1}{2}} D X^{-\frac{1}{2}}\right) = \log|X| + Tr\left(\underline{\underline{X^{-1}}} D\right)$

$$\nabla f = X^{-1}$$

Convexity proof of the function:

Restrict to a line: $X = Z + tV$

where: Z and V are symmetric.

Not necessarly positive definite.

We look for t's for which $Z + tV > 0$.

We assume $Z > 0$, i.e $t=0$ is in the domain.

$$\Rightarrow g(t) = \log|Z + tV| =$$

$$= \log|Z^{\frac{1}{2}}(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})Z^{\frac{1}{2}}| =$$

$$= \log|I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}}| + \log|Z| =$$

внесли λ_i и получили $1 = \lambda \cdot I$

$$= \sum_i \log(1 + t\lambda_i) + \log|Z|$$

$$\Rightarrow g'(t) = \sum_i \frac{\lambda_i}{1 + t\lambda_i}$$

$$g''(t) = -\sum_i \frac{\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0 \quad \blacksquare$$

פעולות על פונקציות ששומרות על קמירות:

- Non negative weighted sums :

אם f_i קמורות אזי גם $\sum w_i f_i$ קמור.

משקל.

- Composition with affine mapping:

אם $f(x)$ קמור אזי גם $g(x)$ קמור. $g(x) = f(Ax+b)$

הוכחה:

$$g(\theta x + (1-\theta)y) = f(A(\theta x + (1-\theta)y) + b) =$$

הצבה

$$= f(\theta(Ax+b) + (1-\theta)(Ay+b)) \leq$$

הוסף והורד θb

$$\leq \theta f(Ax+b) + (1-\theta) f(Ay+b) \leq$$

f קמורה

$$\leq \theta g(x) + (1-\theta) g(y) \blacksquare$$

- Pointwise maximum and supremum

← אם f_1 וגם f_2, הן 2 פונקציות קמורות אזי גם:

$$f(x) = \max\{f_1(x), f_2(x)\}$$

← מקסימום $f_2(x)$ קמורות לכל $a \in A$ (כאשר A לא חייבת להיות קמורה)

- $g(x) = \max_y f(x,y)$ is convex

1 נגדיר 2 נקודות: X_1 ו X_2 כך שמתקיים:

$$g(X_1) = \max_y f(X_1, y)$$

$$g(X_2) = \max_y f(X_2, y)$$

② נבנה קומבינציה לינארית שלהם:

$$\lambda x_1 + (1-\lambda)x_2$$

ונציב ל- g.

$$g(\lambda x_1+(1-\lambda)x_2) \underset{\text{הגדרה}}{=} \max_y f(\lambda x_1+(1-\lambda)x_2,\ y) \le$$

הגבלה ק.ל של y
f קמורה לכל y ובפרט בנקודה $\lambda y_1+(1-\lambda)y_2$

$$\le \max_{y_1,y_2} f(\lambda x_1+(1-\lambda)x_2,\ \lambda y_1+(1-\lambda)y_2) \le$$

Jensen

$$\le \lambda \max_{y_1} f(x_1,y_1) + (1-\lambda)\max_{y_2} f(x_2,y_2) =$$

$$= \lambda g(x_1) + (1-\lambda)g(x_2) \quad \blacksquare$$

דוגמאות לפונקציה קמורות

• $g(x) = \min_y f(x,y)$

f קמורה במשותף ב-x וב-y.

אזי $g(x)$ קמורה.

① נגדיר 2 נקודות: x_1 ו x_2.

$$g(x_1) = f(x_1, y_1)$$
$$g(x_2) = f(x_2, y_2)$$

$$g(\lambda x_1+(1-\lambda)x_2) \underset{\text{הגדרה}}{=} \min_y f(\lambda x_1+(1-\lambda)x_2,\ y) \le$$

הגבלה של ק.ל של y

$$\le \min_{y_1,y_2} f(\lambda x_1+(1-\lambda)x_2,\ \lambda y_1+(1-\lambda)y_2) \le$$

Jensen's extention

$$\le \lambda \min f(x_1,y_1) + (1-\lambda)\min f(x_2,y_2) = \lambda g(x_1) + (1-\lambda)g(x_2)$$

$\blacksquare$

Schur's lemma:

Let $C > 0$, then $X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \geq 0$ if and only if

$$S = A - BC^{-1}B^T \geq 0 \Rightarrow \begin{bmatrix} AC - BB^T \geq 0 \ \text{(det)} \\ A - BC^{-1}B^T \geq 0 \\ \text{детерминанта} \geq 0 \end{bmatrix}$$

הוכחה:

① A matrix X is positive semidefinite if and only if TXT^T is positive semidefinite for an invertible matrix T.

② If C is invertible then we have the following block Cholesky decomposition:

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} = \begin{bmatrix} I & BC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I & BC^{-1} \\ 0 & I \end{bmatrix}^T$$

③ The matrix $\begin{bmatrix} I & BC^{-1} \\ 0 & I \end{bmatrix}$ is invertible. And a block diagonal matrix is positive semidefinite if and only if its blocks are positive semidefinite.

④ Thus, when $C > 0$, $X \geq 0$ if and only if $S \geq 0$.

⑤ This is a generalization of cauchy Schwarts since:

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \geq 0 \iff \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}^T$$

$$\Rightarrow uu^T - uv^T[vv^T]^{-1} vu^T \geq 0.$$

Which is exactly Cauchy Shwarts if u and v are row vectors

■

תוספת: פעולות ששומרות על קמירות של פונקציות

הרכבה:

$$f(x) = h(g(x))$$

$$f'(x) = h'(g(x))g'(x)$$

$$f''(x) = h''(g(x))(g'(x))^2 + h'(g(x))g''(x).$$

מסקנות:

1) f קמורה אם: * h קמורה + לא יורדת.
   * g קמורה.

2) f קמורה אם: * h קמורה + לא עולה
   * g קעורה.

3) אם $g(x)$ קמורה, אזי $e^{g(x)}$ קמורה.

4) אם $g(x)$ קעורה, אזי $\log(g(x))$ קעורה.

פרספקטיבה של פונקציה:

If $f(x)$ is convex, then so its perspective function $g(x,t) = tf\left(\frac{x}{t}\right)$ for $t>0$.

הוכחה:

Assume : $epi(f) = \{(x,z) : f(x) < z\}$ is convex.

Now consider: $egi(g) = \left\{ \begin{bmatrix} x \\ t \\ z \end{bmatrix} : t \geq 0 \ , \ tf\left(\frac{x}{t}\right) \leq z \right\}$

$$= \left\{ \begin{bmatrix} x \\ t \\ z \end{bmatrix} : t > 0, \quad f\left(\frac{x}{t}\right) \leq \frac{z}{t} \right\} =$$

поделили на t

$$= \left\{ (x, t, z) : t > 0, \left(\frac{x}{t}, \frac{z}{t}\right) \in \operatorname{epi} f \right\}$$

Thus epi g is the inverse image of epi f under the perspective function. ■

Example: The perspective of $f(x) = x^T x$ is $g(x,t) = \frac{x^T x}{t}$ which is convex in both x and t.

חזרה על אלגברה לינארית:

We already reviewed EVD. Most square matrices can be diagonalized as: $A = UDU^{-1}$.

When they are also symmetric they can always be diagonalized as: $A = UDU^T$, where U is orthonormal.

A positive definite matrix in S^m_{++} takes us from the subspace of its eigenvectors in R^m to the same subspace in R^m. A positive semidefinite matrix is still symmetric. It does the same thing but may shrink the subspace. Some eigenvectors go to zero.

$$\text{positive semidefinite } R^n = C(A) \overset{\perp}{\oplus} N(A)$$

Given $A \succeq 0$, any vector $x \in R^n$ can be written as

$$x \in R^n \iff x = x_1 + x_2 , \quad x_1 \in C(A) , \quad x_2 \in N(A) , \quad x_1^T x_2 = 0.$$

In terms of EVD we have:

$$A = [U_1 \; U_2] \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} [U_1 \; U_2]^T$$

and U_1 is a basis for $C(A)$, whereas U_2 is a basis for $N(A)$.

The four fundamental supspaces of linear algebra. Given a $m \times n$ matrix A of rank r, we have the following subspaces:

① **Column Space:** $C(A) \in R^m$ -
all the combinations of columns of A, dimension r.

$$C(A) = \{ Ax : x \in \mathbb{R}^n \}$$

② **Null space:** $N(A) \in \mathbb{R}^n$ -
all the vectors such that $Ax = 0$, dimension $n-r$.

$$N(A) = \{ x \in R^n : Ax = 0 \}$$

③ **Row space:** $C(A^T) \in R^n$ -
all the combinations of rows of A, dimension r.

④ **Left null space:** $N(A^T) \in R^m$ -
all the vectors such that $x^T A = 0$, dimension $m-r$.

A fundamental result:

$$R^m : C(A) = [N(A^T)]^{\perp}$$

The null space is the orthogonal complement of the row space

$$R^n : C(A^T) = [N(A)]^{\perp}$$

The left null space is the orthogonal complement of the column space.

The main tool for working with non square, non-symmetric matrices is singular value decomposition.

$$(*) \quad A = UDV^T$$

where U and V are square unitary matrices and $D = \operatorname{diag}\{d_i\}$ is rectangular diagonal matrix with non negative singular values.

This is very similar to EVD, except that now 4 subspaces are involved in possibly different dimensions:

$$A = \begin{bmatrix} \overset{m\times r}{U_1} & \overset{m\times m-r}{U_2} \end{bmatrix} \begin{bmatrix} \overset{r\times r}{\Lambda} & \overset{r\times n-r}{0} \\ \overset{m-r\times r}{0} & \overset{m-r\times n-r}{0} \end{bmatrix} \begin{bmatrix} \overset{n\times r}{V_1} & \overset{n\times n-r}{V_2} \end{bmatrix}^T$$

Конец.

Convex Optimization

Standard form:

$$\min_{x \in S} f_0(x)$$

$$\text{s.t } f_i(x) \leq 0$$

$$\text{s.t } h_i(x) = 0$$

The problem is convex if S is a convex set,
f_i are convex functions
and $h_i(x)$ are affine functions.

Теперь мы покажем несколько трюков чтобы превратить задачу которую мы получаем в convex:

1 החלפת משתנים:

$$X = \phi(z), \quad Z = \phi^{-1}(x)$$

The problem is equivalent to: (подставим вместо x).

$$\min_{\phi(z) \in S} f_0(\phi(z))$$

$$\text{s.t } f_i(\phi(z)) \leq 0$$

$$\text{s.t } h_i(\phi(z)) = 0.$$

Indeed, when we find the optimal z^* then the optimal x^* is $x^* = \phi(z^*)$.

Example: Amplitude and phase recovery from a noisy sine function (frequency f is known).

$$\min_{A,\phi} \sum |y_i - A\cos(2\pi f x_i + \phi)|^2$$

$$\downarrow$$

Non convex least squares!

Now, using a simple change of variables from polar coordinates to cartesian:

$$a_1 = A\cos(\phi)$$
$$a_2 = A\sin(\phi)$$

We obtain:

$$A\cos(2\pi f x_i + \phi) = a_1\cos(2\pi f x_i) - a_2\sin(2\pi f x_i)$$

And a convex Least squares:

$$\min_{a_1,a_2} \sum_i \left| y_i - a_1\cos(2\pi f x_i) + a_2\sin(2\pi f(x_i)) \right|^2$$

② טרנספורמציות מונוטוניות

← בדרך כלל עוזרות להפוך את הבעיה ליותר קלה, לאו דווקא קמורה.

Example: quadratic minim 2 tion vs. norm minimization.

$$\min_x (y-Hx)^T(y-Hx) = \min_x \|y-Hx\|_2^2$$

$$\Updownarrow$$

$$\min_x \|y-Hx\|_2$$ – минимум один и тот же.

– Интерфэйсы в мире этот трюк делают взяв корень
То-есть мало кто правда решает квадратные,
а берут корень и решают обычно.

$$\downarrow$$

3 העלמת אילוצים לינאריים:

$$\min_x f(x)$$
$$s.t \quad Ax=b$$

מאלגברה לינארית, אנחנו יודעים שיש 3 אפשרויות ל: $Ax=b$.

1) אין פתרון - למערכת משוואות ולכן גם לבעיית אופטימיזציה.

2) פתרון יחיד - ______ " ______.

3) יש אינסוף פתרונות (מבין כולם צריך לבדוק ולקחת את המינימלי).

במקרה השלישי:

$$x = x_1 + x_2$$
$$Ax_1 = b$$
$$x_2 \in N(A).$$

ואז הבעיה לגמרי שקולה:

$$\min_{x_2} f(x_1 + x_2)$$
$$s.t \quad x_2 \in N(A)$$

$$\min_z f(x_1 + Fz)$$

- כאשר x_1 זה קבוע פתרון כלשהו
- בעיה בלי אילוצים.

Example: $\min_{x \in \mathbb{R}^n} f(x_1, x_2, \dots, x_n)$

$$s.t \quad \sum_{i=1}^{n} x_i = 0$$

→ Сумма каждого вектора ноль. Поэтому я дим какое-то значение, которое не '0', а потом отнимем.

$$\min_{z \in \mathbb{R}^{n-1}} f\left(z_1, z_2, \dots, z_{n-1}, -\sum_{i=1}^{n-1} z_i\right)$$

4 הוספת אילוצים:

$$\min_{x} f(Ax+b)$$

$\Updownarrow$

$$\min_{x,z} f(z)$$
$$\text{s.t} \quad z = Ax+b.$$

משפט: אל מינימום גלובלי (משפט תנאי מספיק והכרחי)

אם נתון: $f(x)$ גזירה ברציפות.

$f(x)$ קמורה.

S קבוצה קמורה.

$$\min_{x\in S} f(x)$$

אזי: $x\in S$ הוא מינימום גלובלי אמ"מ $\forall y\in S: \nabla^T f(x)(y-x) \geq 0$

הוכחה:

$\Rightarrow$ נניח כי הנגזרת היא חיובית.

אזי, נניח ש $x,y\in S$ והאי שוויון מתקיים.

$$f(y) \geq f(x) + \nabla^T f(x)(y-x)$$

נובע מקמירות של f.

אבל גם אזי מתקיים:

$$f(x) + \nabla^T f(x)(y-x) \geq f(x)$$

הנגזרת חיובית

מה שמוכיח כי $f(x)$ הוא מינימום ■

$\Leftarrow$ נניח כי x הוא מינימום גלובלי, אבל שהתנאי לא מתקיים (קטן מ-0)

ואז נגיע לסתירה.

$$y\in S: \quad \nabla f^T(x)(y-x) < 0$$

נתבונן בנקודה:

$$z(t) = (1-t)x + ty \qquad t\in[0,1]$$

הנקודה פיזבילית מכיוון שגם x ו y.

$\downarrow$

← עבור t קטן נקבל כי:

$$f(z(t)) < f(x(t))$$

מכיוון ש:

$$\left.\frac{df(z(t))}{dt}\right|_{t=0} = \nabla f^T(x)(y-x) < 0$$

- Если увидим что спускаемся, то стира.
- $z(0)$ это и есть x. Если двинемся, то уменьшимся а это бред, так как x и так минимум.

■

Two famous special cases:

- **Unconstrained convex optimization:**

A necessary and sufficient condition for optimality is:

$$\nabla f(x) = 0$$

Since the condition must hold for

$y = x + z$ and $y = x - z$ for any z.

That's:

$z = y - x$ and $z = -(y-x)$

According to the theorem we know that x is global minimum iff $\nabla^T f(x)(y-x) \geq 0$. $\forall y \in S$.

$$\nabla f^T(x)(y-x) =$$
$$= \nabla f^T(x) z \geq 0 \quad \forall z$$

$$\nabla f^T(x) z \leq 0 \quad \forall z$$

This holds only if gradient is identically 0.

■

↓

• Convex scalar minimization on an interval I

$I = \{a \leq x \leq b\}$, a necessary and sufficient condition for optimality is $x \in I$ and
$$f'(x)(y-x) \geq 0, \ \forall y \in I.$$

• Now if x is strictly inside the inteval, then as before the condition reduces to $f'(x) = 0$.

• If it is on the border, e.g, $x=b$ then for all. $y \in I$ we have $y-x \leq 0$, and the condition reduces to $f'(x) \leq 0$, which means that the end point is a global minimizer.

Convex Optimization Problems

1. Least Squares:

$$\min_x \|Ax - b\|_2^2$$

This is of course an unconstrained convex optimization problem (with and without the square as it is a norm or positive definite quadratic form).

Least squares' optimality conditions are:

$$\nabla f(x) = A^T(Ax-b) = 0.$$

* Как на пред. странице.
Просто обнуляем градиент.

→ For an unconstrained problem (i.e $m=p=0$) the condition of the "deriv >0" reduces to necessary and sufficient condition $\nabla f_0(x) = 0$.

Case 1:

A is square invertible matrix.

$Ax = b$ has a unique solution which satisfies the optimality conditions

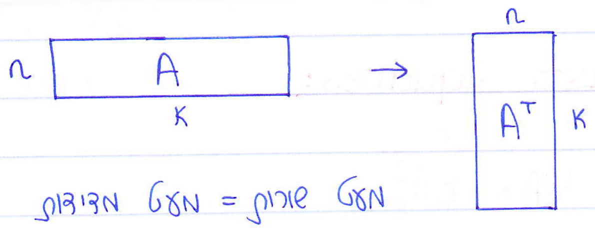
$$x = A^{-1} b$$

Case 2:

Less equations that unknowns.

$Ax = b$ has infinite solutions all of which are optimal as they satisfy the optimality conditions.

→ The Hessian has zero eigenvalues as $A^T A$ is a large matrix with small rank. (אין מספיק משוואות , $n < k$)



מעל שורות = מעל עמודות

Rank = is the dimension of the vector space generated (or spanned) by its columns.

→ However, AA^T is a small matrix, usually of full rank and invertible, and it is easy to verify that one solution is:

$$x_0 = \underbrace{A^T}_{k \times n} \underbrace{(AA^T)^{-1}}_{n \times n} b = A^+ b.$$

A^+ : is known as the Moore-Penrose generalized inverse (aka pseudo-inverse) of A.

נבדוק:

$$A^T(Ax - b) = A^T(A(A^T(AA^T)^{-1}b) - b) = A^T(\cancel{AA^T(AA^T)^{-1}}b - b) =$$

$$= A^T(b - b) = 0.$$

↓

This is one solution and the rest can be characterized as:

$$x = x_0 + z \quad , \quad z \in N(A).$$

Case 3:

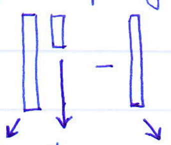
More equations that unknowns.
$Ax = b$ has no solutions, but we can still search for the least squares approximation. We need to solve:

$$A^T A x = A^T b$$

Почему? : $\nabla f = A^T \underbrace{(Ax-b)}_{\text{не обнуляется}} = 0$

именно это нужно решить?

В форме рисунков получается:



Наш A Наш x Наш b

Но это все не может обнулится.
Least squares предлагает это "опустить".

- "Опустить" = A^T (транспоз) $\boxed{A^T}$.

$$Ax = b$$

↓ Опускаем

$$A^T A x = A^T b$$

$$x = \underbrace{(A^T A)^{-1} A^T}_{\text{pseudo inverse}} b$$

$$x = A^+ b.$$

To conclude:

In all cases the LS solution is $x = A^+ b$.
It coinsides with the inverse if A is square and invertible, but otherwise has different meanings.

↓

What happens when the rows|columns are NOT linearly independent?

- The solution is still the $A^+ b$ and the pseudo-inverse always exists and is unique.

→ To gain intuition we will see the scalar case.

$$\min_x |ax-b|^2$$

$$\Big\downarrow \text{נגזרת}$$

$$2a(ax-b)=0$$

$$\Big\downarrow \text{חלוק}$$

$$ax-b=0$$

$$\Big\downarrow \text{חילוק}$$

$$x = b/a.$$

To be consistent with previous solutions, we define

$$x^+ = \begin{cases} 1/x & x \neq 0 \\ 0 & x = 0 \end{cases}$$

⇒ The general case:

$$Ax = b \qquad \text{Опускаем } A = UDV^T \text{ (according to SVD)}$$

$$A^T A x = A^T b$$

$$\Big\downarrow \text{SVD}$$

$$VD^T \cancel{U^T} \cancel{U} D V^T x = VD^T U^T b$$

$$\cancel{V} D^T D V^T x = \cancel{V} D^T U^T b$$

$$D^T D \underbrace{V^T x}_{\tilde{x}} = D^T \underbrace{U^T b}_{\tilde{b}} \text{ - diagonalized the system.}$$

$$\frac{d_i}{d_i^2} = \frac{1}{d_i} = d_i^+ \qquad d_i^2 \tilde{x}_i = d_i \tilde{b}_i \qquad (A^+ = A^T (AA^T)^{-1})$$

$$\Big\downarrow$$

$$\tilde{x}_i = d_i^+ \tilde{b}_i$$

$$\Big\downarrow$$

$$\tilde{x}_i = d_i^+ \tilde{b}_i$$

$$\tilde{x} = D^+ \tilde{b}$$

$\downarrow$

Подставим обратно:

$$V^T x = D^+ U \; b$$

$$x = \underbrace{V D^+ U^T}_{A^+} b = A^+ b \quad , \Rightarrow A^+ = V D^+ U^T$$

$$x = A^+ b. \quad \blacksquare$$

$\downarrow$

2. Problems with equality Constraints only

$$\min_{x : Ax = b} f(x)$$

① The optimality conditions are:

$$\nabla f(x)^T (y-x) \geq 0 \quad \forall y \quad Ay = b, \; y \in S$$

② Since x is feasible, then $y = x + v$ where $v \in N(A)$

and therefore : $y - x = v$

$$\nabla f(x)^T v \geq 0, \; \forall v \in N(A)$$

③ If a linear function is nonnegative on a subspace, then it must be zero on the subspace.

So it follows that :

$$\nabla f_0(x)^T v = 0 \text{ for all } v \in N(A)$$

In other words:

$$\nabla f_0(x) \perp N(A)$$

null space $\perp$ row space

④ Using the fact that $N(A)^{\perp} = C(A^T)$, this optimality condition can be expressed as $\nabla f_0(x) \in C(A^T)$. i.e, there exists a $v \in \mathbb{R}^p$ such that :

$$\nabla f_0(x) + A^T v = 0$$

⑤ Together with the requirment $Ax = b$ (i.e that x is feasible), this is a classical Lagrange multiplier optimality condition. ■

3. פתרון מינימום נורמה

Least-squares solution of linear equations

① The most common least-norm problem involves the Euclidean of ℓ_2-norm. By squaring the objective we obtain the equivalent problem.

$$\text{minimize } \|x\|_2^2$$
$$\text{s.t} \quad Ax = b$$

② Introducing the dual variable $v \in \mathbb{R}^m$, the optimality conditions are:

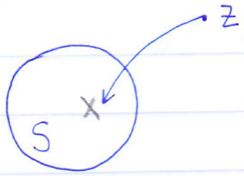
$$2x^* + A^T v^* = 0, \quad Ax^* = b.$$

③ This is a pair of linear equations, and readily solved.
From the first equation:
$x^* = -\frac{1}{2} A^T v^*$, substituting it into the second equation we obtain:

$$-\frac{1}{2} A A^T v^* = b \quad \text{and conclude:}$$

$$v^* = -2(AA^T)^{-1} b.$$
$$x^* = -\frac{1}{2} A^T \cdot (-2)(AA^T)^{-1} b =$$
$$= A^T (AA^T)^{-1} b = A^+ b. \quad \blacksquare$$

4. Projections:



Общая идея:

Хотим найти элемент в группе S, который больше всего похож на элемент z.

① Now we define the projection of z onto a convex set S.

$$P_S(z) = \underset{x \in S}{\operatorname{argmin}} \|x - z\|$$

② We will start with projections onto a subspace (which is of course a convex set)

$$S = \{x : x = Ay \text{ for some } y\}$$

כאשר S הוא ה- column space של A.

③ Where A is a tall matrix (more rows than columns)

$$\min_{\{x : x = Ay \text{ for some } y\}} \|x - z\|$$

④ And now we get: (трюк = перевести в условия)

$$\min_{x,y} \|x - z\|$$
$$\text{s.t} \quad x = Ay$$

⑤ We eliminate the linear constraints:

$$\min_{y} \|Ay - z\|$$

⑥ And the solution is clearly

$$y^* = A^+ z$$

⑦ And in terms of x:

$$x^* = Ay^* = AA^+z$$

⑧ So that:

$$P_s(z) = AA^+z \quad \blacksquare$$

מקרי אורתוגונלית של הטלה על $S = C(A)$ (c. space)

① The optimality conditions read:

$$(x^* - z)^T (x - x^*) \geq 0, \ \forall x \in C(A)$$

② $x \in C(A)$

It means that $x - x^*$ also $\in C(A)$

③ And basically we have:

$$(x^* - z)^T \geq 0, \ \forall x \in C(A)$$

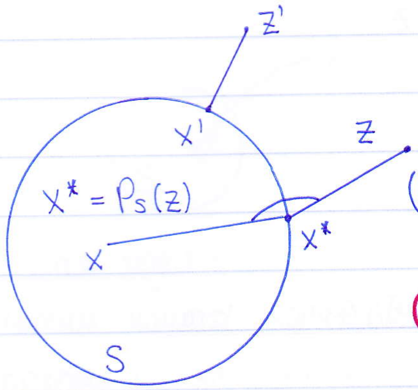
④ This is true for x and $-x$ so

$$(x^* - z)^T x = 0, \ \forall x \in C(A)$$

Which means that the error (gradient) is orthogonal to the range of A. Otherwise we can improve. $\blacksquare$

↓

משפט על הטלות:



תהיה S קבוצה לא ריקה, סגורה וקמורה.

① לכל z קיימת הטלה והיא יחידה

② x^* (המינימום של הבעיה) הינו הטלה $(x^*=P_S(z))$ אמ"מ

$$(\#,\#) \quad (z-x^*)^T(x-x^*) \le 0 \quad \forall x\in S$$

③ ההטלה הינה רציפה ולא מרחיבה:

$$(\#) \quad \|P_S(x)-P_S(y)\|_2 \le \|x-y\|_2 \quad \forall x,y.$$

הכוונה* ⇐ $z \; - \; z' \quad \le \|x-x'\|$

מה הבנו מהשיעור:

(א) הזווית תמיד תהיה קהה מכיוון שמכפלה הפנימית היא שלילית.

(ב) רציפה (משמעות): אם נזיז את z אז יזוז את x^*.

(ג) מרחיבה (משמעות): מרחק בין x^* לבין x' קטן ממרחק בין z ל-z'.

הוכחה:

① We will only give a sketch of a proof.
Skip existence.
Uniqueness is due to strict convexity.

② Non-expansive:
Consider some x and y.
Then $P_S(x)$ and $P_S(y)$ are in S and the optimality conditions yield: (פעם אחת נציב $y^*=P_S(y)$ ופעם שנייה $x^*=P_S(x)$)

$$(z-x^*)^T(x-x^*) \le 0$$

$$\downarrow$$

$$(P_S(y) - \underbrace{P_S(x)}_{x^*})^T(x-P_S(x)) \le 0$$

for every z
unique projection

$$(P_S(x) - \underbrace{P_S(y)}_{y^*})^T(y-P_S(y)) \le 0$$

$$\downarrow$$

Add this up to obtain:

$$\left(\overset{a}{P_S(x)} - \overset{b}{P_S(y)}\right)^T \left(x - \overset{a}{P_S(x)} - y + \overset{b}{P_S(y)}\right) \leq 0$$

$$(a-b)^T(x-a-y+b) \leq 0$$
$$(a^T-b^T)(x-a-y+b) \leq 0$$
$$a^Tx - a^Ta - a^Ty + a^Tb - b^Tx + b^Ta + b^Ty - b^Tb \leq 0 \quad / \cdot (-1)$$
$$-a^Tx + a^Ta + a^Ty - a^Tb + b^Tx - b^Ta - b^Ty + b^Tb \geq 0$$
$$a^Ta - a^Tb - b^Ta + b^Tb \geq a^Tx - a^Ty - b^Tx + b^Ty$$
$$\|a-b\|_2^2 \geq y(b^T - a^T) - x(b^T - a^T)$$
$$\|a-b\|_2^2 \geq (b^T - a^T)(y-x)$$

$$\|P_S(x) - P_S(y)\|_2^2 \geq (P_S(y) - P_S(x))^T(y-x)$$
$$\|P_S(y) - P_S(x)\|_2^2 \leq (P_S(y) - P_S(x))^T(y-x)$$

Cauchy schwarts: $a^Tb \leq \|a\| \cdot \|b\|$.

$$\|P_S(y) - P_S(x)\|_2^2 \leq (P_S(y) - P_S(x))^T(y-x) \underset{C.S}{\leq} \|P_S(y) - P_S(x)\| \cdot \|y-x\|$$

$$\|P_S(y) - P_S(x)\|_2^{\cancel{2}} \leq \cancel{\|P_S(y) - P_S(x)\|} \cdot \|y-x\|$$

$$\|P_S(y) - P_S(x)\| \leq \|y-x\|_2. \quad \blacksquare$$

Example: Projection onto the semidefinite cone (with frobenius norm).

$$S = \{X : X \succcurlyeq 0\}$$

p.s.d קבוצה של

$$Y = UDU^T \qquad (\text{s.v.d})$$

$$X = \underset{X \succcurlyeq 0}{\operatorname{argmin}} \|X - Y\|^2$$

Now, $X \succcurlyeq 0$ is optimal iff :

$$\operatorname{Tr}\{(Y-X)(Z-X)\} \leq 0, \quad \forall Z \preccurlyeq 0.$$

Lets guess : $X = U[D]_+U^T$ then $X > 0$ and:

$$\operatorname{Tr}((Y-X)(Z-X)) = \underbrace{\operatorname{Tr}((Y-X)Z)}_{A} - \underbrace{\operatorname{Tr}((Y-X)X)}_{B} \leq 0.$$

A: חיובי x שלילי = שלילי

B: (if $y > 0$ mo $y^* = y$ $= 0$) $=$ zero
(if $y < 0$ mo $y^* = 0$).

$$B \Rightarrow \operatorname{Tr}((Y-X)X) = \operatorname{Tr}(YX) - \operatorname{Tr}(X^2) =$$

$$= \operatorname{Tr}(\cancel{U}DU^TUD_{(+)}\cancel{U^T}) - \operatorname{Tr}(\cancel{U}D_+U^TUD_{(+)}\cancel{U^T}) =$$

הצבה

$$= \operatorname{Tr}(DD_+) - \operatorname{Tr}(DD_+) = 0. \quad \blacksquare$$

↓

בוחן פתע:

① אם A ו B 2 מטריצות מתחלפות (כלומר: $BA = AB$).
אזי: $A, B > 0$ מתקיים: $Tr(AB) \geq 0$.

הוכחה:

ניקח שורש: $B = LL^T$.

$$Tr(ALL^T) = Tr(L^T A L) =$$
$$= \sum l_i^T A l_i \geq 0. \blacksquare$$

② אם $B > A \geq 0$, אזי $B^{-1} < A^{-1}$.

הוכחה:

- אם כן מתחלפות: $D_B > D_A$

$$A = U D_A U^T$$
$$B = U D_A U^T$$

ואזי: $\frac{1}{D_{A_{ii}}} < \frac{1}{D_{B_{ii}}}$

מכיוון שמשתמשים

במטריצה שלה אותו U. $\blacksquare$

← מוספת. כולם מתחלפים עם מטריצה I ולכן בעזרת שורש
אפשר להגיע אחת מהן I ולהוכיח

Linear Programming

- Linear functions are concave and therefore the solutions to LP always lie on extreme points.

Standard form:
$$\min_{x} \ C^T X$$
$$\text{s.t} \quad GX \leq h$$
$$AX = b$$

Some examples:

- Показываем какие задачи могут быть выражены как LP.

1. Power Control:

המטרה היא להקטין את המצבוע של הכוח כאשר מעבירים מידע.

עפ"י הנוסחא שלומדים ברשתות אנחנו נקבל כי:

$$\log\left(1 + \frac{H_{ii}\overset{\text{הספק}}{p_i}}{\underset{\text{רעש}}{\sigma^2} + \underset{\text{רעש ממקור אחר}}{\sum H_{ij}P_j}}\right) \geq \underset{\text{מה שמשלמים (נתון).}}{\gamma_i}$$

$$\Rightarrow \min_{p_i} \sum_i p_i$$
$$\text{s.t} \quad \log\left(1 + \frac{h_{ii}p_i}{\sum_{j \neq i} p_j h_{ij} + \sigma^2}\right) \geq \gamma_i$$
$$p_i \geq 0$$

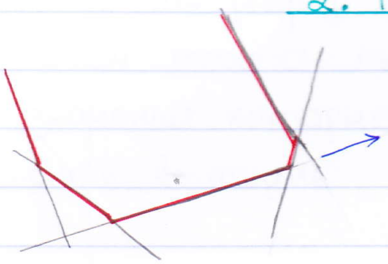
$$\Rightarrow \min_{p_i} \sum_i p_i$$
$$\text{s.t} \quad \frac{h_{ii}p_i}{\sum_{j \neq i} p_j h_{ij} + \sigma^2} \geq e^{\gamma_i} - 1$$
$$p_i \geq 0$$

$$\Rightarrow \quad \min_{p_i} \sum_i p_i$$

$$\text{s.t} \quad h_{ii} p_i \sum_{j\neq i} p_j h_{ij} + \sigma^2 \geq \sum_{j \neq i} p_j (e_i^{\gamma} - 1) h_{ij} + (e_i^{\gamma} - 1)\sigma^2$$

$$p_i \geq 0 \quad \blacksquare$$

2. Piecewise linear minimization



Мы возьмём максимум по всем этим линиям. (так поэтому и получим те что выше).

The problem: $\min_x f(x)$

$$\left\{ \min_x \max_i a_i^T x + b_i \right.$$

can be reformulated:

$$\begin{cases} \min\limits_{x,t} & t \\ \text{s.t} & \max\limits_i a_i^T b + b_i \leq t \end{cases}$$

$$\begin{cases} \min\limits_{x,t} & t \\ \text{s.t} & a_i^T x + b_i \leq t \quad \forall i \end{cases}$$

$$\boxed{\begin{cases} \min\limits_{x,t} & t \\ \text{s.t} & Ax + b \leq t\mathbb{1} \end{cases} \quad \blacksquare}$$

Note: can be written as

$$\min_x \|Ax + b\|_\infty$$

$\downarrow$

3. L_1 minimization

$$\min_x \|Ax - b\|_1$$

→ Эта норма наказывает нас меньше в случае ошибки. Хотим чтобы ошибка была sparse. Норма ℓ_1 хитра.

$$\begin{cases} \min\limits_{x,t} & 1^T t \\ \text{s.t} & t_i = |a_i^T x + b_i| \ \forall i \end{cases}$$

– Relaxing the equality to an inequality :

$$\begin{cases} \min\limits_{x,t} & 1^T t \\ \text{s.t} & t_i \geq |a_i^T x + b_i| \ \forall i \end{cases}$$

– Open the $|\cdot|$:

$$\begin{cases} \min\limits_{x,t} & 1^T t \\ \text{s.t} & -t_i \leq |a_i^T x + b_i| \leq t_i \quad t_i \geq 0 \end{cases}$$

■

→ Together with the previous example both L_1 and L_∞ norm minimizations can be expressed as LP.

↓

4. Linear Fractional Programming

* Общая схема работы:

Мы добавляем больше параметров и тем самым увеличиваем наш мхум.

→ Всегда ищем хасам тахтон - что означает приблизить раздутый радиус к мхуму.
Хасам тахтон потому-что min(x) даст теперь значения которые меньше и им не разрешено быть в мхуме.

→ Мы добавляем илуцим чтобы:
"Приподнять в воздух и потом бросить".
В начале нам будут не нужны, но когда бросим кширут, они понадобятся.

Начало:

$$\begin{cases} \min\limits_{x:\, e^T x + f > 0} \dfrac{c^T x + d}{e^T x + f} \\ \text{s.t} \quad Gx \leq h \\ \qquad\ \ Ax = b \end{cases}$$

אובג'קטיב הופך ללינארים: מוסיפים משתנים ומוסיפים את
כל המשוואות לינאריים.

① נגדיר משתנים: $y = \dfrac{x}{e^T x + f}$, $z = \dfrac{1}{e^T x + f}$

↓

→ В конце мы выбросим 'y' и 'x' так как они не приятные для работы и вообще не линиарные.
Но до этого, вытащим из них по максимум информации (линии на рисунке).

We will add the following constraints:

* $z > 0$
* $y = zx \quad , \quad x = y/z.$
* $y \underset{\text{הגדרה}}{=} \dfrac{x}{e^T x + f} \underset{\text{הצבת } x}{=} \dfrac{y/z}{e^T y/z + f} = \dfrac{y}{e^T y + fz} \Rightarrow$

$$\Rightarrow e^T y + fz = 1.$$

* $Gx < h$

$G \cdot y/z < h$

$Gy < hz$

* $Ax = b$

$A\, y/z = b$

$Ay = bz$

New problem: dpopping the x:

$$\min_{y,z} \quad c^T y + dz$$

$$\text{s.t} \quad \begin{aligned} &Gy < hz \\ &Ay = bz \\ &z > 0 \\ &e^T y + fz = 1. \end{aligned}$$

⇐ בזה סיימנו את כל הדוגמאות כאשר LP מביא ⇒
⇐ פתרון מדוייק. כעת, נראה בעיות בהם LP הוא מקרב ⇒

Compressed Sensing

- A different example where we look for an approximation of a non-convex combinatorial optimization.

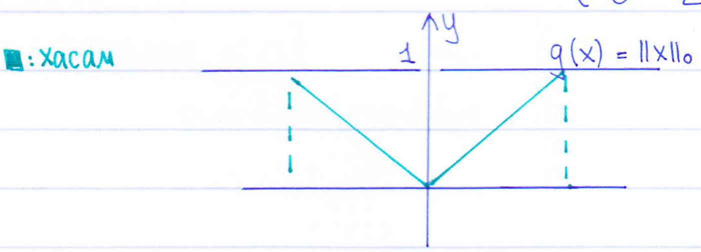
- The problem is again solving a linear system of equations $Ax = b$. But we assume less equations that unknowns, hence infinite solutions.
We already learned the minimum L2 norm solution $x = A^+ b$.

- In some settings (which are very common) we have additional prior knoweledge that x is sparse.
Thus we would like to solve:

$$\min_x \|x\|_0 \quad \text{s.t} \quad Ax = b.$$

where:

$$\|x\|_0 = \text{card}(x), \quad \text{card}(z) = \begin{cases} 1 & z \neq 0 \\ 0 & z = 0 \end{cases}$$



Как и раньше: цель найти хасам махмон
(Найти функцию под нашей функцией).

- The vector is just the sum, whenever it's one.
- The envelope of scalar card(x) on $[-1,1]$ is $|x|$.
- The envelope of vector card(x) on $\{x : \|x\|_\infty \leq R\}$ is $\frac{1}{R}\|x\|_1$.

Thus we propose to solve:

([illegible]) $$\begin{cases} \min\limits_{x} \|x\|_1 \\ \quad \text{s.t } Ax = b. \end{cases}$$

$\Downarrow$ Can be formulated as:

$$\begin{cases} \min\limits_{x,t} \ 1^T t \\ \text{s.t} \quad Ax = b \\ \qquad -t \leq x \leq t. \ \blacksquare \end{cases}$$

$\downarrow$

Second Order Cone Programming

(LP → SOCP → SDP).

$$\min_x f^T x$$

$$\text{s.t } \|A_i x + b_i\| \leq c_i^T x + d_i \quad \forall i$$

① Taking $A_i = b_i = 0$ leads to a simple linear programm.

② Taking $c_i = 0$ and $d_i \geq 0$ leads to a simple quadratic program (QP).

- But it is more general than QP as it allows one negative eigenvalue as long as rhs is positive.

- LP ∋mo $x \geq 0$, Semi Definite ∋mo $M > 0$
  A SOCP ∋mo

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leq_{\text{SOCP}} 0 \Rightarrow \left\| \begin{matrix} x_1 \\ x_2 \end{matrix} \right\| \leq x_3$$

Example: 1 Quadratic Set

$$\{x^T A^T A x + b^T x + c \leq 0\} \underset{\text{שקול לגמרי}}{\iff} \left\| \begin{matrix} \frac{1 + b^T x + c}{2} \\ Ax \end{matrix} \right\|_2 \leq \frac{1 - b^T x - c}{2}$$

נכון כאן

הוכחה:

① SOCP = Quad + LIN

$$\left\| \begin{matrix} x_1 \\ x_2 \end{matrix} \right\| \leq x_3 \iff \left\| \begin{matrix} x_1 \\ x_2 \end{matrix} \right\|^2 \leq x_3^2 \quad \text{and} \quad x_3 \geq 0.$$

② Всегда поднимай в квадрат обе стороны.
③ И требуй чтобы правая часть была больше 0.

④ Вот:

* $\frac{1-b^Tx-c}{2} \geq 0 \Rightarrow \boxed{b^Tx + c \leq 1}$ אילוץ ראשון

* Поднять в квадрат:

$$\left(\frac{1+b^Tx+c}{2}\right)^2 + x^TA^TAx \leq \left(\frac{1-b^Tx-c}{2}\right)^2$$

$\boxed{x^TA^TAx + b^Tx + c \leq 0}$ אילוץ שני

כאשר האילוץ הראשון כלול בתוך אילוץ השני ולכן חזרנו לבעיה המקורית.

■

Example 2: Hyperbolic Constraints

$$\left\{ \begin{array}{lcl} w & : & w^Tw \le xy \\ x & : & x \ge 0 \\ y & : & y \ge 0 \end{array} \right\}$$

x,y סקלרים ו-w וקטור

1) הקבוצה קמורה מכיוון שאם נחלק אותה ב y נקבל: $\frac{w^Tw}{y} \le x$
שהפרספקטיב קטן ממשהו לינארי.

2) בנוסף w^Tw זה ריבועי. וריבועי מחולק בלינארי גם קמור.

3) נוכיח כי: הבעיה שקולה לבעיית SOCP: $\left\{ \left\| \begin{matrix} 2w \\ x-y \end{matrix} \right\| \le x+y \right\}$

* Поднимем в квадрат:

$$(2w)^2 + (x-y)^2 \le (x+y)^2$$

$$4w^2 + \cancel{x^2} - 2xy + \cancel{y^2} \le \cancel{x^2} + 2xy + \cancel{y^2}$$

$$\cancel{4}w^2 \le \cancel{4}xy$$

$$\underset{\text{квадрат}}{0 \le} w^Tw \le xy.$$

* Больше 0: $x+y \ge 0.$

$$\left\{ \begin{array}{l} x+y \ge 0 \\ xy \ge 0 \\ w^Tw \le xy \end{array} \right\} \Leftrightarrow x,y \ge 0$$

Example 3: Minimizing the harmonic mean Of positive functions over a polytope:

סכום הרמוני :

$$\begin{cases} \min\limits_{x} \sum \dfrac{1}{a_i^T x + b_i} \\ \text{s.t} \quad a_i^T x + b_i > 0 \\ \qquad\quad c_i^T x + d_i \geq 0 \end{cases}$$

נוסיף משתנה :

$$\begin{cases} \min\limits_{x, t_i} \sum t_i \\ \text{s.t} \quad a_i^T x + b_i > 0 \\ \qquad\quad c_i^T x + d_i \geq 0 \\ \qquad\quad t_i = \dfrac{1}{a_i^T x + b_i} \iff t_i(a_i^T x + b_i) = 1. \end{cases}$$

It's only better to decrease t_i, so we can make the equality an inequality:

$$\begin{cases} \min\limits_{x, t_i} \sum\limits_i t_i \\ \text{s.t} \quad a_i^T x + b_i > 0 \\ \qquad\quad c_i^T x + d_i \geq 0 \\ \qquad\quad t_i(a_i^T x + b_i) \geq 1 \\ \qquad\quad t_i \geq 0. \end{cases}$$

As SOCP:

$$\left\| \begin{matrix} 2 \cdot 1 \\ t_i - (a_i^T x + b_i) \end{matrix} \right\| \leq t_i + (a_i^T x + b_i) \quad \blacksquare$$

Трюк SOCP:

$$\|x\|_2 \leq yz \Rightarrow \left\| \begin{matrix} 2x \\ y-z \end{matrix} \right\| \leq y+z$$

* Всегда ищем где норма.
* Ее без квадрата умножаем на 2.
* То что больше нормы – разбиваем на минус и плюс.

Robust Linear Programming:

בע"ה קמורה:

$$\begin{cases} \min_x \; c^T x \\ \text{s.t } a_i^T x \leq b_i \;\; \forall a_i \in S_i \\ \text{where:} \\ S_i = \{\bar{a}_i + P_i u : \|u\|_2 \leq 1\} \end{cases}$$

- The constraint:

$$a_i^T x \leq b_i \quad \forall a_i \in S_i$$

Из всей группы берем максимум.
- is equivalent to

$$\max_{a_i \in S_i} a_i^T x \leq b_i$$

работает и в другую сторону.

подставляем
- or equivalent to

$$\max_{u : \|u\|_2 \leq 1} (\bar{a}_i + P_i u)^T x \leq b_i$$

Открываем скобки, и то что не относится к 'u' выносим.

- $\bar{a}_i^T x + \left\{ \max_{u : \|u\|_2 \leq 1} u^T P_i^T x \right\} \leq b_i$

$$\left\{ \max_{u:\, \|u\|_2 \leq 1} u^T P_i^T x \right\} b_i$$

$$\Downarrow$$

Using Cauchy Schwartz:

$$u^T P_i^T x \leq \|u\| \cdot \|P_i^T x\| \underset{\substack{\downarrow \\ u = u^*}}{=} \|P_i^T x\|$$

achievable by choosing

$$u = \frac{P_i^T x}{\|P_i^T x\|}$$

Thus, we get SOCP:

$$\min_x \quad c^T x$$

$$\text{s.t} \quad \bar{a}_i^T x + \|P_i^T x\| \leq b_i$$

■

XI. Semidefinite programming

SDP :

A higher level in the conic programming hierarchy: LP, SOCP and SDP. Basically an LP with a different definition of the inequalities.

$$\begin{array}{ll} \min_x & c^T x \\ \text{s.t.} & F_0 + \sum_i x_i F_i \succeq 0 \\ & Ax = b \end{array}$$

LMI = Linear matrix inequality

Multiple LMIs via a single block diagonal LMI.

LP is special case by choosing scalar F_0 and F_i.

SOCP is special case by noting that

$$\{x, t \ : \ \|x\|_2 \leq t\} = \left\{ x, t \ : \ \begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0 \right\}$$

Example: matrix fractional program can be cast as as SOCP and therefore also as SDP (which is much more natural yet more expensive)

$$\left\{ \begin{array}{ll} \min_x & (Ax+b)^T \left(I + B\text{diag}\{x\} B^T\right)^{-1} (Ax+b) \\ \text{s.t.} & x \geq 0 \end{array} \right.$$

Adding variables

$$\left\{ \begin{array}{ll} \min_{x,t} & t \\ \text{s.t.} & \underbrace{t}_{A} \geq \underbrace{(Ax+b)^T}_{B} \underbrace{\left(I + B\text{diag}\{x\} B^T\right)^{-1}}_{C^{-1}} \underbrace{(Ax+b)}_{B^T} \\ & x \geq 0 \end{array} \right.$$

Now using Schur's lemma : $(A - BC^{-1}B^T \geqslant 0) \Leftrightarrow \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \geqslant 0$

$$\left\{ \begin{array}{ll} \min_{x,t} & t \\ \text{s.t.} & \begin{bmatrix} t & (Ax+b)^T \\ (Ax+b) & \left(I + B\text{diag}\{x\} B^T\right) \end{bmatrix} \succeq 0 \\ & x \geq 0 \end{array} \right.$$

■

A. *Minimization of maximum eigenvalue (symmetric)*

$$A(X) = A_0 + \sum_i x_i A_i \qquad \text{symmetric}$$

We want to solve

$$\min_x \lambda_{\max}(A(x))$$

$$\begin{cases} \min_{x,t} & t \\ \text{s.t.} & \lambda_{\max}(A(x)) \leq t \end{cases}$$

переход на "для каждого"

$$\begin{cases} \min_{x,t} & t \\ \text{s.t.} & \lambda_i(A(x)) \leq t \ \forall\, i \end{cases}$$

$$\{\lambda_i(A) \leq t \ \forall\, i\} = \{A \leq tI\}$$

Ответ:

$$\boxed{\begin{cases} \min_{x,t} & t \\ \text{s.t.} & A(x) \preceq tI \end{cases}}$$

B. *Minimization of spread of eigenvalues (symmetric)*

$$A(X) = A_0 + \sum_i x_i A_i \qquad \text{symmetric}$$

We want to solve

$$\min_x \lambda_{\max}(A(x)) - \lambda_{\min}(A(x))$$

$$\begin{cases} \min_{x,t_1,t_2} & t_1 - t_2 \\ \text{s.t.} & \lambda_{\max}(A(x)) \leq t_1 \\ & \lambda_{\min}(A(x)) \geq t_2 \end{cases}$$

Ответ:

$$\begin{cases} \min_{x,t_1,t_2} & t_1 - t_2 \\ \text{s.t.} & t_2 I \preceq A(x) \preceq t_1 I \end{cases}$$

C. *Matrix norm minimization (non-symmetric)*

$$A(X) = A_0 + \sum_i x_i A_i \qquad \text{non symmetric}$$

We want to solve

$$\min_x \|A(x)\|_2$$

where $\|X\|_2$ is the maximum singular value of X, is the square root of the maximum eigenvalue of $X^T X$, and also the maximum eigenvalue if $X \succeq 0$. This is convex because it is a norm.

$$\min_x \quad \lambda_{\max}\left(A^T(x)A(x)\right)$$

$$\begin{cases} \min_{x,t} & t^2 \\ \text{s.t.} & \lambda_{\max}\left(A^T(x)A(x)\right) \leq t^2 \end{cases}$$

$$\begin{cases} \min_{x,t} & t \\ \text{s.t.} & A^T(x)A(x) \preceq t^2 I \end{cases}$$

determinanta

$$\{A^T A \preceq t^2 I\} \overset{\downarrow}{=} \left\{ A, t \;:\; \begin{bmatrix} tI & A \\ A^T & tI \end{bmatrix} \succeq 0 \right\}$$

$$\Downarrow$$

Ombem: $$\boxed{\begin{cases} \min_{x,t} & t \\ \text{s.t.} & \begin{bmatrix} tI & A(x) \\ A^T(x) & tI \end{bmatrix} \succeq 0 \end{cases}}$$

D. *Fastest mixing Markov chain on a graph*

Fastest mixing Markov chain on a graph - also related to fastest distributed computation of mean via consensus.

We consider an undirected graph $G = \{V, E\}$ with $|V| = n$. We allow self-loops, i.e., $(i, i) \in E$. We define a Markov chain with states V and transitions E with probabilities P_{ij} for $(i, j) \in E$.

$$
\begin{array}{rl}
P = P^T & \text{undirected and symmetric} \\
1^T P = 1^T & \text{probabilities sum to one} \\
P \geq 0 & \text{probabilities are not negative} \\
P_{ij} = 0 \ (i,j) \notin E & \text{transitions only on edges}
\end{array}
$$

The first two properties show that $\frac{1}{n}1$ is an equilibrium stationary distribution and this is the largest eigenvalue.

Convergence to the stationary distribution is determined by the second largest (in magnitude) eigenvalue, denoted by r. This can be either λ_2 or $-\lambda_n$ where

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Thus,

$$r = \max\{\lambda_2, -\lambda_r\}$$

When $r = 1$ convergences is not promised. When $r < 1$ the smaller r is, the faster the mixing. Thus, we want to solve

$$
\begin{array}{ll}
\min_P & \max\{\lambda_2, -\lambda_r\} \\
\text{s.t} & P = P^T \\
& P \geq 0 \\
& 1^T P = 1^T \\
& P_{ij} = 0 \ (i,j) \notin E
\end{array}
$$

The principal eigenvalue is λ_1 so we cannot use the spectral norm as is. We need to first remove the first eigenvalue. Its eigenvector is $\frac{1}{\sqrt{n}}1$ so we can project P on the orthogonal complement

$$[1]^\dagger = \frac{1^T}{1^T 1}$$

and the projection operator is

$$Q = I - \frac{11^T}{1^T 1}$$

so that

$$\begin{aligned} r &= \|QPQ\|_2 \\ &= \|P - \frac{11^T}{1^T 1}\|_2 \end{aligned}$$

and we get

$$\begin{array}{ll} \min_P & \|P - \frac{11^T}{1^T1}\|_2 \\ \text{s.t} & P = P^T \\ & P \geq 0 \\ & 1^T P = 1^T \\ & P_{ij} = 0 \; (i,j) \notin E \end{array}$$

$$\begin{array}{ll} \min_{P,t} & t \\ \text{s.t} & |\lambda_i \left(P - \frac{11^T}{1^T1} \right)| \leq t \quad \forall\, i \\ & P = P^T \\ & P \geq 0 \\ & 1^T P = 1^T \\ & P_{ij} = 0 \; (i,j) \notin E \end{array}$$

$$\begin{array}{ll} \min_{P,t} & t \\ \text{s.t} & -tI \preceq P - \frac{11^T}{1^T1} \preceq tI \\ & P = P^T \\ & P \geq 0 \\ & 1^T P = 1^T \\ & P_{ij} = 0 \; (i,j) \notin E \end{array}$$

Duality:

1. The Lagrange dual function

Consider the following problem:

$$p^* = \begin{cases} \min\limits_x f_0(x) \\ \text{s.t} \quad f_j(x) \le 0, \ j = 1 \dots m \\ \qquad h_j(x) = 0, \ j = 1 \dots p \end{cases}$$

Define the Lagrangian:

$$L(x; \lambda, \mu) = f_0(x) + \sum_{j=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

וקטור באורך m של כופלי לגרנג' או פרמטרים דואליים

באורך p

- Множитель Лагранжа: созданы для того чтобы решать задачу с шуцим.

And the dual function:

$$g(\lambda, \mu) = \inf_x L(x; \lambda, \mu)$$

This may be hard to solve (in closed form or in non-convex) and where the art comes in by adding more constrains.

משפטים, הגדרות

Weak duality :משפט

$$g(\lambda,\mu) \leq p^* \text{ for any } \lambda \geq 0 \text{ and } \mu.$$

Мы же ведь хотим подняться.
Так тут мы говорим о том что мы будем оптимальными или меньше.

Всегда спрашивает на экзамене!

:הוכחה

$$p^* \underset{\text{הגדרה}}{=} f_0(x^*) \geq f_0(x^*) + \sum \underbrace{\lambda_i}_{\geq 0 \text{ לפי המשפט}} \underbrace{f_i(x^*)}_{\leq 0} + \sum \mu_i \underbrace{h_i(x^*)}_{= 0 \text{ x* פיזבלי}} =$$

מכיון ש x^* הוא פיזבלי (מקיים s.t).

(Еще пошутил что все забывают, что פיזי = פיזלי . חיובי)

$$\underset{\text{הגדרה}}{=} L(x^*;\lambda,\mu) \geq \inf_x L(x;\lambda,\mu) = g(\lambda,\mu)$$

■

Notations: We say that λ and μ are dual feasible if $\lambda \geq 0$ and $g(\lambda,\mu) > -\infty$.

The maximal lower bound

$$g(\lambda^*,\mu^*) = d^* = \max_{\mu,\lambda \geq 0} g(\lambda,\mu)$$

is called the dual program (and the original is primal).

↓

תכונות של הבעיה הדואלית

① כאשר L לא חסום : $g(\lambda,\mu) = -\infty$.

② הבעיה הדואלית היא קמורה תמיד!

הוכחה $\Rightarrow g(\lambda,\mu) = \inf_x L(x;\lambda,\mu) =$

$$= \inf_x f_0(x) + \sum \lambda_j f_j(x) + \sum \mu_j h_j(x)$$

In addition we know that min concave is concave, thus the written above is always concave

He said:" L has 2 types of parameters primal and dual.

- For every x, $L(x;\lambda,\mu)$ is concave in $\begin{bmatrix}\lambda\\ \mu\end{bmatrix}$.
- That's why $\min_x$ is also concave in $\begin{bmatrix}\lambda\\ \mu\end{bmatrix}$.

■

Как перевести проблему и найти дуальную?

① Дана задача: Linear programming:

$$\min_x \; c^T x$$
$$\text{s.t } Ax = b$$
$$x \geq 0$$

② Мы должны найти правильные λ и μ.
По определению μ идет там где равенство.
В то време как λ там где функция <u>меньше</u> 0.

Это значит что μ подходит для $Ax=b$.
А λ пока не подходит. Поэтому меняем
$-x \leq 0$.

Получается: $L(x;\lambda,\mu) =$

$= f_0(x) + \sum^{m} f_i \lambda_i + \sum h_i \mu_i =$

$= c^T x - \lambda^T x + \mu^T (Ax-b) =$

$= c^T x + \sum \lambda_i (-x_i) + \sum \mu_i (a_i^T x - b_i)$

• $g(\lambda,\mu) = \inf L(x;\lambda,\mu) =$

$$= \inf_x \underbrace{[c + A^T \mu - \lambda]}_{\alpha}{}^T x - \mu^T b.$$

вынесли x

Трюк: $\min_x \alpha x = \begin{cases} 0 & \alpha = 0 \\ -\infty & \alpha \neq 0. \end{cases}$

результат, а не сам x!

Поздравляю! Твой первый трюк в дуальной задаче!

• $g(\lambda,\mu) = \begin{cases} -\mu^T b & c + A^T \mu - \lambda = 0 \\ -\infty & \text{иначе} \end{cases}$

③ В конце, мы получаем дуальную задачу которая выглядит таким образом:

$$\begin{cases} \max\limits_{\mu,\lambda\geq 0} & -\mu^T b \\ \text{s.t} & c + A^T\mu - \lambda = 0 \end{cases} \blacksquare$$

Strong Duality : משפט

If the problem is convex and Slater's condition (constraint qualification) holds, i.e there exists a strictly feasible point, then strong duality holds and $d^* = p^*$

הוכחה:

We will prove $p^* \leq g(\lambda^*)$ because last lesson we saw weak duality ($p^* \geq g(\lambda^*)$) thus we will prove the equation.

① $g(\lambda^*) \geq p^*$

② $\inf\limits_x f_0(x) + \sum \lambda_i^* f_i(x) \geq p^*$

③ goal:

$f_0(x) + \sum \lambda_i^* f_i(x) \geq p^* \quad \forall x$

if inf is bigger than p^*, so other answers definately.

$\downarrow$

④ For this purpose, we define two convex sets

$$A = \{(u,t) \mid \exists x \ f_i(x) \le u_i \ , \ f_0(x) \le t\}$$

וקטור $m\times1$.

$$B = \{(0,t) \mid t < p^*\}$$

⑤ The set A captures all the solutions for the optimization plus the points which are worse than them, in fact we have.

$$p^* = \min t \quad \text{s.t } (0,t) \in A$$

⑥ A and B don't intersect, and

לפי משפט ההפרדה קיימים $(\tilde{\lambda},\mu) \neq 0$ ו-α כך ש:

$$(u,t) \in A \rightarrow \tilde{\lambda}^T \cdot u + \mu t \ge \alpha$$

$$(u,t) \in B \rightarrow \tilde{\lambda}^T \cdot u + \mu t \le \alpha$$

⑦ If $(u,t) \in A$, then we can add arbitrary positive numbers to them and still be in A. Together with the first property this means $\tilde{\lambda} \ge 0$ and $\mu \ge 0$ (otherwise $\lambda^T u + \mu t$) is unbounded from below.
From the second we have: $\mu t \le \alpha$ & $t < p^*$
Or more simply: $\mu p^* \le \alpha$.

Together: $\mu f_0(x) + \sum_i \tilde{\lambda}_i f_i(x) \ge \alpha \ge \mu p^* \quad \forall x.$

$\downarrow$

נחלק את אי שוויון ב- μ ונגדיר: $\lambda_i = \frac{\tilde{\lambda}_i}{\mu}$

$$f_0(x) + \sum \lambda_i f_i(x) \geq p^* \quad \forall x$$

$\mu \neq 0$ מכיוון שלפי סלייטר: $f_i(\tilde{x}) < 0 \Leftarrow$

ולכן: $0 + \sum \tilde{\lambda}_i f_i(\tilde{x}) \geq 0$

$$\tilde{\lambda}_i = 0$$

■

Saddle point Interpretation:

$$p^* = \min_{x} \max_{\lambda \geq 0, \mu} L(x; \lambda, \mu)$$

$$d^* = \max_{\lambda \geq 0, \mu} \min_{x} L(x; \lambda, \mu)$$

- The first inner maximization is exactly f_0 with the domain defined as the feasible set.

$$\max_{\lambda \geq 0, \mu} L(x; \lambda, \mu) = \begin{cases} 0 & x \text{ is feasible} \\ \infty & \text{otherwise} \end{cases}$$

The second inner minimization is exactly the definition of the dual problem

In general: max min $\leq$ min max
And that's the weak duality. Strong duality is the other way around and its solution is a saddle point:

$$L(x^*; \lambda, \mu) \leq L(x^*; \lambda^*, \mu^*) \leq L(x; \lambda^*; \mu^*)$$

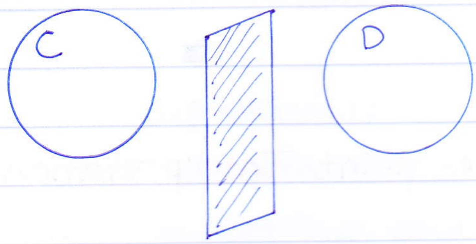
$\forall x \geq 0, \mu$
$\forall \lambda \geq 0, \mu.$

"Saddle point говорит что max min и также min max это одно и то-же".

משפט ההפרדה:

יהיו C ו D קבוצות קמורות וזרות.
אז קיים $a \neq 0$ ו b כך ש:

$$x^T a \leq b \quad \forall x \in C$$
$$x^T a \geq b \quad \forall x \in D$$



$$f(x) = x^T a - b$$

על מישור.

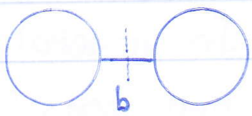
הוכחה:

① We only prove a special case where there exist two points $c \in C$ and $d \in D$ that are closest to each other in terms of Euclidean norm i.e.:

$$\|c-d\| = \inf_{\substack{c \in C \\ d \in D}} \|x-y\|$$

② In this case we define

$$a = d-c \quad , \quad b = \frac{\|d\|^2 - \|c^2\|}{2}$$



And show that these define a seperating hyperplane.
Specifically, let us show that $a^T x - b \geq 0$ on D

③ Our heperplane is :

$$f(x) = a^T x - b = (d-c)^T x - \frac{\|d\|^2 - \|c\|^2}{2} =$$

$$= (d-c)^T (x-d) + \frac{1}{2}\|d - c\|^2$$

④ נניח בשלילה שקיים $u \in D$ עבורו $f(u) < 0$

(כאשר מה שנכון $x^T a - b \geq 0$).(בגלל ששייך ל-D).

ואז מתקיים:

$$(d-c)^T (u-d) < 0$$

אבל אם נעשה צעד קטן מ-d לכיוון u נראה שנקטין (צעד: $t(u-d)$)

את המרחק שלו מנקודה מכיוון שבחרנו את d להיות

כנקודה הכי קרובה.

נראה זאת ע"י נגזרות:

$$\frac{d}{dt} \left\| d + t(u-d) - c \right\|^2 \Big|_{t=0} =$$

$$= 2(d + t(u-d) - c)^T (u-d) \underset{t=0}{=} 2(d-c)^T (u-d) = 0.$$

■

Farkas Lemma:

This is a theory of alternatives - the logic behind duality.
Basically gives us certificates for impossibility results.
Much stronger than showing possibility which is just
by showing an example. S
Similarly, to the fact that lower bounds to minimization
problems are stronger than upper bounds.

Theorem:

Consider the two sets of inequalities

I) $Ax \leq 0$, $c^T x < 0$

II) $A^T y + c = 0$, $y \geq 0$.

Then, only one of them holds, i.e each system is feasible if and only if the other is not.

Proof:

(P) $\min c^T x$ s.t $Ax \leq 0$

(D) $\max 0$ s.t $A^T y + c = 0$, $y \geq 0$.

(P) is homogenous, therefore

$p^* = -\infty$ if (I) is feasible.

$p^* = 0$ if it isn't feasible.

(D) is also either $d^* = 0$ if (II) is feasible

or $d^* = -\infty$ if it isn't feasible.

Due to $p^* = d^*$, feasible (I) means infeasible (II)
and vice verse.

→ Теперь, с этой страницы, мы будем рассматривать множество примеров. Некоторые из них будут с лекций, а некоторое из д/з. Возле каждой будет отмечено **откуда** она. ←

E-optimal experiment design : Ex 3

Derive the dual program of:

$$\begin{cases} \min\limits_{x} & \operatorname{eig}_{max}\left[\left(\sum\limits_{i=1}^{p} x_i a_i a_i^T\right)^{-1}\right] \\ \text{s.t} & x_i \geq 0 \quad \forall i \\ & \sum\limits_{i=1}^{p} x_i = 1. \end{cases}$$

Solution:

1. We use: $\operatorname{eig}_{max}(A^{-1}) = \dfrac{1}{\operatorname{eig}_{min}(A)}$

2. Equivalent to:

$$\min_{x, t \geq 0} \quad \frac{1}{t}$$

([illegible]) $\quad \text{s.t} \quad \sum\limits_{i=1}^{p} x_i a_i a_i^T \succeq tI$

$$x_i \geq 0 \quad \forall i$$

$$\sum_{i=1}^{p} x_i = 1.$$

↓

③ The lagrangian is:

$$L = \frac{1}{t} - \mathrm{Tr}\left\{Z\left(\sum x_i a_i a_i^T - tI\right)\right\} - \sum \lambda_i x_i + \mu \sum x_i - \mu$$

④ Minimizing over $t \geq 0$ yields:

$$\min_{t \geq 0} \frac{1}{t} + t\,\mathrm{Tr}\{Z\} = \begin{cases} 2\sqrt{\mathrm{Tr}\{Z\}} & Z \geq 0 \\ -\infty & \text{else} \end{cases}$$

⑤ Minimizing over x yields $-\infty$ unless:

$$-a_i^T Z a_i - \lambda_i + \mu = 0$$

Noting that $\lambda_i \geq 0$ we get:

$$\begin{cases} \max\limits_{Z,\mu} & 2\sqrt{\mathrm{Tr}\{Z\}} - \mu \\ \text{s.t} & a_i^T Z a_i \leq \mu \quad \forall i \\ & Z \geq 0 \end{cases}$$

⑥ The problem can be further simplified by defining $W = \frac{1}{\mu} Z$ where $\mu \geq 0$ due to the constraints. Solving for μ yields a simple SDP

$$\begin{cases} \max\limits_{W} & \mathrm{Tr}\{W\} \\ \text{s.t} & a_i^T W a_i \leq 1 \quad \forall i \\ & W \geq 0. \end{cases}$$

∎

Boolean Linear Programming: Ex 3

$$\min_x \; c^T x$$
$$Ax \leq b$$
$$x_i \in \{0,1\}$$

→ Prove that the linear relaxation using $0 \leq x_i \leq 1$ and the dual relaxation using constraints $x_i(1-x_i) = 0$, are equivalent

① Linear Relaxation is:

$$\min_x \; c^T x$$
$$\text{s.t} \quad Ax \leq b$$
$$0 \leq x_i \leq 1.$$

② The lagrangian is:

$$L = c^T x + y^T(Ax-b) + u^T(x-1) - v^T x$$
$$L = c^T x + y^T A x - y^T b + u^T x - u^T - v^T x$$

③ $g(\lambda,\mu) = \inf L(x; \lambda, \mu) =$

$$= \inf_x \; (c^T + y^T A - v^T + u^T)x - y^T b - u^T 1$$

④ and the dual:

$$\begin{cases} \max\limits_{y \geq 0 \; v \geq 0 \; u \geq 0} & -y^T b - u^T 1 \\ \text{s.t} & c + A^T y + u \geq 0 \end{cases}$$

⑤ or just:

$$\begin{cases} \max\limits_{y \geq 0 \; u \geq 0} & -y^T b - u^T 1 \\ \text{s.t} & c + A^T y + u \geq 0 \end{cases}$$

↓

(6) On the other hand, the dual relaxation is:

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{s.t} \quad & Ax \leq b \\ & x_i - x_i^2 = 0 \end{aligned}$$

(7) The Lagrangian is:

$$L = c^T x + y^T A x - y^T b - \sum_i \lambda_i x_i + \sum_i \lambda_i x_i^2$$

(8)

minimizing over x gives the dual function

$$g = \begin{cases} -y^T b - \frac{1}{4} \sum_i \frac{(c_i + a_i^T y - \lambda_i)^2}{\lambda_i} & \lambda \geq 0 \\ -\infty & \text{else} \end{cases}$$

(9) We need to maximize this function with respect to the dual variables.

$$\sup_{\lambda_i \geq 0} - \frac{(q - \lambda_i)^2}{4\lambda_i} = \begin{cases} q & q \leq 0 \\ 0 & \text{else} \end{cases} = \min\{0, q\}$$

(10) So the dual program is simplified to:

$$\begin{cases} \max_{y} & -y^T b + \sum_i \min\{0, c_i + a_i^T y\} \\ \text{s.t} & y \geq 0. \end{cases}$$

(11) That can be written as

$$\begin{cases} \max_{y,u} & -y^T b - 1^T u \\ \text{s.t} & y \geq 0 \\ & u \geq 0 \\ & c + A^T y + u \geq 0. \quad \blacksquare \end{cases}$$

Projection on Simplex : Ex 3

Derive the dual of the following problem.
Use only one Lagrange multiplier, so that the dual will be a simple line search.

$$\min_{x} \|y - x\|^2$$
$$\text{s.t} \quad x \geq 0$$
$$1^T x = 1.$$

① Now the problematic (non-separable) constraints is $1^T x = 1$ so we put a Lagrange multiplier on it

$$\max_{\mu} \min_{x \geq 0} \|y - x\|^2 + \mu - \mu 1^T x$$

which can also be written as:

$$\max_{\mu} \mu + \sum_i f_i(y_i; \mu)$$

where

$$f_i(y_i; \mu) = \min_{x_i \geq 0} |y_i - x_i|^2 - \mu x_i$$

has a simple closed solution:

$$x_i = \begin{cases} y_i + \frac{\mu}{2} & \text{if } y_i + \frac{\mu}{2} \geq 0 \\ 0 & \text{else} \end{cases}$$

■

LP with box constraints: Notebook

$$\begin{cases} \min_x \ c^T x \\ \text{s.t} \ \ Ax = b \\ \qquad -1 \leq x \leq 1 \end{cases}$$

$$\Downarrow$$

$$\begin{cases} \min_x \ c^T x \\ \text{s.t} \ \ Ax = b \ \Rightarrow \ Ax - b = 0 \\ \qquad x \leq 1 \ \Rightarrow \ x - 1 \leq 0 \\ \qquad x \geq -1 \ \Rightarrow \ -1 - x \leq 0 \end{cases}$$

$$L = c^T x + \lambda^T (Ax - b) + \mu^T (x - 1) + z^T(-1 - x).$$

- Однако это нам даст не очень красивую дуальную проблему, поэтому можно записать примальную задачу несколько иначе:

$$\begin{cases} \min \ c^T x \\ x : \{x : -1 \leq x \leq 1\} \\ \ \text{s.t} \ \ Ax = b \end{cases}$$

$$L(x; \lambda) = c^T x + \lambda (Ax - b)$$

$$g(\lambda) = \min_{-1 \leq x \leq 1} c^T x + \lambda^T (Ax - b)$$

Трюк: $-|a| = \min_{-1 \leq x \leq 1} ax$

$$\Downarrow$$

$$\min_{-1 \leq x \leq 1} -\lambda^T b + x^T (A^T \lambda + c)$$

$$g(\lambda) = -\lambda^T b - \| A^T \lambda + c \|_1$$

And so we get the dual problem:

$$\min \ c^T x \ \ \text{s.t} \ \ Ax = b \ , \ \|x\|_\infty = 1. \ \blacksquare$$

Minimum norm solution: Notebook

$$\min_{x} \quad x^T x$$
$$\text{s.t} \quad Ax = b$$

$$L(x;\lambda) = L = x^T x + \lambda^T (Ax - b)$$

$$g(\lambda) = \min_{x} \; x^T x + \lambda^T A x - \lambda^T b$$

$$g(\lambda) = \min_{x} \; x^T x + x^T A^T \lambda - b^T \lambda$$

$$\frac{\partial L(x;\lambda)}{x} = 2x + A^T \lambda = 0$$

$$x = -\frac{1}{2} A^T \lambda$$

Plugging in this solution yields:

$$g(\lambda) = L\left(-\tfrac{1}{2} A^T \lambda ; \lambda\right) =$$

$$= x^T x + \lambda^T (Ax - b) = -\frac{1}{4} \lambda^T A A^T \lambda - b^T \lambda$$

And the dual problem:

$$\max_{\lambda} \; -\frac{1}{4} \lambda^T A A^T \lambda - b^T \lambda \quad \blacksquare$$

$-|X|^{1/n}$ function : Ex1

Prove that the function $-|X|^{1/n}$ is convex on the set of $n \times n$ positive definite matrices.

הוכחה:

① We will prove that $\det^{1/n}(X)$ is concave on $\{X \in S^n : X > 0\}$ by restricting it on a line.
This means than negation of the function is convex.

② The line : $g(t) = f(Z + tV)$
where $Z > 0$ and $V \in S^n$.
We need to show that $g(t)$ is concave in t for all $t \in \{t : Z + tV > 0\}$

③ $$g(t) = \det^{1/n}(Z + tV) =$$

$$= \det^{\frac{1}{n}}(Z) \cdot \det^{\frac{1}{n}}\left(I + tZ^{-\frac{1}{2}} V Z^{-\frac{1}{2}}\right) =$$

$$= \det^{\frac{1}{n}}(Z) \cdot \left(\prod 1 + t\lambda_i\right)^{1/n}$$

④ Affine transformations preserve concavity and the geometric mean is concave. → this we need to prove:

↓

⑤ Geometric mean $f(x) = (\prod_i x_i)^{\frac{1}{n}}$ is concave on $\mathbb{R}^n_{++}$.

The Hessian is:

$$\nabla^2 f(x) = \frac{-\prod_{i=1}^n x_n^{1/n}}{n^2} \cdot \left(\begin{bmatrix} n \cdot \frac{1}{x_1^2} & & 0 \\ & \ddots & \\ 0 & & n \cdot \frac{1}{x_n^2} \end{bmatrix} - \begin{bmatrix} \frac{1}{x_1} \\ \vdots \\ \frac{1}{x_n} \end{bmatrix} \begin{bmatrix} \frac{1}{x_1} & \cdots & \frac{1}{x_n} \end{bmatrix} \right)$$

To show concavity, we prove that the Hessian is negative semidefinite:

$$v^T \nabla^2 f(x) v = \frac{-\prod_{i=1}^n x_n^{1/n}}{n^2} \left(n \cdot \sum_i \frac{v_i^2}{x_i^2} - \left(\sum_i \frac{v_i}{x_i} \right)^2 \right) =$$

$$= -\frac{\prod_{i=1}^n x_n^{1/n}}{n^2} \left(\left(\sum_i a_i^2 \right) \left(\sum_i b_i^2 \right) - \left(\sum_i a_i b_i \right)^2 \right) \leq 0$$

Where we define $a_i = 1$ and $b_i = \frac{v_i}{x_i}$ and the inequality holds due to Cauchy Schwartz.

■

Problem no.4 : Ex 1

Consider the set of points whose distance to a point 'a' does not exceed a fixed fraction 'θ' of the distance to another point b, i.e :

$$\{ x \mid \|x-a\|_2 \leq \theta \|x-b\|_2 \}$$

for some 'a' and 'b'. Is it convex?

הוכחה:

$$\|x-a\|_2 \leq \theta \|x-b\|_2 \quad \uparrow^2$$

$$\|x-a\|_2^2 \leq \theta^2 \|x-b\|_2^2$$

$$x^Tx - a^Tx - x^Ta - a^Ta - \theta^2(x^Tx - b^Tx - x^Tb - b^Tb) \leq 0$$

$$x^Tx - a^Tx - x^Ta - a^Ta - \theta^2 x^Tx + \theta^2 b^Tx + \theta^2 x^Tb + \theta^2 b^Tb \leq 0$$

$$(1-\theta^2)x^Tx + a^Ta - x(a^T - \theta^2 b^T) - x^T(a-\theta^2 b) - \theta^2 b^Tb \leq 0$$

$$\Rightarrow (1-\theta^2)x^Tx + a^Ta - x(a-\theta^2 b)^T - x(a-\theta^2 b)^T - \theta^2 b^Tb \leq 0$$

$$(1-\theta^2)x^Tx + a^Ta - 2x(a-\theta^2 b)^T - \theta^2 b^Tb \leq 0 \quad /: (1-\theta^2)$$

$$x^Tx + \frac{a^Ta}{1-\theta^2} - \frac{2x(a-\theta^2 b)^T}{1-\theta^2} - \frac{\theta^2 b^Tb}{1-\theta^2} \leq 0$$

$$x^Tx - \frac{2x(a-\theta^2 b)^T}{1-\theta^2} \leq \frac{\theta^2 b^Tb}{1-\theta^2} - \frac{a^Ta}{1-\theta^2} \quad /+ c^Tc$$

$$\begin{cases} x^Tx - \frac{2x(a-\theta^2 b)^T}{1-\theta^2} + c^Tc \leq \frac{\theta^2 b^Tb - a^Ta}{1-\theta^2} + c^Tc \\ c = \frac{a-\theta^2 b}{1-\theta^2} \end{cases}$$

$$\to \|x-c\|_2^2 \leq \frac{\theta^2 b^Tb - a^Ta}{1-\theta^2} + c^Tc \quad /\sqrt{}$$

$$\begin{cases} \|x-c\|_2 \leq R & \to \text{this is ball, and it's convex.} \blacksquare \\ R = \sqrt{\frac{\theta^2 b^Tb - a^Ta}{1-\theta^2} + c^Tc} \end{cases}$$