Lecture 3. Convex Sets

3.1 Affine sets

Affine sets are generalization of lines. Given two points $x, y \in \mathbb{R}^n$, the line passing through x, y can be represented by

$$\ell = \{x + \theta(y - x) \mid \theta \in \mathbb{R}\}.$$

Note that $x + \theta(y - x) = (1 - \theta)x + \theta y$. So we have the following definition of *affine combination*.

Definition (*Affine combination*)

Given $x_1, \ldots, x_m \in \mathbb{R}^n$, $\theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m$ is an *affine combination* of $x_1, \ldots, x_m$ if $\theta_1 + \ldots + \theta_m = 1$.

A set is *affine* if it is closed under affine combinations.

Definition (*Affine set*)

A set S is an *affine set*, if for all $m \geq 1$, for all m points $x_1, x_2, \ldots, x_m \in S$, any affine combination of $x_1, \ldots, x_m$ is still in S.

Example

- A line is an affine set;
- $\mathbb{R}^n$ is an affine set;
- Given $\boldsymbol{w} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, the hyperplane $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{w}^\mathsf{T}\boldsymbol{x} + b = 0\}$ is an affine set;
- In general, given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, the solution set of the system of linear equations $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ is an affine set.

Note that if $m = 1$, the solution set S is a hyperplane. If $m > 1$ and $\boldsymbol{A} \neq \boldsymbol{0}$, S is the intersection of m hyperplanes.

Proof

Given $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$, we have $\boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{b}$. So for any $\theta \in \mathbb{R}$, $\boldsymbol{A}(\theta\boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2) = \theta\boldsymbol{A}\boldsymbol{x}_1 + (1-\theta)\boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{b}$.

Why can we only verify affine combinations of two points in S? Suppose we have an affine combination $\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$ for 3 points x_1, x_2, x_3. Since $\theta_1 + \theta_2 + \theta_3 = 1$, clearly there must exists two of them such that their sum is non-zero. Assume that $\theta_1 + \theta_2 \neq 0$. Then we have

$$\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 = (\theta_1 + \theta_2)\left(\frac{\theta_1}{\theta_1 + \theta_2} x_1 + \frac{\theta_2}{\theta_1 + \theta_2} x_2\right) + \theta_3 x_3 \,.$$

If any affine combination of two points is still in S, then $\frac{\theta_1}{\theta_1+\theta_2} x_1 + \frac{\theta_2}{\theta_1+\theta_2} x_2$ is in S and thus $\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$ is in S. For an affine combination of more than 3 points, we can rewrite it in a similar way recursively. So it suffices to verify affine combinations of 2 points.

We have shown that the solution to each linear equation is an affine set. Conversely, any affine set is also a solution set to a system of linear equations.

Proposition

Any affine set $\subseteq \mathbb{R}^n$ is the solution set to a system of linear equations.

Proof

If S is an affine set, pick an arbitrary point $x_0 \in S$. Then we claim that the following set

$$S' = S - x_0 \triangleq \{x - x_0 \mid x \in S\}$$

is a linear space. For all $x_1, x_2 \in S'$, we have $x_1 + x_0, x_2 + x_0 \in S$ by definition. Hence, for any $a_1, a_2 \in \mathbb{R}$,

$$a_1 x_1 + a_2 x_2 + x_0 = a_1(x_1 + x_0) + a_2(x_2 + x_0) + (1 - a_1 - a_2)x_0 \in S \,.$$

Therefore, $a_1 x_1 + a_2 x_2 \in S'$.

Since S' can be represented as $\{x \mid Ax = 0\}$, then $S = S' + x_0$ can be represented as $\{x \mid Ax = Ax_0\}$, which is the solution set to $Ax = Ax_0$.

Roughly speaking, *affine* can be viewed as *linear* added by some bias term. Similar to the *linear map*, we can define an *affine map* $f : \mathbb{R}^n \to \mathbb{R}^m$ by $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ for some $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^n$. We can also define *affinely independent points* as follows.

Definition (*Affine independence*)

Given $m+1$ points $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$, we say they are *affinely independent*, if there **does not** exist $\theta_0, \theta_1, \ldots, \theta_m \in \mathbb{R}$ such that $\theta_0 + \theta_1 + \cdots + \theta_m = 0$, and

$$\theta_0 \boldsymbol{x}_0 + \theta_1 \boldsymbol{x}_1 + \cdots + \theta_m \boldsymbol{x}_m = \boldsymbol{0}\,.$$

Equivalently, $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ are *affinely independent*, if and only if $\boldsymbol{x}_1 - \boldsymbol{x}_0, \boldsymbol{x}_2 - \boldsymbol{x}_0, \ldots, \boldsymbol{x}_m - \boldsymbol{x}_0$ are *linearly independent*.

Clearly, there are at most $n+1$ affinely independent points in $\mathbb{R}^n$, since there are at most n linearly independent vectors in $\mathbb{R}^n$.

3.2 Convex sets

Similar to the definition of lines, we can define the *segment* from x to y by

$$s = \{x + \theta(y - x) \mid \theta \in [0, 1]\}\,.$$

Note that the difference between lines and segments is the range of θ. Again, since $x + \theta(y - x) = (1 - \theta)x + \theta y$, we have the following definition.

Definition (*Convex combination*)

Given $x_1, \ldots, x_m \in \mathbb{R}^n$, $\theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m$ is a *convex combination* of $x_1, \ldots, x_m$ if $\theta_1 + \ldots + \theta_m = 1$ and for all $i \in [m]$, $\theta_i \geq 0$.

A set is *convex* if it is closed under convex combinations.

Definition (*Convex set*)

A set S is a *convex set*, if for all $m \geq 1$, for all m points $x_1, x_2, \ldots, x_m \in S$, any convex combination of $x_1, \ldots, x_m$ is still in S.

In particular, we can define the *convex hull* of any set.

Definition (*Convex hull*)

The *convex hull* of a set S is the set of all convex combinations of points in S, namely,

$$\operatorname{conv}(S) \triangleq \left\{ \sum_{i=1}^{m} \theta_i x_i \mid \forall\, i \in [m], \theta_i \geq 0, x_i \in S, \text{ and } \sum_{i=1}^{m} \theta_i = 1 \right\}.$$

Clearly, for any set $S \in \mathbb{R}^n$, its convex hull is a convex set.

For a general set S, if we would like to show that S is convex, using the same argument we used in the section of affine sets, we only need to show that any convex combination of two arbitrary points in S is still in S.

Question

If we would like to determine the *convex hull* of some set S, can we only check convex combinations of any two points? If not, how many points are sufficient?

Tip (*Carathéodory's theorem*)

At most $n+1$ points in $\mathbb{R}^n$ are sufficient. Because $n+2$ points are *affinely dependent*, there exists $\theta_0, \ldots, \theta_{n+1}$ such that $\theta_0 + \cdots + \theta_{n+1} = 0$, and $\theta_0 \boldsymbol{x}_0 + \cdots + \theta_{n+1} \boldsymbol{x}_{n+1} = \mathbf{0}$. Thus,

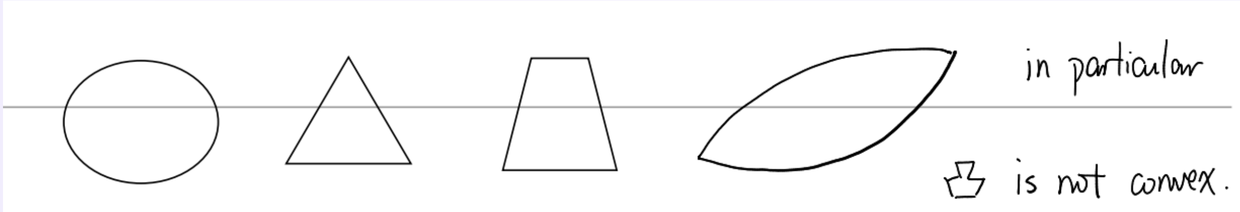
$$\boldsymbol{x}_{n+1} = \frac{\theta_0 \boldsymbol{x}_0}{\theta_0 + \cdots + \theta_n} + \frac{\theta_1 \boldsymbol{x}_1}{\theta_0 + \cdots + \theta_n} + \cdots + \frac{\theta_n \boldsymbol{x}_n}{\theta_0 + \cdots + \theta_n}$$

is a convex combination of $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n$.

3.3 Examples of convex sets

We first give some geometric examples of convex sets.

Example



A particular example of convex sets is the *convex cone*.

Definition (*Conic combination*)

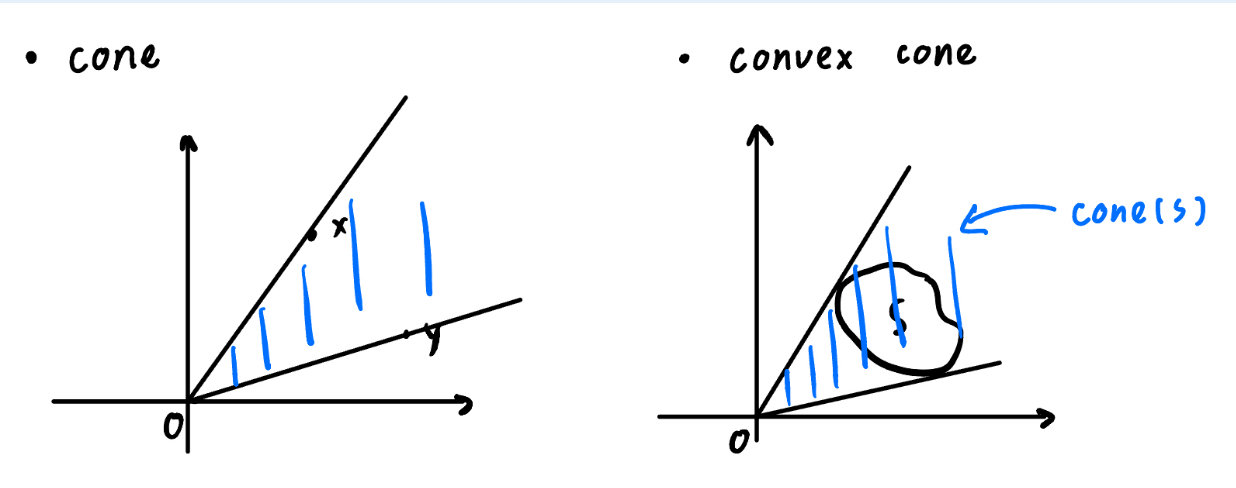
Given $x_1, \ldots, x_m \in \mathbb{R}^n$, $\theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m$ is a *conic combination* of $x_1, \ldots, x_m$ if for all $i \in [m]$, $\theta_i \geq 0$.

A *convex cone* is a set closed under conic combinations. A *convex cone hull* of a set S is the conic combination version of a *convex hull*.

Definition (*Convex cone*)

The *convex cone hull* of a set S is the set of all conic combinations of points in S, namely,

$$\operatorname{cone}(S) \triangleq \left\{ \sum_{i=1}^{m} \theta_i x_i \mid \forall i \in [m], \theta_i \geq 0, x_i \in S \right\}.$$



Clearly, any cone is a convex set.

Another examples include $\mathbb{R}^n$, hyperplanes and halfspaces.

Example

- Affine sets are all convex sets. So $\mathbb{R}^n$ and hyperplane $\{x \mid w^\top x + b = 0\}$ are convex sets.
- A *halfspace* defined by $H \triangleq \{x \mid w^\top x + b \leq 0\}$ (or < 0 for open *halfspace*) is a convex set.
  However, H is not affine unless $H = \mathbb{R}^n$.

Proof (*Convexity of halfspaces*)

For all $x, y \in H = \{x \mid w^\top x + b \leq 0\}$, let $z \triangleq \theta x + \bar{\theta} y$, where $\theta \in [0, 1]$ and $\bar{\theta} = 1 - \theta$. Since

$$w^\top z = w^\top(\theta x + \bar{\theta} y) + b = \theta(w^\top x + b) + \bar{\theta}(w^\top y + b) \leq 0\,,$$

we conclude that z is also in the halfspace H.

Convexity is not only a property of geometric shapes.

Example (*Definite matrices*)

Let $\mathcal{S}_+^n$ and $\mathcal{S}_{++}^n$ denote the set of all *positive semidefinite matrices* and the set of all *positive definite matrices*, respectively, namely,

$$\begin{aligned} \mathcal{S}_+^n &= \{A \in \mathbb{R}^{n\times n} \mid A \succeq 0\}\,, \\ \mathcal{S}_{++}^n &= \{A \in \mathbb{R}^{n\times n} \mid A \succ 0\}\,. \end{aligned}$$

Then both $\mathcal{S}_+^n$ and $\mathcal{S}_{++}^n$ are convex sets.

Proof

For all $A_1, A_2 \in \mathcal{S}_+^n$, let $\theta \in [0, 1]$ and $\bar{\theta} = 1 - \theta$,

1. it's easy to verify that $\theta A_1 + \bar{\theta} A_2$ is symmetric.
2. $\forall v \in \mathbb{R}^n$, $v^\top(\theta A_1 + \bar{\theta} A_2)v = \theta(v^\top A_1 v) + \bar{\theta}(v^\top A_2 v) \geq 0,$.

Example (*Euclidean balls*)

Given $c \in \mathbb{R}^n$, the *Euclidean ball*

$$\{x \mid \|x - c\|_2 \leq r, x \in \mathbb{R}^n\}$$

is a convex set for any $r \in \mathbb{R}_{\geq 0}$.

Proof

For any two points x, y in $\{x \mid \|x - c\|_2 \leq r, x \in \mathbb{R}^n\}$,

$$\begin{aligned} & \|\theta x + \bar{\theta} y - c\|_2 \\ = & \|\theta(x - c) + \bar{\theta}(y - c)\|_2 \\ \leq & \|\theta(x - c)\|_2 + \|\bar{\theta}(y - c)\|_2 \\ = & \theta\|x - c\|_2 + \bar{\theta}\|y - c\|_2 \\ \leq & r\,. \end{aligned}$$

In fact, note that we do not need the norm function to be L^2-norm. We only use the *triangle inequality* and the *absolute homogeneity* in the proof. Hence the norm balls defined by other norm functions are also convex sets.

Convexity-preserving operations

Example (*Ellipsoid*)

The *Ellipsoid* in $\mathbb{R}^2$

$$E = \left\{ (x_1, x_2)^\mathsf{T} \mid \frac{x_1^2}{\lambda_1^2} + \frac{x_2^2}{\lambda_2^2} \leq 1 \right\}$$

is convex.

Why? An idea is to define a norm and the ellipsoid can be viewed as a norm ball. The other viewpoint is that, an ellipsoid is the image of a ball under a linear (or affine) map. To see this, note that

$$\|\boldsymbol{x}\|_2 \leq 1 \iff \boldsymbol{x}^\mathsf{T}\boldsymbol{x} \leq 1 \iff \boldsymbol{\Lambda x} \in E, \quad \text{where } \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

In general, given an invertible $Q \in \mathbb{R}^{n\times n}$, the set $\{\boldsymbol{x} \mid \boldsymbol{x}^\top Q^\top Q\boldsymbol{x} \le 1\}$ gives an ellipsoid.

Now we show that an affine map is a *convexity-preserving operation*.

Proposition

Suppose $C \subseteq \mathbb{R}^n$ is a convex set, $f : \mathbb{R}^n \to \mathbb{R}^m$ is an affine map. Then

$$f(C) \triangleq \{f(\boldsymbol{x}) \mid \boldsymbol{x} \in C\}$$

is convex.

Proof

Without loss of generality, assume $f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ for some $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Then for all $\boldsymbol{y}_1, \boldsymbol{y}_2 \in f(C) = \boldsymbol{A}C + \boldsymbol{b}$, there exists $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$ such that $\boldsymbol{y}_1 = f(\boldsymbol{x}_1)$ and $\boldsymbol{y}_2 = f(\boldsymbol{x}_2)$.
For all $\theta \in \mathbb{R}$, since C is convex, $\theta\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2$ is also in C. Therefore,

$$\theta\boldsymbol{y}_1 + \bar{\theta}\boldsymbol{y}_2 = \theta(\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}) + \bar{\theta}(\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}) = \boldsymbol{A}(\theta\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2) + \boldsymbol{b} \in \boldsymbol{A}C + \boldsymbol{b}\,,$$

which yields that $f(C)$ is also convex.

Proposition (*Convexity-preserving operations*)

The following operations preserve the convexity:

- (*Affine map*) If C is convex, f is an affine map, then $f(C)$ is convex.
- (*Intersection*) If C and D are both convex, then $C \cap D$ is also convex.
  - This property works for infinite sets intersection.
  - Unfortunately, union is **not** a convexity-preserving operation.
- (*Cartesian product*) If $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^m$ are both convex, then their cartesian product

  $$C \times D \triangleq \{(x_1, x_2) \mid x_1 \in C, x_2 \in D\}$$

  is also convex.
- (*Minkowski addition*) If $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^n$ are both convex, then their Minkowski sum

$$C + D \triangleq \{x_1 + x_2 \mid x_1 \in C, x_2 \in D\}$$

is also convex.

- $C - D$ is also convex.

Polyhedron and polytope

Definition (*Polyhedron and polytope*)

A polyhedron (多面体) is the intersection of some halfspaces:

$$P = \{\boldsymbol{x} \mid \forall i, \boldsymbol{w}_i^\top \boldsymbol{x} + b_i \le 0\}.$$

A polytope (多胞体) is a bounded polyhedron.

Tip

- Affine sets are polyhedra. (Because $\boldsymbol{w}^\top \boldsymbol{x} + b = 0$ is equivalent to $\boldsymbol{w}^\top \boldsymbol{x} + b \le 0 \wedge \boldsymbol{w}^\top \boldsymbol{x} + b \ge 0$.)
- Halfspaces are polyhedra.
- Polyhedra are convex sets.

In particular, we define the *simplex* (单纯形) as "simplest" polytope:

- the 0-simplex is just a point;
- the 1-simplex is a segment;
- the 2-simplex is a triangle;
- the 3-simplex is a tetrahedron;
- ……
  Specifically, a *k-simplex* is a k-dimensional polytope which is the convex hull of its $k+1$ vertices. More formally, suppose the $k+1$ points $u_0, \dots, u_k$ are *affinely independent*. Then the simplex determined by them is their convex hull

$$C = \left\{\theta_0 u_0 + \cdots \theta_k u_k \mid \sum_{i=0}^{k} \theta_i = 1 \text{ and } \theta_i \ge 0 \text{ for all } i = 0, 1, \dots, k\right\}.$$

The *standard simplex* or *probability simplex* is the k dimensional simplex in

$\mathbb{R}^{k+1}$ whose $k+1$ vertices are the $k+1$ standard unit vectors in $\mathbb{R}^{k+1}$. Namely, the standard k-simplex is given by

$$\Delta_k \triangleq \left\{ \boldsymbol{x} = (x_0, \ldots, x_k)^\mathsf{T} \mid \sum_{i=0}^{k} x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i = 0, 1, \ldots, k \right\}.$$

For example, the standard 2-simplex is the triangle whose vertices are $(0,0,1)^\mathsf{T}$, $(0,1,0)^\mathsf{T}$ and $(1,0,0)^\mathsf{T}$.

Question

Why are simplexes polyhedra?

Suppose $S = \text{conv}(\boldsymbol{u}_0, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) \subseteq \mathbb{R}^n$ is a n-simplex. Then $\boldsymbol{x} \in S$ if and only if there exists $\theta_0, \theta_1, \ldots, \theta_n$ such that $\sum_{i=0}^{n} \theta_i \boldsymbol{u}_i = \boldsymbol{x}$, $\sum_{i=0}^{n} \theta_i = 1$ and $\theta_i \geq 0$ for all $i = 0, 1, \ldots, n$. Equivalently, we have

$$\boldsymbol{x} = \boldsymbol{u}_0 + \sum_{i=1}^{n} \theta_i (\boldsymbol{u}_i - \boldsymbol{u}_0).$$

Now let $\boldsymbol{y} \triangleq (\theta_1, \theta_2, \ldots, \theta_n)^\mathsf{T}$ and $\boldsymbol{B} = (\boldsymbol{u}_1 - \boldsymbol{u}_0, \boldsymbol{u}_2 - \boldsymbol{u}_0, \ldots, \boldsymbol{u}_n - \boldsymbol{u}_0) \in \mathbb{R}^{n \times n}$. Clearly $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}$. Thus, S can be equivalently written as

$$\begin{aligned} S &= \left\{ \theta_0 \boldsymbol{u}_0 + \ldots + \theta_n \boldsymbol{u}_n \mid \sum_{i=0}^{n} \theta_i = 1, \theta_i \geq 0 \right\} \\ &= \left\{ \boldsymbol{u}_0 + \boldsymbol{B}\boldsymbol{y} \mid \sum_{i=1}^{n} y_i \leq 1, y_i \geq 0 \right\}. \end{aligned}$$

Note that $\boldsymbol{u}_1 - \boldsymbol{u}_0, \ldots, \boldsymbol{u}_n - \boldsymbol{u}_0$ are n linearly independent vectors (since $\boldsymbol{u}_0, \ldots, \boldsymbol{u}_n$ are affinely independent). So $\boldsymbol{B}$ has full rank and is invertible. Let $\boldsymbol{A} = \boldsymbol{B}^{-1}$. For any $\boldsymbol{x} \in S$, $\boldsymbol{x} = \boldsymbol{u}_0 + \boldsymbol{B}\boldsymbol{y}$ for some $\boldsymbol{y}$. Thus $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{u}_0 + \boldsymbol{B}\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{u}_0 + \boldsymbol{y}$, which yields that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}\boldsymbol{u}_0$. Note that the constraints for $\boldsymbol{y}$ are $\sum y_i \leq 1$ and $y_i \geq 0$. Denote $\boldsymbol{A}$ by

$$\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)^\mathsf{T} = \begin{pmatrix} \boldsymbol{a}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{a}_n^\mathsf{T} \end{pmatrix}.$$

We obtain that $y_i = \boldsymbol{a}_i^\top \boldsymbol{x} - \boldsymbol{a}_i^\top \boldsymbol{u}_0$. Overall, S can be written as

$$S = \left\{ \boldsymbol{x} \mid \sum_{i=1}^{n} (\boldsymbol{a}_i^\top \boldsymbol{x} - \boldsymbol{a}_i^\top \boldsymbol{u}_0) \leq 1, \text{ and } \boldsymbol{a}_i^\top \boldsymbol{x} - \boldsymbol{a}_i^\top \boldsymbol{u}_0 \geq 0 \text{ for all } i = 1, 2, \ldots, n \right\},$$

which gives that S is a polytope.
In fact, note that $\boldsymbol{a}_i^\top (\boldsymbol{u}_i - \boldsymbol{u}_0) = 1$ and $\boldsymbol{a}_i^\top (\boldsymbol{u}_j - \boldsymbol{u}_0) = 0$ for all $i \neq j$. This argument has a simple geometric explanation.