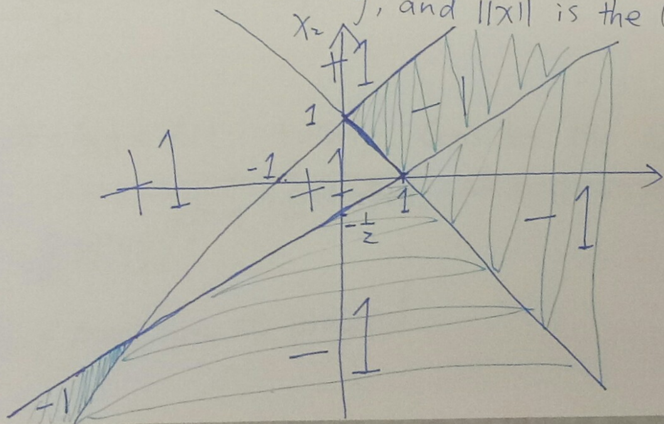
1. • For $H_1 = \{x \mid a^T x \le b\} \subseteq H_2 = \{x \mid \tilde{a}^T x \le \tilde{b}\}$, we need those two hyperplanes $a^T x = b$ and $\tilde{a}^T x = b$ to be "parallel" intuitively. This means there exists $\lambda > 0$ such that $a = \lambda \tilde{a}$. Besides, we need hyperplane $a^T x = b$ to be "lower" than $\tilde{a}^T x = \tilde{b}$ so that H_1 is contained in H_2. This translates to $\frac{b}{\lambda} \le \tilde{b}$, or $b \le \lambda \tilde{b}$. Hence, the condition can be summarized as $\exists \lambda > 0$ s.t. $\begin{cases} a = \lambda \tilde{a}, \text{ and} \\ b \le \lambda \tilde{b} \end{cases}$

   • When two half spaces are equal, it implies $H_1 \subseteq H_2$ and $H_2 \subseteq H_1$. Hence, the condition is $\exists \lambda > 0$, s.t. $\begin{cases} a = \lambda \tilde{a}, \text{ and} \\ b = \lambda \tilde{b} \end{cases}$.

2. Let $x_1 \in \mathbb{R}^n$ and $x_2 \in \mathbb{R}^n$ be two points on hyperplane $a^T x = b_1$ and $a^T x = b_2$, respectively. In other words, $\begin{cases} a^T x_1 = b_1 \\ a^T x_2 = b_2 \end{cases}$. Now the distance b/t those two parallel hyperplane can be computed by $\left| \frac{\langle x_1 - x_2, a \rangle}{\|a\|} \right| = \left| \frac{a^T (x_1 - x_2)}{\|a\|} \right| = \frac{|b_1 - b_2|}{\|a\|}$, where $\langle x, y \rangle$ is the dot product b/t x and y, and $\|x\|$ is the (L_2) norm of vector x.

3.



area in this color means areas of x_1, x_2 that will be assigned -1

Other areas will be assigned $+1$.

5.1(a) $\min \|Ax-b\|_1$
s.t. $\|x\|_\infty \le 1$

1). Remove absolute value in the constraints by introducing and ~~to~~ restricting a new variable

$$\begin{cases} t \ge |x_i|, \forall i=1,\dots,n \\ t \le 1 \\ t \ge 0 \end{cases} \Rightarrow \begin{cases} -t \le x_i \le t, \forall i=1,\dots,n \\ t \le 1 \\ t \ge 0 \end{cases}$$

You can see that the possible values for $x \in R^n$ is exactly the same as in the orignial constraints;

2) Remove absolute value in the objective by introducing and restricting ~~a~~ new variable k_i $\forall i=1,\dots,m$ so that $k_i \ge |a_i^T x - b_i|$, $\forall i=1,\dots,m$. This adds to the constraints as $\begin{cases} -k_i \le a_i^T x - b_i \le k_i \\ k_i \ge 0 \end{cases}$ $\forall i=1,\dots,m$ and allows us to rewrite the objective as $\min \sum_{i=1}^m k_i$. The equivalence b/t this new objective and the orignal objective can be seen by this: fix a $x_0 \in R^n$ in the constraints, each k_i is lower-bounded by $|a_i^T x_0 - b_i|$, so, minimizing $\sum_{i=1}^m k_i$ is equal to the original objective evaluated at x_0 (i.e. $\|Ax_0 - b\|_1$). Hence, if we try out each and every possible value of x in the constraints, we can find the minimal ~~$a_i^T x - b_i$~~ $\|Ax-b\|_1$.

$$\min_{\substack{x \in R^n \\ t \in R \\ k \in R^m}} \sum_{i=1}^m k_i$$

$$\begin{aligned} \text{s.t. } & -t \le x_i \le t, \forall i=1,\dots,n \\ & 0 \le t \le 1 \\ & -k_i \le a_i^T x - b_i \le k_i, \forall i=1,\dots,m \\ & 0 \le k_i, \quad \forall i=1,\dots,m \end{aligned}$$

(b). see next page

continued: 5. (b). Similarly to (a), we just need to rewrite both the constraints and objective by introducing additional variables. ~~(as well as additional constraints)~~

$$\begin{array}{l} \min \|x\|_1 \\ \text{s.t. } \|Ax-b\|_\infty \le 1 \end{array} \xrightarrow[\text{let } t \ge |a_i^T x - b_i|,\ \forall i=1,\dots,m]{\text{let } k_i \ge |x_i|,\ \forall i=1,\dots,n} \begin{array}{l} \min\limits_{\substack{x\in R^n,\\ t\in R,\\ k\in R^n}} \sum_{i=1}^{n} k_i \\ \text{s.t. } -t \le a_i^T x - b_i \le t,\ \forall i=1,\dots,m \\ \quad 0 \le t \le 1 \\ \quad -k_i \le x_i \le k_i,\ \forall i=1,\dots,n \\ \quad 0 \le k_i \end{array}$$

(c). $$\min (\|Ax-b\|_1 + \|x\|_\infty) \xrightarrow[\text{let } t \ge |x_i|,\ \forall i=1,\dots,n]{\text{let } k_i \ge |a_i^T x - b_i|,\ \forall i=1,\dots,m} \begin{array}{l} \min\limits_{\substack{k\in R^m,\\ t\in R}} \left(\sum_{i=1}^{m} k_i + t\right) \\ \text{s.t. } -k_i \le a_i^T x - b_i \le k_i,\ \forall i=1,\dots,m \\ \quad 0 \le k_i,\ \forall i=1,\dots,m \\ \quad -t \le x_i \le t,\ \forall i=1,\dots,n \\ \quad 0 \le t \end{array}$$

~~s.t.~~

6. (a) $$\min \sum_{i=1}^{m} \max\{0, a_i^T x + b_i\} \xrightarrow[\text{let } t_i \ge a_i^T x + b_i,\ \forall i=1,\dots,m]{t_i \ge 0} \begin{array}{l} \min \sum_{i=1}^{m} t_i \\ \text{s.t. } t_i \ge a_i^T x + b_i,\ \forall i=1,\dots,m \\ \quad t_i \ge 0,\ \forall i=1,\dots,m \end{array}$$

(b). Let $A = A_0 + x_1 A_1 + \cdots + x_p A_p$. Consider the objective function itself first:

$\max\limits_{\|y\|_1=1} \|Ay\|_1$ ~~$= \max\limits_{\|y\|_1=1} \|\sum_{i=1}^{n} a_i^T y_i\|_1$, where a_i is the i^{th} column of A~~. can be written as

$\min\limits_{\|y\|_1=1} -\sum_{i=1}^{n} |a_i^T y| \dots\dots$

I don't know how to proceed.

7. $\min\ r$

s.t. $r \geq t_k, \forall k=1,\dots,p$

$r \geq 0$

$t_k \geq l_{ki}, \forall k=1,\dots,p\ \forall i=1,\dots,n$

$t_k \geq 0, \forall k=1,\dots,p$

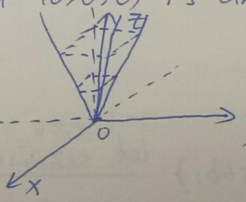
$l_{ki} \geq \sum_{j=1}^{n} S_{kij}, \forall k=1,\dots,p\ \forall i=1,\dots,n$.

$l_{ki} \geq 0, \forall k=1,\dots,p, \forall i=1,\dots,n$

$S_{kij} \leq (I - A_k X)_{ij} \leq S_{kij}, \forall k=1,\dots,p\ \forall i=1,\dots,n, \forall j=1,\dots,n$

$S_{kij} \geq 0, \forall k=1,\dots,p, \forall i=1,\dots,n, \forall j=1,\dots,n$

8. Yes, intuitively, you can easily see that $(0,0,0)$ is an extreme point from my (bad) drawing of the polyhedron:



Formally, we can first rewrite this polyhedron as $\{x \in R^3 \mid Ax \leq b\}$ where $A = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 0 & -1 \\ 0 & -1 & -1 \\ 0 & 1 & -1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$.

Then we can easily see that the first three rows of ~~row~~ A are linearly independent. Hence, this polyhedron has an extreme point.

9. In class, we've proven the equivalence of the first definition of extreme point (minimal face of a pointed polyhedron) and $\operatorname{rank}\left(\begin{bmatrix} A_{J(x)} \\ C \end{bmatrix}\right) = n$ (rank test), so, now I am going to prove the equivalence of the $\operatorname{rank}\left(\begin{bmatrix} A_{J(x)} \\ C \end{bmatrix}\right)=n$ and the second definition (no convex combination).

$$\operatorname{rank}\left(\begin{bmatrix} A_{J(x)} \\ C \end{bmatrix}\right) = n \iff \begin{array}{l} x \text{ cannot be written as convex} \\ \text{combination of any other two points} \\ y, z \text{ in } P \end{array}$$

1). $\Longrightarrow$ direction: (prove by contradiction)

Assume $x = \lambda y + (1-\lambda) z$, for some $y, z \in P$, $\lambda \in [0,1]$. Then $a_i^T x = b_i$, $\forall i \in J(x)$ can be written as $a_i^T(\lambda y + (1-\lambda) z) = \lambda a_i^T y + (1-\lambda) a_i^T z = b_i$, $\forall i \in J(x)$. We also know that $\begin{array}{l} a_i^T y \le b_i \\ a_i^T z \le b_i \end{array}$ $\forall i \in J(x)$ since $y, z \in P$. Hence, we have $\begin{cases} a_i^T y = b_i \\ a_i^T z = b_i \end{cases}$, $\forall i \in J(x)$.

Consider then the vector $y - z \neq 0$. We have $\begin{cases} a_i^T (y-z) = 0, \forall i \in J(x) \\ C(y-z) = 0 \end{cases} \Rightarrow \begin{bmatrix} A_{J(x)} \\ C \end{bmatrix}(y-z) = 0$.

This means the nullspace of $\begin{bmatrix} A_{J(x)} \\ C \end{bmatrix} \neq \{0\}$, so, the rank of it is less than n. $\lightning$

2). $\Longleftarrow$ direction: (prove by contradiction)

If $\operatorname{rank}\left(\begin{bmatrix} A_{J(x)} \\ C \end{bmatrix}\right) < n$, then $\exists v$ s.t. $\begin{cases} a_i^T v = 0, \forall i = J(x) \\ Cv = 0 \end{cases}$. Recall that for x, we have $\begin{cases} a_i^T x = b_i, \forall i \in J(x) \\ a_i^T x < b_i, \forall i \notin J(x) \\ Cx = d \end{cases}$. Then I can alway find two points $x + \lambda v$, $x - \lambda v$ for sufficiently small λ s.t. both points are in P. We can then write $x = \frac{1}{2}(x + \lambda v) + \frac{1}{2}(x - \lambda v)$, a convex combination of another two points in P. $\lightning$

Hence, the two statements are equivalent.

10. ~~(3) $\Longrightarrow$ (1) : If we have a basic feasible solution v, it means, by definition that $\begin{bmatrix} A_{J(v)} \\ c \end{bmatrix}$ contains n linearly independent vectors, which mean $\operatorname{rank}\left(\begin{bmatrix} A_{J(v)} \\ c \end{bmatrix}\right) = n$. I just proved in problem 9 (the previous one) that $\operatorname{rank}\left(\begin{bmatrix} A_{J(v)} \\ c \end{bmatrix}\right) = n$ implies v cannot be written as a convex combination of any other two points in P. Hence, v is an extreme point by definition.~~

~~(1) $\Longrightarrow$ (2) :~~

(1) $\Longrightarrow$ (3) : I just proved this in the last problem. That is, if v is an extreme point, then $\operatorname{rank}\left(\begin{bmatrix} A_{J(v)} \\ c \end{bmatrix}\right) = n$, which means there are n linearly independent rows, which, by definition means v is a basic feasible solution.

(3) $\Rightarrow$ (2) : If v is a basic feasible solution, then let's assume $\{a_1, \dots, a_n\}$ to be the active constraints. Let $c = -\sum_{i=1}^{n} a_i$ and we can see that $c^T v = \sum_{i=1}^{n} (-a_i^T v) = -\sum_{i=1}^{n} a_i^T v$ $= -(b_1 + b_2 + \dots + b_n) \leq -\sum_{i=1}^{n} a_i^T u$, for any $u \in P$, $u \neq v$.

Hence, v is a vertex.

(2) $\Longrightarrow$ (1) : If v is a vertex, then $\exists\, c \in \mathbb{R}^n$ s.t. $c^T v < c^T k$, for any $k \in P$, $k \neq v$. We want to show that this means v cannot be written as a convex combination of another points in P. Assume it can, that is, $v = \lambda y + (1-\lambda) z$, for $y, z \in P$. Then $c^T v = c^T(\lambda y + (1-\lambda) z) = \lambda c^T y + (1-\lambda) c^T z > \lambda c^T v + (1-\lambda) c^T v = c^T v$ ↯

Hence, v is an extreme point.

Therefore, all three statements are equivalent.

11. (a) Let's rewrite the constraints that doubly stochastic matrices satisfy as

$$P=\{x\in\mathbb{R}^{n^2} \mid Ax\le b,\ Cx=d\} \text{ where } x=\begin{bmatrix} x_{11}\\ \vdots \\ x_{1n}\\ x_{21}\\ \vdots\\ x_{2n}\\ \vdots\\ x_{n1}\\ \vdots\\ x_{nn}\end{bmatrix}_{n^2\times 1},\quad A=-I_{n^2\times n^2},\quad b=0_{n^2\times 1},\quad C=\begin{bmatrix} 1\cdots 1 & 0\cdots 0 & \cdots & 0\cdots 0\\ 0\cdots 0 & 1\cdots 1 & \cdots & 0\cdots 0\\ & & \vdots & \\ 0\cdots 0 & 0\cdots 0 & \cdots & 1\cdots 1\\ 1\cdots 0 & 1\cdots 0 & \cdots & 1\cdots 0\\ 01\cdots 0 & 01\cdots 0 & \cdots & 01\cdots 0\\ & & \vdots & \\ 0\cdots 1 & 0\cdots 1 & \cdots & 0\cdots 1\end{bmatrix}_{2n\times n^2},\quad d=\mathbf{1}_{2n\times 1}$$

(I.e. we convert each matrix as a long column vector in $\mathbb{R}^{n^2}$)

~~Then we can see that $\#\text{nullspace}\left(\begin{bmatrix}A\\C\end{bmatrix}\right)=\#\text{nullspace}\left(\begin{bmatrix}-I\\C\end{bmatrix}_{(n^2+2n)\times n^2}\right)=n^2$. This means that~~

Let's look at $\text{rank}\left(\begin{bmatrix}A\\C\end{bmatrix}\right)=\text{rank}\left(\begin{bmatrix}-I\\C\end{bmatrix}\right)=n^2$. Thus the null space of $\begin{bmatrix}A\\C\end{bmatrix}=\{0\}$. This is saying the linearity space of P is just $\{0\}$. Then, by definition, P is a pointed polyhedron.

(b). In class, we proved a fact that for a polyhedron $\{x\in\mathbb{R}^n \mid x\ge 0,\ Cx=d\}$, an extreme point $\hat{x}$ has at most rank(C) non-zero elements.

We can apply this to our polyhedron P and we know that the extreme point $\hat{x}$ of P has at most rank(C) non-zero elements. Here, ~~four~~ for our special C as described above, rank(C) is $2n-1$ because the sum of rows sums is equal to sum of column sums.

Therefore, if $\hat{x}$ is an EP of P, ~~the~~ then it must have one row with exactly one nonzero element (with value 1) and one colum with exactly one nonzero element (with value 1). WLOG, let this row (and column) be the first row (and first column) of $\hat{x}$. We can then look at a smaller matrix formed by excluding the first row and first column. The resulted matrix has to be an extreme point for the set of doubly stochastic matrices in $\mathbb{R}^{(n-1)\times(n-1)}$. With induction, this ensures that $\hat{x}$ has exactly one 1 in each row and column.