Machine Learning Homework 3

PB22061259 Liu Pei

Exercise 1: Affine Sets

1

(a)

It is a subspace to it is a affine set:

> Because U is a subspace,so any vectors $\mathbf{x_1}, \mathbf{x_2}$ in U,where $\mathbf{x_1} \neq \mathbf{x_2}$,we have $\forall\theta \in \mathbb{R}, \theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \in U$,whice means U is a affine set.

It is a affine set to it is a subspace:

> Because U is a affine set,so any vectors $\mathbf{x_1}, \mathbf{x_2}$ in U ,we have $\forall\theta \in \mathbb{R}, \theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \in U$.
> Because $\mathbf{0} \in U$,o we can let $\mathbf{x_2} = \mathbf{0}$
> Then we can get $\forall\mathbf{x_1} \in U, \forall\theta \in \mathbb{R}, \theta\mathbf{x_1} \in U$.
> Then ,let $\theta = 0.5$,we can get $\forall\mathbf{x_1}, \mathbf{x_2} \in U, 0.5(\mathbf{x_1} + \mathbf{x_2}) \in U, \mathbf{x_1} + \mathbf{x_2} \in U$.

So,U is a affine set if it is a subspace.
In summary,U is a affine set if and only if it is a subspace.

(b)

The question is also equivalent to the following:

> $\text{If } U \subset \mathbb{R}^n \text{ is an affine set, there is a unique subspace } V \subset \mathbb{R}^n \text{ such that } V = U - \mathbf{u}$ $\text{for any } \mathbf{u} \in U.$

Firstly ,we can show that V is a subspace:

> Because U is a affine set,so any vectors $\mathbf{u}$ in U ,we have the vector $\mathbf{u}$ is in U so that $\mathbf{0} \in V$

Then,we consider $\forall\mathbf{v_1} = (\mathbf{u_1} - \mathbf{u}) \in V$,we have $\forall\alpha \in \mathbb{R}, \alpha\mathbf{v_1} = [\alpha(\mathbf{u_1} - \mathbf{u}) = \alpha\mathbf{u_1} + (1-\alpha)\mathbf{u_1}] - \mathbf{u} \in V$,whice means $\forall\alpha \in \mathbb{R}, \alpha\mathbf{v_1} \in V$.
Then ,we consider $\forall\mathbf{v_1}, \mathbf{v_2} \in V$:

$$\begin{aligned}\mathbf{v_1}+\mathbf{v_2} &= \mathbf{u_1}-\mathbf{u}+\mathbf{u_2}-\mathbf{u}\\ &= 2[0.5(\mathbf{u_1}+0.5\mathbf{u_2})-\mathbf{u}]\end{aligned}$$

So $\mathbf{v_1}+\mathbf{v_2}$ is in V ,which means V is a subspace.

Secondly,we can show that V is unique:

If we have two different subspaces V_1 and V_2 such that $V_1=U-\mathbf{u}$ and $V_2=U-\mathbf{u'}$,where $\mathbf{u},\mathbf{u'}\in U$,then we can get:

There is a vector $\mathbf{v_1}$ such that $\mathbf{v_1}\in V_1$ and $\mathbf{v_1}\notin V_2$.

$$2\mathbf{v_1}=2\mathbf{u_1}-2\mathbf{u}=[2\mathbf{u_1}-2\mathbf{u}+\mathbf{u'}]-\mathbf{u'}=[2\mathbf{u_1}-(2\mathbf{u}-\mathbf{u'})]-\mathbf{u'}$$

We have $2\mathbf{u}-\mathbf{u'}\in U$,so $[2\mathbf{u_1}-(2\mathbf{u}-\mathbf{u'})]\in U$.

So,$[2\mathbf{u_1}-(2\mathbf{u}-\mathbf{u'})]-\mathbf{u'}$ in V_2.

It contradicts with the fact that $\mathbf{v_1}\in V_1$ and $\mathbf{v_1}\notin V_2$.So,the suspution is wrong,which means V is unique.

2

(a)

We consider $\forall\mathbf{x_1},\mathbf{x_2}\in C$,we have

$$\forall\theta\in\mathbb{R},\mathbf{A}(\theta\mathbf{x_1}+(1-\theta)\mathbf{x_2})=\theta\mathbf{A}\mathbf{x_1}+(1-\theta)\mathbf{A}\mathbf{x_2}=\theta\mathbf{b}+(1-\theta)\mathbf{b}=\mathbf{b}$$

Which means

$$\forall\mathbf{x_1},\mathbf{x_2}\in C,\forall\theta\in\mathbb{R},\theta\mathbf{x_1}+(1-\theta)\mathbf{x_2}\in C$$

So C is a affine set.

(b)

We consider

> If $U\subset\mathbb{R}^n$ is an affine set, there is a unique subspace $V\subset\mathbb{R}^n$ such that $U=V+\mathbf{u}$ for any $\mathbf{u}\in U$.

And,for the subspace V, also can be a matrix A's null space.i.e.

There exists $\mathbf{A}\in\mathbb{R}^{m\times n}$ and $\mathbf{b}\in\mathbb{R}^m$

$$A\mathbf{x}=\mathbf{0},\forall\mathbf{x}\in V,\text{ where } m\le n.$$

If we let $\mathbf{b} = A\mathbf{u}$,
Then ,we have

$$\text{set } C = \{\mathbf{x} : \mathbf{Ax} = \mathbf{b}\} = U$$

Exercise 2: Convex Sets

1

(a)

Suppose $\mathbf{cl}C$ is not a convex set, then there exists $\mathbf{x}, \mathbf{y} \in \mathbf{cl}C$ such that $\exists\theta \in [0, 1]$ such that $\theta\mathbf{x} + (1-\theta)\mathbf{y} \notin \mathbf{cl}C$.
So,we can get :The sequences $\{x_n\}$ and $\{y_n\}$ in C converge to $\mathbf{x}$ and $\mathbf{y}$ respectively.
Because C is convex set,so

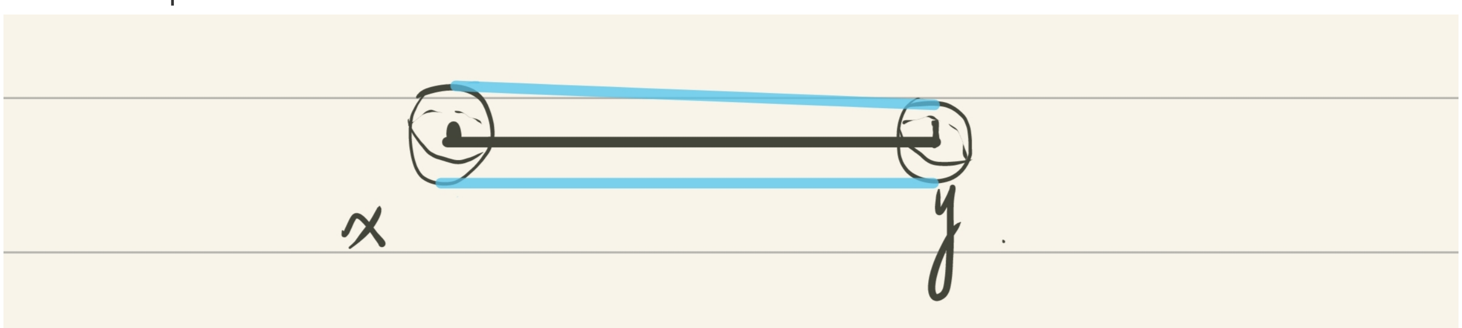
$$\theta\mathbf{x_n} + (1-\theta)\mathbf{y_n} \in C, \forall\theta \in [0, 1]$$

Then we can get the sequence $\{\theta\mathbf{x_n} + (1-\theta)\mathbf{y_n}\}$ converges to $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathbf{cl}C$.
It is ridiculous,so $\mathbf{cl}C$ is not a convex set.

Here,we are going to prove that $\mathbf{int}C$is a convex set:

> $\forall\mathbf{x}, \mathbf{y} \in \mathbf{int}C$,their open balls $\mathbf{B}(\mathbf{x}, \mathbf{r_x})$ and $\mathbf{B}(\mathbf{y}, \mathbf{y_x}) \in C$
> Because C is a convex, we choose a point $\mathbf{z} \in \mathbf{B}(\mathbf{x}, \mathbf{r_x})$ and $\mathbf{w} \in \mathbf{B}(\mathbf{y}, \mathbf{r_y})$,we have the line between $\mathbf{z}$ and $\mathbf{w}$ is also in C.

Look at this picture:



> Then we can get any point in the line between $\mathbf{x}, \mathbf{y}$ should have a open ball is also in C.

Which means $\mathbf{int}C$ is also a convex set.

(b)

According to the definition ,we can say

$$\forall \mathbf{x},\mathbf{y} \in \mathbf{relint} C, \forall \theta \in [0,1], \theta\mathbf{x} + (1-\theta)\mathbf{y} \in C$$

If $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathbf{relint} C$,we have $\mathbf{relint} C$ is a convex set.
So,we are goint to prove that $\theta\mathbf{x} + (1-\theta)\mathbf{y}$ is a relative interior point of C.

$$\forall \mathbf{z} \in C, \exists \epsilon > 0, \text{such that } \mathbf{x} - \epsilon(\mathbf{z}-\mathbf{x}), \mathbf{y} - \epsilon(\mathbf{z}-\mathbf{y}) \in C$$

So,we have $\theta(\mathbf{x} - \epsilon(\mathbf{z}-\mathbf{x})) + (1-\theta)(\mathbf{y} - \epsilon(\mathbf{z}-\mathbf{y})) \in C$

$$[\theta\mathbf{x} + (1-\theta)\mathbf{y}] - \epsilon(\mathbf{z} - [\theta\mathbf{x} + (1-\theta)\mathbf{y}]) \in C$$

Which means $\theta\mathbf{x} + (1-\theta)\mathbf{y}$ is a relative interior point of C.
Therefore, $\mathbf{relint} C$ is a convex set.

(c)

Let the intersection of C_i be C,we have:

$$\forall \mathbf{x},\mathbf{y} \in C, \forall \theta \in [0,1], \theta\mathbf{x} + (1-\theta)\mathbf{y} \in C_i$$

So,we can say $\theta\mathbf{x} + (1-\theta)\mathbf{y}$ in the intersection of C_i,whice means $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in C$.
So,the intersection of C_i is also a convex set.

(d)

Let the set be C'.
$\forall \mathbf{y_1} = A\mathbf{x_1} + b$ and $\forall \mathbf{y_2} = A\mathbf{x_2} + b \in C', \text{then } \mathbf{x_1}, \mathbf{x_2} \in C.$
$\forall \theta \in [0,1], \text{we have:}$
$\theta\mathbf{y_1} + (1-\theta)\mathbf{y_2} = \theta(A\mathbf{x_1} + b) + (1-\theta)(A\mathbf{x_2} + b) = A[\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}] + b$
Because C is a convex set,so $\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}$ is in C,so $\theta\mathbf{y_1} + (1-\theta)\mathbf{y_2}$ in C'.
Therefore, C' is a convex set.

(e)

Let the set be C'.
$\forall \mathbf{x_1} = A\mathbf{y_1} + b$ and $\forall \mathbf{x_2} = A\mathbf{y_2} + b \in C, \text{then } \mathbf{y_1}, \mathbf{y_2} \in C'.$
$\forall \theta \in [0,1], \text{we have:}$
$B(\theta\mathbf{y_1} + (1-\theta)\mathbf{y_2}) + \mathbf{b} = \theta(B\mathbf{y_1} + \mathbf{b}) + (1-\theta)(B\mathbf{y_2} + \mathbf{b}) = \theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \in C$
So $\theta\mathbf{y_1} + (1-\theta)\mathbf{y_2}$ is inC'

Therefore, C' is a convex set.

2

(a)

The interior set is ϕ,the relative interior set is $\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\}$.

(b)

The interior set is ϕ,the relative interior set is $\{\mathbf{A} \in S_{++}^n : \mathrm{Tr}(\mathbf{A}) = 1\}$.

(c)

The interior set is ϕ,the relative interior set is $\{\mathbf{A} \in S_{++}^n : \mathrm{Tr}(\mathbf{A}) = 1\}$.

Exercise 3: Relative Interior and Interior

1

Firstly,we should find out what $\mathbf{x}_0 + r\mathbf{v}, \text{ for any } \text{ and } \|\mathbf{v}\|_2 \leq 1$ represents.

> $\|\mathbf{x}_0 + r\mathbf{v} - \mathbf{x}_0\|_2 = \|r\mathbf{v}\|_2 \leq r$,so we can say $\mathbf{x}_0 + r\mathbf{v} \in \mathbf{B}(\mathbf{x_0}, \mathbf{r})$.

Then we try to prove $\mathbf{x}_0 + r\mathbf{v} \in \mathbf{aff}C$:

> $\mathbf{v} = \mathbf{v_1} - \mathbf{x}_0, \mathbf{v_1} \in \mathbf{aff}C$,we have $\mathbf{x}_0 + r\mathbf{v} = (1-r)\mathbf{x}_0 + r\mathbf{v_1}$
> We have,$\mathbf{v_1} \in \mathbf{aff}C, \mathbf{x}_0 \in C$,that is
>
> $$\mathbf{v_1} = \sum_{i=0}^{n} \theta_i \mathbf{x_i}, \sum_{i=0}^{n} \theta_i = 1$$
>
> Then we have $(1-r)\mathbf{x}_0 + r\mathbf{v_1} = (1-r)\mathbf{x}_0 + r\sum_{i=0}^{n} \theta_i \mathbf{x_i}$
> Because $1 - r + r\sum_{i=0}^{n} \theta_i = 1$,so $\mathbf{x}_0 + r\mathbf{v} \in \mathbf{aff}C$.

Therefore, $\mathbf{x}_0 + r\mathbf{v} = \mathbf{aff}C \cap \mathbf{B}(\mathbf{x_0}, \mathbf{r})$.
And,we know $\mathbf{x}_0 + r\mathbf{v} = \mathbf{aff}C \cap \mathbf{B}(\mathbf{x_0}, \mathbf{r}) \in C$,so $\mathbf{x}_0 \in \mathbf{relint}C$.

Then,we are going to prove that if $\mathbf{x}_0 \in \mathbf{relint}C$,then :
$\text{There exists } r > 0 \text{ such that } \mathbf{x}_0 + r\mathbf{v} \in C \text{ for any } \mathbf{v} \in \mathbf{aff}C - \mathbf{x}_0 \text{ and } \|\mathbf{v}\|_2 \leq 1$.

It is easy to see that $\mathbf{aff}C \cap \mathbf{B}(\mathbf{x_0}, \mathbf{r}) \in C$,so there exists $r > 0, \mathbf{x}_0 + r\mathbf{v} \in C$ for any $\mathbf{v} \in \mathbf{aff}C - \mathbf{x}_0$ and $\|\mathbf{v}\|_2 \leq 1$.

2

(a)

For any $\mathbf{y} \in C$,the $(\mathbf{x} - \mathbf{y})$ can represents any direction from $\mathbf{x}$ to the points in $\mathbf{aff} C$.
And,$\mathbf{v} \in \operatorname{aff} C - \mathbf{x}_0$ in question 1 represents a direction :from $\mathbf{x}_0$ to the points in $\mathbf{aff} C$
So,we can say $\mathbf{x} + \gamma(\mathbf{x} - \mathbf{y})$ represents a point move away from $\mathbf{x}$ by direction from $\mathbf{x}$ to the points in C
.
So.$\mathbf{x} + \gamma(\mathbf{x} - \mathbf{y})$ represents the same as question 1 $\mathbf{x}_0 + r\mathbf{v}$ if we let max$||\gamma(\mathbf{x} - \mathbf{y})||_2 = r$
So,we have the conclusion that the condition of question 2 is equivalent to the condition of question 1.

(b)

If $\mathbf{y} \in \mathbf{relint} C$,then $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in \mathbf{relint} C$.
But if $\mathbf{y} \notin \mathbf{relint} C$,which means $\mathbf{y}$ is a boundary point of C,whice means,$\mathbf{B(y, r)} \cap \mathbf{C} \neq \phi$ then,
$\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} = \mathbf{y} + \lambda(\mathbf{x} - \mathbf{y})$.
Let's talk about the case what $\mathbf{y} + \lambda(\mathbf{x} - \mathbf{y})$ represents:

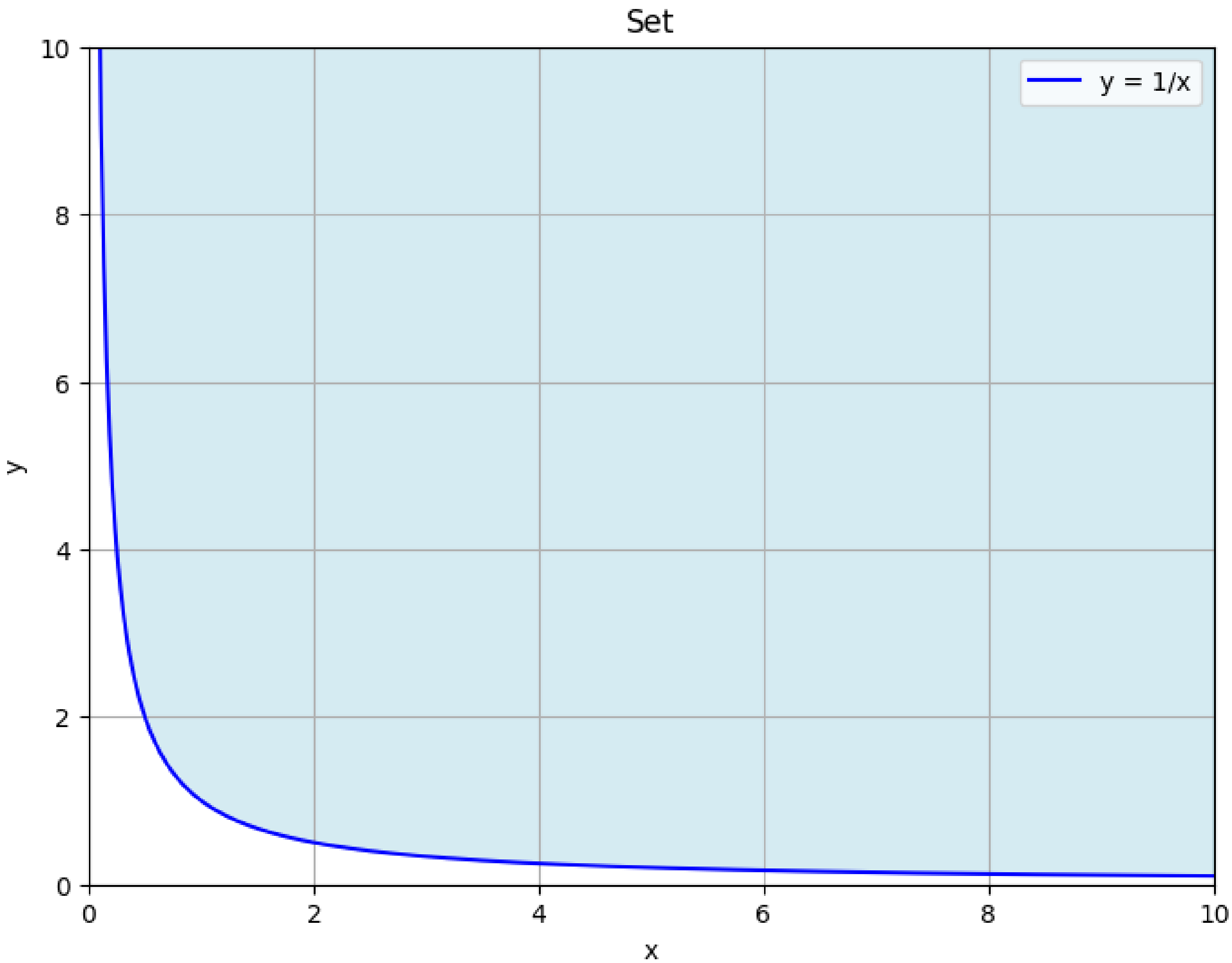
> $\mathbf{y} + \lambda(\mathbf{x} - \mathbf{y})$ is a point move away from $\mathbf{y}$ by direction from $\mathbf{y}$ to the interior points in C.

Exercise 4: Supporting Hyperplane

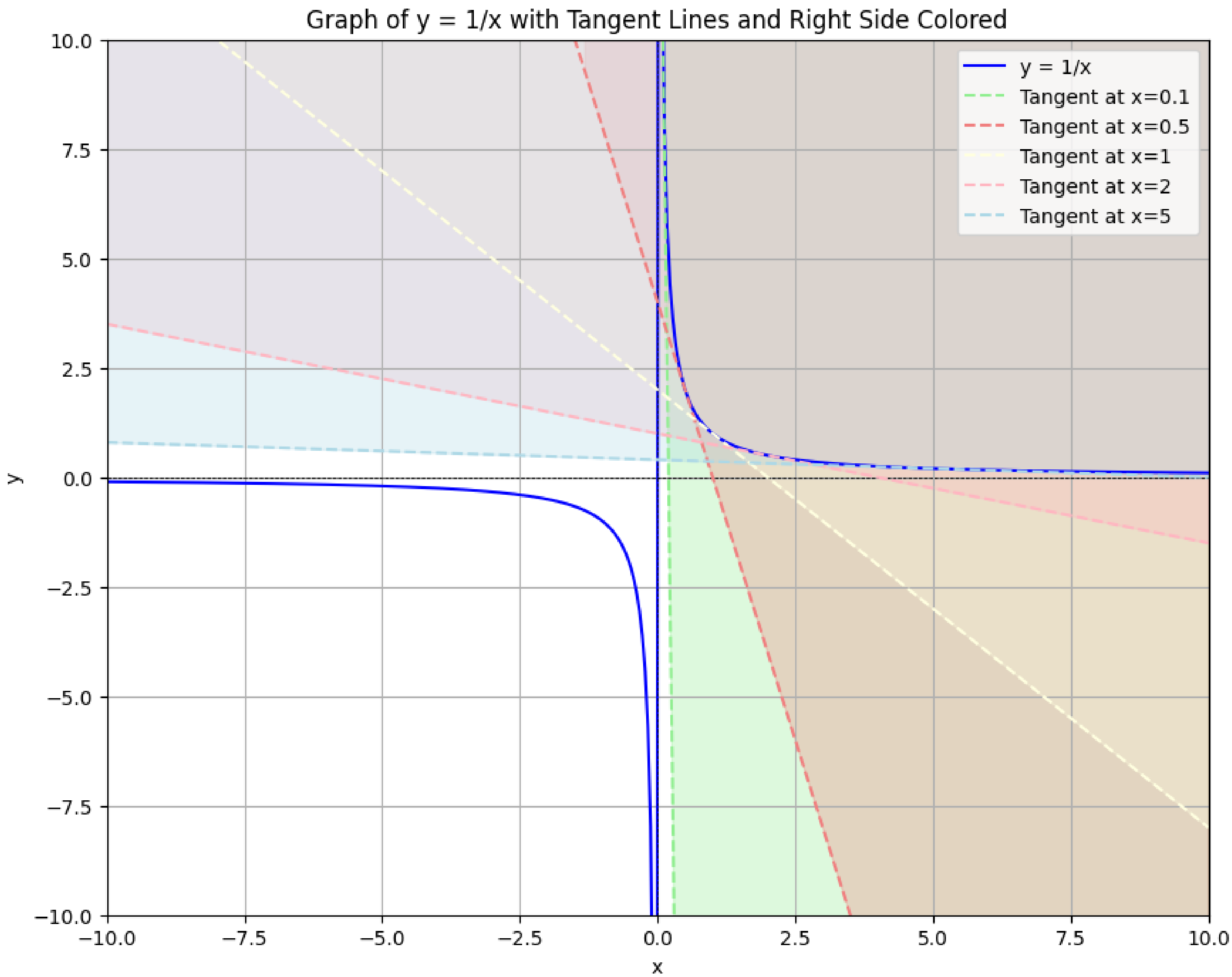
1

(a)

$\{\mathbf{x} \in \mathbb{R}_+^2 \mid x_1x_2 \geq 1\}$
We can easily to draw the set beacuse it is in $\mathbb{R}_+^2$, and the set is the part above the curve $y = \frac{1}{x}$



When we draw the tangent line,we can find that every tangent line can be a supporting hyperplane which sepeartes the whole space into two parts,and if we choose the above parts and find the intersection,we can easily get the set.



(b)

$C = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_\infty \leq 1\}$
First,we should find the boundary of C.
We can easily to find that the boundary of C is $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}_i = 1\}$.

> $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a boundary point of C if and only if x_i = 1 for some i.

Then,we are going to prove that C is a convex set.

> C is a convex set because if $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0,1]$,then $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} = (\lambda x_i + (1 - \lambda y_i), \ldots, \lambda x_n + (1 - \lambda y_n))$.
> $\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\|_\infty \leq \lambda + (1-\lambda) = 1$,so $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in C$

WLOG,we can suppose $(1, x_1, \ldots, x_n)$ is a boundary point of the set.
When we consider this theorem

> Theorem 4 (Supporting Hyperplane Theorem). Let C $\subseteq \mathbb{R}^n$ *be a nonempty convex set, and*

> $\mathbf{x}_0 a point in its boundary \mathbf{bd} C. Then, there exists a hyperplane supporting C at \mathbf{x}_0.$

We can easily to find the hyperplanes supporting C at $\mathbf{x}_0$:
Let's consider a sequence $\mathbf{x}_k = (1 + 1/n, x_1, \dots, x_n), \mathbf{x}_k \notin \mathbf{cl}\, C,\ k = 1, 2, \dots,$ and $\mathbf{x}_k \to \mathbf{x}_0$
We then construct a sequence of unit norm vectors by

$$\mathbf{a}_k = \frac{\mathbf{x}_k - \Pi_{\bar{C}}(\mathbf{x}_k)}{\|\mathbf{x}_k - \Pi_{\bar{C}}(\mathbf{x}_k)\|}.$$

Notice that $\mathbf{cl}\ C$ is a closed convex set and $\mathbf{x}_k \notin \mathbf{cl}\ C,\ k = 1, 2, \dots$ By Theorem 3, we have

$$\langle \mathbf{a}_k, \mathbf{y} \rangle \le \langle \mathbf{a}_k, \mathbf{x}_k \rangle,\ \forall\, \mathbf{y} \in C.$$

As $\|\mathbf{a}_k\| = 1$ for all $k = 1, 2, \dots,$, there exists a converging subsequence. Without loss of generality, we assume that $\mathbf{a}_k \to \mathbf{a}$. Passing to the limit on both sides of the above inequality, we have

$$\langle \mathbf{a}, \mathbf{y} \rangle \le \langle \mathbf{a}, \mathbf{x}_0 \rangle,\ \forall\, \mathbf{y} \in C,$$

So,we get a hyperplane $H_{(\mathbf{a},\mathbf{x}_0)}$ supporting C at $\mathbf{x}_0$.
Similarly,$\text{find a sequence } (\mathbf{x}_k) \text{ with } \mathbf{x}_k \notin \mathbf{cl}\ C, k = 1, 2, \dots, \text{ and } \mathbf{x}_k \to \mathbf{x}_0$,then we can get another hyperplane.By this way,we can easily to find the hyperplanes supporting C at all the boundary points of C.

2

(a)

First,let's show that the set $(\mathbf{a}, b) \in \mathbb{R}^{n+1}$ denoted as S is convex cone.
For any $(\mathbf{a_1}, b_1), (\mathbf{a_2}, b_2) \in S, \theta_1, \theta_2 \in [0, +\infty)$,we have:

1. For $\mathbf{x} \in C$:
$$\begin{aligned}(\theta_1 \mathbf{a_1} + \theta_2 \mathbf{a_2})^T x &= \theta_1 \mathbf{a_1}^T x + \theta_2 \mathbf{a_2}^T x \\ &\le \theta_1 b_1 + \theta_2 b_2 \quad (\text{since } \theta_1 \mathbf{a_1}^T x \le b_1 \text{ and } \theta_2 \mathbf{a_2}^T x \le b_2)\end{aligned}$$
2. For $\mathbf{x} \in D$:
$$\begin{aligned}(\theta_1 \mathbf{a_1} + \theta_2 \mathbf{a_2})^T x &= \theta_1 \mathbf{a_1}^T x + \theta_2 \mathbf{a_2}^T x \\ &\ge \theta_1 b_1 + \theta_2 b_2 \quad (\text{since } \theta_1 \mathbf{a_1}^T x \ge b_1 \text{ and } \theta_2 \mathbf{a_2}^T x \ge b_2)\end{aligned}$$

So, S is a cone.

Exercise 5: Farkas' Lemma

1

(a)

Firstly,let's show that $\mathbf{cone}A$ is a closed set.
Suppose $\mathbf{cone}A$ is not a closed set,then there exists a sequence $\{\mathbf{x}_k\}_{k=1}^{\infty}$ such that $\mathbf{x}_k \in \mathbf{cone}A$ and $\mathbf{x}_k \to \mathbf{x}$,$\mathbf{x} \notin \mathbf{cone}A$.
We can write $\mathbf{x}_k = \sum_{i=1}^{n} \alpha_i^k \mathbf{a}_i, \alpha_i^k \geq 0$.
But,we know $\forall \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i \mathbf{a}_i \in \mathbf{cone}A$.
So,we can assert that $\lim_{k\to\infty} \mathbf{x}_k \in \mathbf{cone}A$.
Then,we can show that $\mathbf{cone}A$ is a closed set.

Then,we can show that $\mathbf{cone}A$ is a convex set.
For any $\mathbf{x}, \mathbf{y} \in \mathbf{cone}A, \lambda \in [0,1]$,we have:

$$\lambda \mathbf{x} + (1-\lambda)\mathbf{y} = \sum_{i=1}^{n} (\lambda \alpha_i^{(x)} + (1-\lambda)\alpha_i^{(y)})\mathbf{a}_i \in \mathbf{cone}A$$

So, $\mathbf{cone}A$ is a convex set.

2

Let $\mathbf{b} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i, \alpha_i \geq 0, x = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$.
Then,we have:$\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} > \mathbf{0}$

3

We have $\mathbf{cone}A$ is a closed convex set,and $\mathbf{b} \notin \mathbf{cone}A$.
Then,sccording to the sepration theorem,there exists a superplane $\mathbf{H}_{(\mathbf{y},\alpha)}$ such that separates $\mathbf{cone}A$ and $\mathbf{b}$.
WLOG,we can suppose $\mathbf{b}^{\mathbf{T}}\mathbf{y} \leq 0$,then we have:
$\forall \mathbf{x} \in \mathbf{cone}A, \mathbf{x}^{\mathbf{T}}\mathbf{y} \geq 0$,which contains $\mathbf{A}^{\mathbf{T}}\mathbf{y} \geq \mathbf{0}$.
But,we know $\mathbf{cone}A$ is a closed convex set,so $\mathbf{b}$ can't be the boundary point of $\mathbf{cone}A$.
So,we can assert that there exists $\mathbf{y} \in \mathbb{R}^m$,such that $\mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top \mathbf{y} < 0$.

4

This is the same as the previous part.When $\mathbf{b} \notin \mathbf{cone}A$,we have the first situation.
When $\mathbf{b} \in \mathbf{cone}A$,we have the second situation.