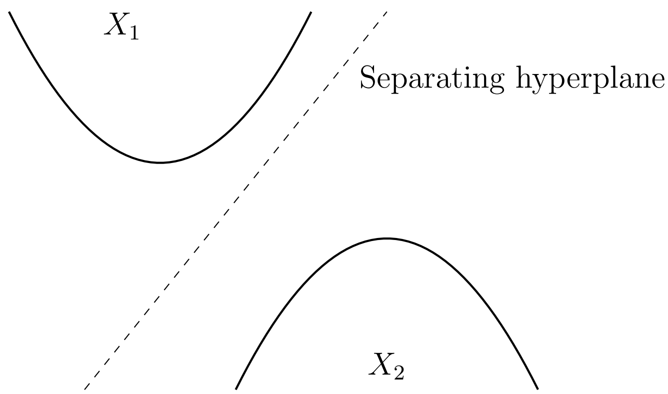
IIT Hyderabad **Scribed By: Vaibhav Agrawal (2021101019)**
- **Akshat Sanghvi (2021101094)**

Optimization Methods(CS1.404) **Lecture #8**
Instructor: Dr. Naresh Manwani 3^{rd} **Feb 2025**

Outline. *This scribe examines various properties of convex sets, particularly the separating hyperplane theorem and supporting hyperplane theorem, along with their various applications.*

1 Separating Hyperplane Theorem



1.1 Introduction

The Separating Hyperplane Theorem is a fundamental result in convex analysis with applications in optimization, machine learning, and other fields. It states that if two convex sets are disjoint, then there exists a hyperplane that separates them. This theorem has both geometric and analytical interpretations, making it a powerful tool for understanding the relationship between convex sets.

1.2 Statement

Theorem: Let X_1 and X_2 be two disjoint, non-empty convex subsets of $\mathbb{R}^d$. Then there exists a non-zero vector $w \in \mathbb{R}^d$ and a scalar b such that

$$w^T x \geq b \quad \forall x \in X_1$$

and

$$w^T x \leq b \quad \forall x \in X_2$$

Thus, the hyperplane $w^T x = b$ separates X_1 and X_2.

1.3 Proof

Proof: Consider the distance minimization problem:

$$\min_{x_1 \in X_1, x_2 \in X_2} \|x_1 - x_2\|$$

Since X_1 and X_2 are closed and non-empty, and at least one is bounded, this problem has a solution. Let $\bar{x}_1 \in X_1$ and $\bar{x}_2 \in X_2$ be such that

$$\|\bar{x}_1 - \bar{x}_2\| = \inf_{x_1 \in X_1, x_2 \in X_2} \|x_1 - x_2\| > 0$$

since X_1 and X_2 are disjoint. Define $w = \bar{x}_1 - \bar{x}_2$. Since the sets are disjoint, $w \neq 0$. Note that $\bar{x}_1$ is the closest point to X_2 not in X_2, and vice versa. Thus, using the closest point property:

$$\begin{aligned}
(\bar{x}_1 - \bar{x}_2)^T (x_1 - \bar{x}_1) \geq 0 \\
(\bar{x}_2 - \bar{x}_1)^T (x_2 - \bar{x}_2) \geq 0
\end{aligned}$$

Let $b = \frac{1}{2} w^T (\bar{x}_1 + \bar{x}_2)$. Then for any $x_1 \in X_1$:

$$\begin{aligned}
w^T x_1 - b &= (\bar{x}_1 - \bar{x}_2)^T x_1 - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)^T (\bar{x}_1 + \bar{x}_2) \\
&= (\bar{x}_1 - \bar{x}_2)^T (x_1 - \bar{x}_1) + \frac{1}{2} (\bar{x}_1 - \bar{x}_2)^T (\bar{x}_1 - \bar{x}_2) \\
&\geq 0
\end{aligned}$$

Similarly, for any $x_2 \in X_2$:

$$\begin{aligned}
w^T x_2 - b &= (\bar{x}_1 - \bar{x}_2)^T x_2 - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)^T (\bar{x}_1 + \bar{x}_2) \\
&= (\bar{x}_1 - \bar{x}_2)^T (x_2 - \bar{x}_2) - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)^T (\bar{x}_1 - \bar{x}_2) \\
&\leq 0
\end{aligned}$$

Thus, the hyperplane $w^T x = b$ separates X_1 and X_2. $\square$

1.4 Strong Separation

If the sets X_1 and X_2 are closed, then they are strongly separable if there exists a hyperplane $w^T x = b$ and an $\epsilon > 0$ such that:

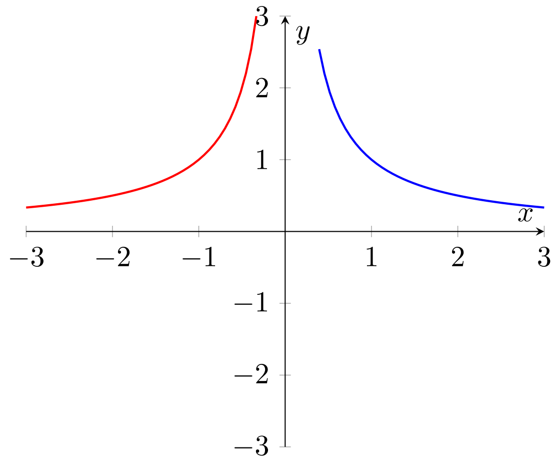
$$\begin{aligned} w^T x &\geq b + \epsilon \quad \forall x \in X_1 \\ w^T x &\leq b - \epsilon \quad \forall x \in X_2 \end{aligned}$$

1.5 Strict Separation

If the sets X_1 and X_2 are closed, then they are strictly separable if there exists a hyperplane $w^T x = b$ such that

$$\begin{aligned} w^T x &> b \quad \forall x \in X_1 \\ w^T x &< b \quad \forall x \in X_2 \end{aligned}$$

Geometric Idea

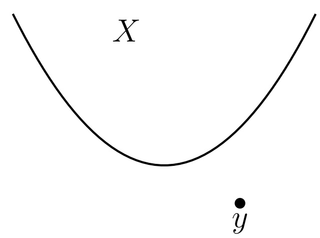


These curves represent the following sets:

$$X_1 = \{(x, y) | y = \frac{1}{x}\}$$
$$X_2 = \{(x, y) | y = \frac{-1}{x}\}$$

These curves are *not* strongly separable. To demonstrate this, we use an argument based on symmetry. If these are strongly separable, then using symmetry, $x = 0$ represents a separating hyperplane, i.e., $w = 1, b = 0$ is a separating hyperplane. Let's say ϵ satisfies the above conditions, which implies that $x \geq \epsilon \ \ \forall x \in X_1$ and $x \leq -\epsilon \ \ \forall x \in X_2$. However, this is not true for the curves described above. Hence $x = 0$ is not a separating hyperplane. Hence these functions are not strongly separable. However, for $w = 1, b = 0$, we see that $w^T x > 0 \quad \forall x \in X_1$ and $w^T < 0 \quad \forall x \in X_2$. Thus, they are strictly separable.

1.6 Interesting result



Theorem: Let $X \subset \mathbb{R}^d$ be a closed convex set and $y \notin X$. Then, there exists a hyperplane that strictly separates X and y.

Proof: Let $z \in X$ such that z is the closest point in X to y. Using the closest point theorem, we have,

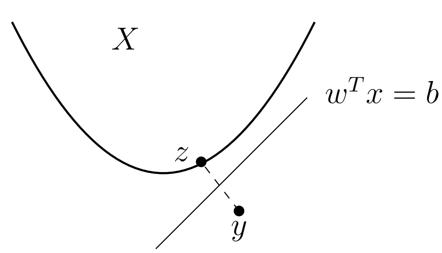
$$(y - z)^T (x - z) \leq 0 \quad \forall x \in X$$

Let $w = y - z$, thus we have $z = y - w$. We can write,

$$\begin{aligned} w^T (x - z) &\leq 0 \\ w^T (x - (y - w)) &\leq 0 \\ w^T (x - y + w) &\leq 0 \\ w^T x - w^T y + ||w||^2 &\leq 0 \end{aligned}$$

Rearranging the terms gives us,

$$w^T x \leq w^T y - ||w||^2 < w^T y \tag{1}$$



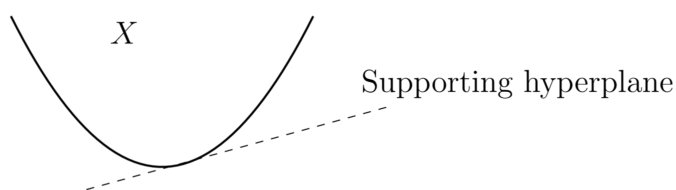
Let us choose $b = w^T(\frac{y+z}{2})$. Using $w = y - z$, we can write,

$$\begin{aligned} b &= \frac{w^T(y+y-w)}{2} \\ &= w^T y - \frac{||w||^2}{2} \end{aligned}$$

We can re-write (1) as follows:

$$\begin{aligned} w^T x \leq w^T y - ||w||^2 < w^T y - \frac{||w||^2}{2} < w^T y \\ w^T x \leq w^t y - ||w||^2 < b < w^T y \\ w^T x < b < w^T y \quad \square. \end{aligned}$$

2 Supporting Hyperplane Theorem



2.1 Statement

Theorem: Let C be a closed convex set in $\mathbb{R}^n$, and let x_0 be a point on the boundary of C. Then there exists a non-zero vector $a \in \mathbb{R}^n$

$$a^T x \leq a^T x_0 \quad \text{for all } x \in C \quad \textbf{and}$$

$$\{x : a^T x = a^T x_0\} \text{ is a supporting hyperplane for C}$$

In other words, there is a hyperplane that passes through x_0 and does not intersect the interior of the set C.

2.2 Proof

Let C be a convex set in $\mathbb{R}^n$, and let $y^* \in \partial C$ be a boundary point of C. Consider a sequence $\{y_k\} \notin \text{interior}(C)$ such that $y_k \to y^*$.

From convexity and the **Closest Point Theorem**, we have:

$$(x - x_k)^T (y_k - x_k) \leq 0, \quad \forall x \in C.$$

where $x_k \in C$ is the closest point corresponding to y_k.

Define the direction vectors:

$$v_k = \frac{y_k - x_k}{\|y_k - x_k\|},$$

Dividing by $\|y_k - x_k\|$, we obtain:

$$v_k^T(x - x_k) \leq 0, \quad \forall x \in C.$$

Since v_k is a unit vector (i.e., $\|v_k\| = 1$), the sequence $\{v_k\}$ is bounded. By the **Bolzano–Weierstrass theorem**, there exists a convergent subsequence $\{a_{i_k}\}$ such that:

$$\lim_{i \to \infty} a_{i_k} = \bar{a}.$$

For each i, we have:

$$a_{i_k}^T(x - x_{i_k}) \leq 0, \quad \forall x \in C.$$

Taking the limit as $i \to \infty$, we get:

$$\bar{a}^T(x - y^*) \leq 0, \quad \forall x \in C.$$

Defining $b = \bar{a}^T y^*$, we conclude that the hyperplane:

$$H = \{x \mid \bar{a}^T x = b\}$$

is a supporting hyperplane at y^*, satisfying:

$$\bar{a}^T x \leq b, \quad \forall x \in C. \quad \square$$

Therefore there exists $x_0 \in \partial C$ such that $a^T x \leq a^T x_0 \quad \text{for all } x \in C$

Bolzano-Weierstrass Theorem: *Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.*
Supporting Hyperplane Theorem: *The closest point in a closed convex set to an external point is unique.*

3 Applications

Support Vector Machines (SVMs)

In machine learning, SVMs use the Separating Hyperplane Theorem to classify data by finding the hyperplane that maximizes the margin between two classes in a feature space. This hyperplane is called the maximum-margin hyperplane and is critical for improving the generalization of the classifier.

Dimensionality Reduction of Features

Separating hyperplanes help define discriminative subspaces for algorithms such as Linear Discriminant Analysis (LDA).

Convex Loss Function Design

In neural network optimization, convex loss functions (such as hinge loss and logistic loss) leverage the Supporting Hyperplane Theorem to ensure that gradients correspond to hyperplanes supporting the level sets of the loss surface. This property guides optimization algorithms like gradient descent toward stable convergence.

4 Conclusion

The Separating Hyperplane Theorem is a fundamental result in convex analysis, establishing that two disjoint convex sets can be separated by a hyperplane. This idea is central to many fields, particularly in machine learning, where it forms the basis for algorithms like support vector machines that classify data points by finding the optimal separating boundary. The Supporting Hyperplane Theorem complements this by ensuring that for any point on the boundary of a convex set, there exists a hyperplane that touches the set at that point without intersecting its interior. This result is essential for optimization problems, as it guarantees that optimal solutions often lie on the boundaries of feasible regions. Together, these theorems provide a geometric framework that supports critical applications in areas such as optimization, machine learning, and decision-making processes.