Mathematical Tools in Computer Science, Winter 2006		
Lecture 6		
Lecturer: Nati Linial		*Scribe: Noa Eidelstein*

We start with definition and examples of convex sets. Then we give some more definitions and prove a separation theorem.

1 Convex sets

Definition 1 *A set $S \subseteq \mathbb{R}^d$ is called* **convex** *if for all* $\mathbf{x}, \mathbf{y} \in \mathbf{S}$ *and* $0 \leq \alpha \leq 1$, *holds:* $\alpha\mathbf{x} + (\mathbf{1} - \alpha)\mathbf{y} \in \mathbf{S}$.

Convex programming: Given convex set $S \subseteq \mathbb{R}^d$ and function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$, find $min_{x \in S} f(x)$.

Definition 2 *Let* $S \subseteq \mathbb{R}^d$ *and* $f : S \longrightarrow \mathbb{R}$. *A point* $x \in S$ *is* **global minima** *of* f *if* $f(x) \leq f(x')$ *for all* $x' \in S$. *A point* $x \in S$ *is* **local minima** *of* f *if there exists* $R > 0$ *such that* $f(x) \leq f(x')$ *for all* $x' \in S, ||x' - x|| \leq \mathbb{R}$ *(* R *can be very small). Similarly, we can define* **global maxima** *and* **local maxima** .

Convex domains are good since a convex function has no local minima on a convex set, other than the global optima.

Claim 3 *A convex function on a convex region has the property that local optima is global optima.*

Proof Let $S \subseteq \mathbb{R}^d$ convex set and $f : S \longrightarrow \mathbb{R}$ be a convex function. Suppose, by contradiction, that $\mathbf{z} \in \mathbf{S}$ is local minima of f, but is not global minima. This means, that there exists $\mathbf{y} \in \mathbf{S}$ such that $f(\mathbf{y}) < \mathbf{f}(\mathbf{z})$. Look at the line between $\mathbf{z}$ and $\mathbf{y}$, denoted by L. All the line is in S, by convexity of the set S. A point $\mathbf{x}$ is on the line L if $\mathbf{x} = \lambda\mathbf{y} + (\mathbf{1} - \lambda)\mathbf{z}$ and the value of $f(\mathbf{x})$ for every point on L is strictly less than $f(\mathbf{z})$, by the convexity of f. This contradict the claim that $\mathbf{z}$ is local minima since points on the line L appear in every radius R around $\mathbf{z}$. The same proof holds for maxima. ■

1.1 Important examples of convex sets

1. **Linear Space:** A linear subspace $C \subseteq \mathbb{R}^n$ is closed under addition and scalar multiplication. Equivalently C can be defined as the solution set for a homogenous linear equations $C = \{\mathbf{x} \in \mathbb{R}^{\mathbf{n}} | < \mathbf{a_i}, \mathbf{x} >= \mathbf{0}\}$ for $i = 1, ..., n$, or $C = \{\mathbf{x} : \mathbf{Ax} = \mathbf{0}\}$.

2. **Affine Space:** The set of solutions of a system of linear equations: $\{\mathbf{x} : \mathbf{Ax} = \mathbf{b}\}$ defines an affine subspace of $\mathbb{R}^n$. Equivalently $C \subseteq \mathbb{R}^n$ is an affine set C if for every $\mathbf{x}, \mathbf{y} \in \mathbf{C}$ and for every real number λ, $\lambda\mathbf{x} + (\mathbf{1} - \lambda)\mathbf{y} \in \mathbf{C}$. An affine set C is a translated linear subspace, i.e., there is a linear subspace $L \subseteq \mathbb{R}^n$ such that $C - \mathbf{x} = \mathbf{L}$, where

$C - \mathbf{x} = \{\mathbf{y} - \mathbf{x} : \mathbf{y} \in \mathbf{C}\}$. We define the dimension of the affine space C to be the dimension of the linear subspace L since the subspace L and the dimension are independent of the choice of $\mathbf{x}$.

3. **Convex hull:** Let $S \subseteq \mathbb{R}^d$. Since the intersection of convex sets is convex we define the **convex hull** of S as the smallest convex set which contains S:

$$conv(S) = \bigcap \{K : K \supseteq S, K \text{convex}\}.$$

4. **Polyhedron:** A set C is called polyhedron if it is the set of solutions of a finite system of linear inequalities: $\{\mathbf{x} : \mathbf{Ax} \leq \mathbf{c}\}$. An alternative definition of a polyhedron is the intersection of finitely many half spaces.

5. **Polytope:** A set $C \subseteq \mathbb{R}^n$ is called a polytope if it is the convex hull of a finite subset $\{\mathbf{y_1}, ..., \mathbf{y_k}\} \subset \mathbb{R}^\mathbf{n}$, i.e., $C = \{\mathbf{y} \in \mathbb{R}^\mathbf{n} | \mathbf{y} = \sum_{\mathbf{i=1}}^{\mathbf{k}} \lambda_\mathbf{i} \mathbf{y_i},\ \lambda_\mathbf{i} \geq \mathbf{0} \text{ and } \sum_{\mathbf{i=1}}^{\mathbf{k}} \lambda_\mathbf{i} = \mathbf{1}\}$. We denote C by $CONV(\mathbf{y_1}, ..., \mathbf{y_k})$. A basic theorem in discrete geometry states that a set is polytope if and only if it is a bounded polyhedron.

6. **Cone:** A set $C \subseteq \mathbb{R}^n$ is a cone if for every $\mathbf{x}, \mathbf{y} \in \mathbf{C}$ and non-negative reals $\lambda, \mu \in \mathbb{R}$ holds $\lambda \mathbf{x} + \mu \mathbf{y} \in \mathbf{C}$. The cone generated by a finite subset $Y = \{\mathbf{y_1}, ...\mathbf{y_k}\} \subset \mathbb{R}^\mathbf{n}$ is $CONE(Y) = \{\mathbf{y} \in \mathbb{R}^\mathbf{n} | \mathbf{y} = \sum_{\mathbf{i=1}}^{\mathbf{k}} \lambda_\mathbf{i} \mathbf{y_i},\ \lambda_\mathbf{i} \geq \mathbf{0}\}$.

   Examples of cones:

   - The non-negative orthant $\mathbb{R}^d$: $\{\mathbf{x} \in \mathbb{R}^\mathbf{n} : \mathbf{x} \geq \mathbf{0}\}$. Clearly this is the same as $CONE(\mathbf{e_1}, ..., \mathbf{e_n})$ where $\mathbf{e_i} = (\mathbf{0}, .., \mathbf{0}, \mathbf{1}, \mathbf{0}, ..., \mathbf{0})$.
   - The set PSD_n of the symmetric Positive Semi Definite $n \times n$ matrices: $PSD_n = \{A \in M_{n \times n}(\mathbb{R}) : A^\top = A,\ \mathbf{uAu}^\top \geq \mathbf{0} \text{ for every } \mathbf{u} \in \mathbb{R}^\mathbf{n}\}$. C is a cone since for all $A, B \in PSD_n$, $\lambda, \mu \geq 0$, and $\mathbf{u} \in \mathbb{R}^\mathbf{n}$ holds: $\mathbf{u}(\lambda \mathbf{A} + \mu \mathbf{B})\mathbf{u}^\top = \lambda(\mathbf{uAu}^\top) + \mu(\mathbf{uBu}^\top) \geq \mathbf{0}$.
   - **The cut cone:**

$$C_n = \{A \in M_{n \times n}(\mathbb{R}) : \exists \mathbf{x_1}, ..., \mathbf{x_n} \in \mathbb{R}^\mathbf{N} \text{ for some } \mathbf{N} \text{ with } \mathbf{a_{ij}} = ||\mathbf{x_i} - \mathbf{x_j}||_\mathbf{1}\}$$

     C_n is a cone: Let $A \in C_n$ such that $a_{ij} = ||\mathbf{x_i} - \mathbf{x_j}||_\mathbf{1}$, and $\alpha \geq 0$. Then $\alpha A \in C_n$, since $(\alpha A)_{i,j} = ||\alpha \mathbf{x_i} - \alpha \mathbf{x_j}||_\mathbf{1}$. Take $B \in C_n$ such that $b_{ij} = ||\mathbf{y_i} - \mathbf{y_j}||_\mathbf{1}$. Then $A + B \in C_n$. Since we take $\mathbf{z_i} = (\mathbf{x_i}, \mathbf{y_i})$ the concatenation of the vectors $\mathbf{x_i}$ and $\mathbf{y_i}$. Then $||\mathbf{z_i} - \mathbf{z_j}||_\mathbf{1} = ||\mathbf{x_i} - \mathbf{x_j}||_\mathbf{1} + ||\mathbf{y_i} - \mathbf{y_j}||_\mathbf{1} = \mathbf{a_{i,j}} + \mathbf{b_{i,j}}$. Note the we may simply consider vectors in $\mathbb{R}^\infty$. Though we will not prove it the whole construction can be carried out with vectors in $\mathbb{R}^N$ where $N \geq \begin{pmatrix} n \\ 2 \end{pmatrix}$.

7. A basic theorem in convexity which we will not prove states that every convex set in $\mathbb{R}^n$ can be described as a Minkovsky sum of a cone and a polytope. formally A convex set

$$C = \{\mathbf{x} + \mathbf{y} | \mathbf{x} \in \mathbf{CONV}(\mathbf{x_1}, ..., \mathbf{x_{k_1}}),\ \ \mathbf{y} \in \mathbf{CONE}(\mathbf{y_1}, ..., \mathbf{y_{k_2}})\}$$

8. **A Ball:** $\{\mathbf{x} \in \mathbb{R}^\mathbf{n} : ||\mathbf{x}|| = \mathbf{R}\}$, for some $R > 0$.

9. **Elipsoid:** the image of a ball under an affine transformation $\mathbf{x} \rightarrow \mathbf{A}\mathbf{x} + \mathbf{b}$. An elipsoid is a set of the form $E = \{\mathbf{x} \in \mathbb{R}^{\mathbf{d}} : (\mathbf{x} - \mathbf{x_0})\mathbf{B}(\mathbf{x} - \mathbf{x_0})^{\mathbf{t}} \leq \mathbf{1}\}$, where B is a PSD ($B = AA^t$ and then E is the image of the unit ball under the transformation $\mathbf{x} \rightarrow \mathbf{A}^{-\mathbf{1}}(\mathbf{x} + \mathbf{x_0})$).

10. In general, an image of a convex set under an affine transformation, is convex, since affine transformation preserves convexity.

2 Seperating theorems

We denote by $H_{\mathbf{a},\alpha}$ the one dimensional affine space $H_{\mathbf{a},\alpha} = \{\mathbf{x} :< \mathbf{x}, \mathbf{a} >= \alpha\}$. This hyperplane divides $\mathbb{R}^d \backslash H_{\mathbf{a},\alpha}$ to two halfplanes: $H^+_{\mathbf{a},\alpha} = \{\mathbf{x} :< \mathbf{x}, \mathbf{a} > > \alpha\}$ and $H^-_{\mathbf{a},\alpha} = \{\mathbf{x} :< \mathbf{x}, \mathbf{a} > < \alpha\}$. We also denote $\overline{H^+_{\mathbf{a},\alpha}} = \{\mathbf{x} :< \mathbf{x}, \mathbf{a} > \geq \alpha\}$ and $\overline{H^-_{\mathbf{a},\alpha}} = \{\mathbf{x} :< \mathbf{x}, \mathbf{a} > \leq \alpha\}$ be the closure of $H^+_{\mathbf{a},\alpha}$ and $H^-_{\mathbf{a},\alpha}$ respectively.

Definition 4

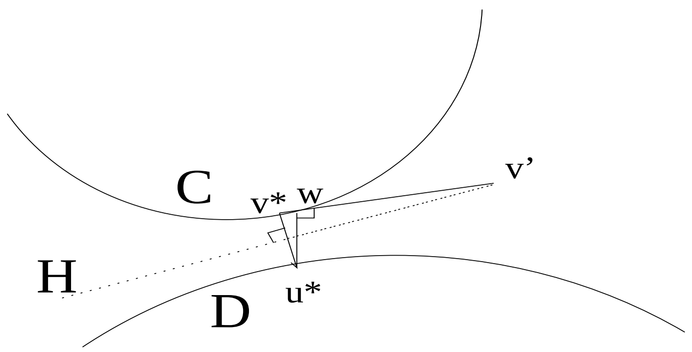
- *We say that $H_{\mathbf{a},\alpha}$ is a supporting hyperplane of a convex body $C \subset \mathbb{R}^d$ if $H_{\mathbf{a},\alpha} \cap C \neq \phi$ and $C \subseteq \overline{H^+_{\mathbf{a},\alpha}}$.*

- *Two disjoint sets $C, D subset \mathbb{R}^d$ are weakly separated if there are $\mathbf{a}, \alpha$ such that $C \subseteq \overline{H^+_{\mathbf{a},\alpha}}$ and $D \subseteq H^+_{\mathbf{a},\alpha}$.*

- *Two disjoint sets $C, D subset \mathbb{R}^d$ are strongly separated if there are $\mathbf{a}, \alpha$ such that $C \subseteq H^+_{\mathbf{a},\alpha}$ and $D \subseteq H^+_{\mathbf{a},\alpha}$.*

Theorem 5 *(**Seperating theorem**) Let $C, D \subseteq \mathbb{R}^d$ be two disjoint closed bounded convex sets. Then, there exists hyperplane $H_{\mathbf{a},\alpha}$ that separates C and D. Meaning that, $C \subseteq H^+_{\mathbf{a},\alpha}$, and $D \subseteq H^-_{\mathbf{a},\alpha}$, i.e., C, D are strongly separated.*

Proof Define $dist(C, D) = inf\{||\mathbf{u} - \mathbf{v}||_\mathbf{2} : \mathbf{u} \in \mathbf{C}, \mathbf{v} \in \mathbf{D}\}$. Since C and D are compact, this infimum is achieved, say at $\mathbf{u}^* \in \mathbf{C}, \mathbf{v}^* \in \mathbf{D}$ (i.e $||\mathbf{u}^* - \mathbf{v}^*||_\mathbf{2} = \mathbf{dist}(\mathbf{C}, \mathbf{D})$). We take H to be the perpendicular bisector to the interval $(\mathbf{u}^*, \mathbf{v}^*)$. In other words, H is the hyperplane that goes through the point $\frac{\mathbf{u}^* + \mathbf{v}^*}{2}$, and orthogonal to the interval $(\mathbf{u}^*, \mathbf{v}^*)$. We will show that H is a separating hyperplane.



Suppose, by contradiction, that H meets D and let $\mathbf{v}' \in \mathbf{D} \cap \mathbf{H}$. Consider the 2-dimensional plane spanned by the points $\mathbf{u}^*, \mathbf{v}^*, \mathbf{v}'$. By convexity of D, the interval $(\mathbf{v}^*, \mathbf{v}') \subseteq \mathbf{D}$. In the triangle $\mathbf{u}^*, \mathbf{v}^*, \mathbf{v}'$ we project $\mathbf{u}^*$ to the interval $(\mathbf{v}^*, \mathbf{v}')$. This height meets the edge $\mathbf{v}^*, \mathbf{v}'$ at the point $\mathbf{w}$. Notice that $\mathbf{w}$ must be between $\mathbf{v}^*$ and $\mathbf{v}'$ (why? H is a perpendicular bisector, thus the triangle $\mathbf{u}^*, \mathbf{v}^*, \mathbf{v}'$ is isosceles: $||\mathbf{v}^* - \mathbf{v}'|| = ||\mathbf{u}^* - \mathbf{v}'||$. Therefore the height from $\mathbf{u}^*$ stays inside the triangle). $\mathbf{v}^*, \mathbf{v}' \in \mathbf{D}$, $\mathbf{w} \in (\mathbf{v}^*, \mathbf{v}')$, and by convexity, $\mathbf{w} \in \mathbf{D}$. Now $\mathbf{u}^* \in \mathbf{C}, \mathbf{w} \in \mathbf{D}$ and their distance is shorter then the distance between $\mathbf{u}^* \in \mathbf{C}, \mathbf{v}^* \in \mathbf{D}$ since the height $(\mathbf{u}^*, \mathbf{w})$ is shorter then the side $(\mathbf{u}^*, \mathbf{v}^*)$. This contradicts the definition of $\mathbf{u}^*, \mathbf{v}^*$ as the points where the minimal distance between C, D is achieved. So we proved that $D \cap H = \phi$, and this means that D is in one side of H, say $D \subseteq H^-$. Similarly, $C \subseteq H^+$. ■

Are the assumptions on compactness really necessary? The following example shows that to guarantee strong separation we need to assume more than C, D are closed: Take $C = \{(x, y) : x, y \geq 0, xy \geq 1\}$, and $D = \{(x, y) : y \leq 0\}$. C and D are disjoint closed convex sets in $\mathbb{R}^2$. It is easy to see that $H = \{(x, y) : y = 0\}$ is the only separating hyperplane with $C \subseteq \overline{H^+}$, and $D \subseteq \overline{H^-}$. However $\overline{D} = \overline{H^-}$.

A variation of the above proof which we will not show:

Theorem 6 *(***Second version of the seperating theorem***) Let $C \in \mathbb{R}^d$ be a compact set, and $D \subseteq \mathbb{R}^d$ convex set that does not intersect C. Then, there exists hyperplane H that strongly separates C and D.*

2.1 Separation theorems in Linear Programming

How can we know that a system $A\mathbf{x} \leq \mathbf{b}$ is not satisfiable? next lecture we will use the separating theorem in order to prove the following theorem.

Theorem 7 *(***Farkas lemma, Duality theorem in linear programming***) For a matrix $A \in M_{m \times n}(\mathbb{R})$ and a vector $\mathbf{b} \in \mathbb{R}^{\mathbf{n}}$ the system $A\mathbf{x} \leq \mathbf{b}$ is not satisfiable if and only if there exists a non-negative vector $\mathbf{u} \in \mathbb{R}^{\mathbf{m}}$ such that $\mathbf{uA} = \mathbf{0}$ and $< \mathbf{u}, \mathbf{b} >< \mathbf{0}$.*

This theorem will be proved next lecture. At the moment we only observe that if such a vector $\mathbf{u}$ exists, then the linear system is infeasible.

Definition 8 *Let $C, D \subseteq \mathbb{R}^d$ sets, and $\alpha, \beta \in \mathbb{R}$. Define* **Minkovsky sum** *of C and D as*

$$\alpha C + \beta D = \{\alpha x + \beta y : x \in C, y \in D\}.$$

Notice that C and D are disjoint iff $0 \notin C - D$. We will use this fact in order to prove Farkas lemma.