Lecture 4: Convex Sets

Rajat Mittal *

IIT Kanpur

This lecture will focus on convex combinations and sets which arise in the study of linear programming. We will look at halfspaces, hyperplans, polytopes/polygons and cones. The lecture will finish with the advantages of considering these convex sets.

The train of thought starts with the observation that if x_1 and x_2 are two feasible solutions of an LP then $\alpha x_1 + \beta x_2$ is also a solution, given α, β are positive and add up to 1.

1 Linear combinations

We denote the set of real numbers as $\mathbb{R}$. Most of the time we will be working with the vector space $\mathbb{R}^n$ and its elements will be called vectors. Remember that a vector space is a set of vectors closed under addition and scalar multiplication.

First we learn how to take *interesting* combinations of a given set of vectors. For vectors $x_1, x_2, \cdots, x_k$; any point y is a linear combination of them iff

$$y = \alpha_1 x_1 + \alpha_2 x_2 \cdots + \alpha_k x_k \quad \forall i,\ \alpha_i \in \mathbb{R}$$

If we restrict α_i's to be positive then we get a conic combination.

$$y = \alpha_1 x_1 + \alpha_2 x_2 \cdots + \alpha_k x_k \quad \forall i,\ \alpha_i \geq 0 \in \mathbb{R}$$

Instead of being positive, if we put the restriction that α_i's sum up to 1, it is called an affine combination

$$y = \alpha_1 x_1 + \alpha_2 x_2 \cdots + \alpha_k x_k \quad \forall i,\ \alpha_i \in \mathbb{R},\ \sum_i \alpha_i = 1$$

When a combination is affine as well as conic, it is called a convex combination.

$$y = \alpha_1 x_1 + \alpha_2 x_2 \cdots + \alpha_k x_k \quad \forall i,\ \alpha_i \geq 0 \in \mathbb{R},\ \sum_i \alpha_i = 1$$

Exercise 1. What is the geometric shape obtained by taking linear/conic/affine/convex combination of two points in $\mathbb{R}^2$?

Exercise 2. What do you get by taking linear combination of three or in general n points?

Hint: Will you get a full dimensional vector space?

1.1 Affine sets

Let's start by defining an affine set.

Definition 1. *Affine set: A set is called affine iff for any two points in the set, the line through them is contained in the set. In other words, for any two points in the set, their affine combinations are in the set itself.*

Theorem 1. *A set is affine iff any affine combination of points in the set is in the set itself.*

* The contents of this lecture note are inspired by the books from Boyd and Vandenberghe, Dantzig and Thapa, Papadimitriou and Steiglitz.

Proof. Exercise. (Use induction) □

Exercise 3. What is the affine combination of three points?

We will try to answer this question methodically. Suppose, the three given points are x_1,x_2 and x_3. Then, any affine combination of these three points can be written as

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3, \ \ \sum_i \alpha_i = 1.$$

We can simplify the expression to,

$$\alpha_1(x_1 - x_3) + \alpha_2(x_2 - x_3) + x_3.$$

Notice that there are no constraints on α_1, α_2. We can think of this sum as,

$$x_3 + \text{plane generated by } x_1 - x_3 \text{ and } x_2 - x_3.$$

In other words, it is a linear subspace shifted by x_3.

Exercise 4. What is the dimension of the linear subspace?

Generalizing this, any affine set S can thought of as an offset x added to some vector space V. Hence,

$$S = \{v + x : \ v \in V\}.$$

You can prove that such a set is affine (exercise). Also, given an affine set V and a point inside it x, it can be shown that $S = \{v - x : \ v \in V\}$ is a vector space. In the assignment you will show that the vector space associated with an affine set does not depend upon the point chosen. So, we can define the dimension of the affine set as the dimension of the subspace.

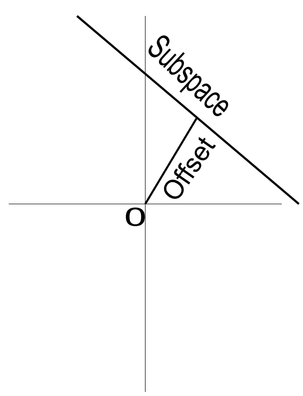

Fig. 1. Example of an affine set

Exercise 5. Show that the feasible region of a set of linear equations is affine.

1.2 Convex set

From the definition of affine sets, we can guess the definition of convex sets.

Definition 2. *A set is called convex iff any convex combination of a subset is also contained in the set itself.*

Theorem 2. *A set is convex iff for any two points in the set their convex combination (line segment) is contained in the set.*

We can prove this using induction. It is left as an exercise.

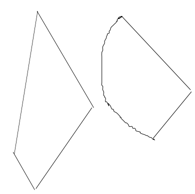

Fig. 2. Example of convex sets

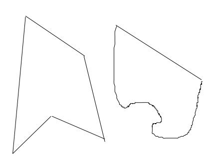

Fig. 3. Example of non-convex sets. How can we make them convex?

Convex hull: Convex hull of a set of points C (denoted $Conv(C)$) is the set of all possible convex combinations of the subsets of C. It is clear that the convex hull is a convex set.

Theorem 3. *$Conv(C)$ is the smallest convex set containing C.*

Proof. Suppose there is a smaller convex set S, then S contains C and hence all possible convex combinations of C.

Hence, S contains $Conv(C)$. But then S is not bigger than $Conv(C)$. This implies $S = Conv(C)$. □

Convex hull of S can also be thought of as the intersection of all convex sets containing S (Prove it).

Exercise 6. What is the convex hull of three vertices of a triangle? What if I add a point inside the triangle? What about outside?

2 Important affine and convex sets

2.1 Lines:

The simplest example of a non-trivial affine set is probably a line in the space $\mathbb{R}^n$. It is the set of all points y of the form

$$y = \alpha x_1 + (1 - \alpha) x_2$$

Where $x_1$and x_2 are two points in the space and $\alpha \in \mathbb{R}$ is a scalar. This gives us the unique line passing through x_1 and x_2. If we constraint α to be in $[0, 1]$, it is called the line segment between x_1 and x_2.

Exercise 7. Prove that the line segment is a convex set.

So, a point is on the line segment between x_1 and x_2 iff it is a convex combination of the given two points. Note that the condition for being a convex set is weaker than the condition for being an affine set. Hence an affine set is always convex too. Since line is an affine set, it is a convex set too. A line segment will be a convex set but not affine.

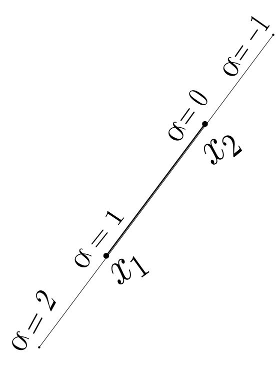

Fig. 4. A line, line segment, points on the line and α's corresponding to them

There is another way to think of a line in $\mathbb{R}^2$. You will prove in the assignment that every point on a line has same inner product with the vector perpendicular to $x_1 - x_2$. In other words, the equation of line can be written as,

$$a^T x - b = 0.$$

Where a is the vector perpendicular to $x_1 - x_2$.

Exercise 8. What is b?

2.2 Hyperplane and Halfspaces

We extend the idea of a line to hyperplanes and halfspaces. A hyperplane H is described by a vector $a \in \mathbb{R}^n$ and a number $b \in \mathbb{R}$

$$H = \{x : a^T x - b = 0\}$$

Exercise 9. Prove that a hyperplane is affine and so convex too.

Suppose we know a point x_0 on the hyperplane H. Then this equation can be changed to $a^T x = b \Rightarrow a^T x = a^T x_0,\ x_0 \in H$. We can view the hyperplane as the set of all points which have same inner product b with the vector a. This gives a very nice geometrical picture of the hyperplane, i.e., all points in H can be expressed as the sum of x_0 and a vector orthogonal to a (we can call it $a^{\perp}$). So, another definition of hyperplane is

$$H = \{x : x = x_0 + a^{\perp}, a^T a^{\perp} = 0\}$$

Note that this definition assumes that we know a point on the hyperplane. But this point is not special in any way, for any point on the hyperplane we can define the hyperplane in the same way. The vector a is called the *normal* vector of the hyperplane and b is called the *offset*. Another way to think of this hyperplane is to take the set of all vectors orthogonal to a (hyperplane, passing through origin) and offset them by the distance $\frac{b}{\|a\|}$.

We studied the generalization of a line in higher dimensional space, it was called a hyperplane. A hyperplane divides the space into two parts, $a^T x \geq b$ and $a^T x \leq b$. Geometrically, they are the two sides of the plane. Is a halfspace affine/convex?

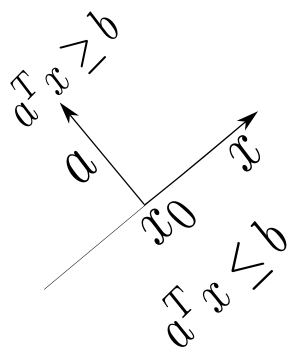

Fig. 5. Halfspace with normal vector a and offset b.

2.3 Polytopes and Polygons

We will mostly be interested in linear equations and linear inequalities as our constraints. We already discussed the set of points which satisfy the set of all constraints is called the *feasible region*. When this constraints are linear inequalities and linear equalities the feasible region is called a polytope. Mathematically, S is a polytope iff

$$S = \{x : a_i^T x - b_i \leq 0 \quad i = 1, 2, \cdots, m \text{ and } c_i^T x - d_i = 0 \quad i = 1, 2, \cdots, p\}$$

A bounded polytope in two dimensions is called a polygon.

Exercise 10. Prove that the polytope is a convex set.

Exercise 11. What are the polytopes in one dimension?

Geometrically, polytopes are intersections of hyperplanes and halfspaces. Remember that a polytope could be bounded or unbounded. Intuitively, polytopes have vertices, edges, planes and hyperplanes as their bounding surface.

A bounded polytope can be thought of as the convex hull of its vertices. Actually a bounded polytope can have an alternate definition as the convex hull of a finite set of points. The statement that these two definitions (the previous one and the original definition in terms of equalities and inequalities) are the same is known as Minkowski-Weyl theorem. The proof of this theorem is out of the scope of this course. Now we look at a special case of polytope.

We haven't defined vertices/extremal points formally till now. It is intuitively clear that a vertex is a corner of the polytope. Formally, A vertex of a polytope is the point which cannot be expressed as the convex combination of two different points in the polytope. This implies that vertex is not *inside* of any line segment joining two points in the convex set.

We saw that the convex hull of a triangle and a point inside the triangle is triangle itself. Suppose we are given a convex set and we want to find the minimal set whose convex combinations will generate the entire set.

By the definition of vertices, all vertices should be in this minimal set. It turns out that the for a bounded polytope this set (set of all vertices) is enough.

Exercise 12. Suppose $S = Conv(x_1, \cdots, x_k)$. Prove that x_i is not extremal/vertex if and only if it can be written as the convex combination of other x_j's.

2.4 Cones

We have seen sets made by vertices, lines, line segments. Now we look at sets generated by *rays*.

Exercise 13. What is a ray?

A set is called a cone iff every ray from origin to any element of the set is contained in the set. Hence, a set C is a cone iff for every $x \in C$ we have $\alpha x \in C, \alpha \geq 0$.

Note 1. A cone is *not* a set which has all possible conic combinations of all its points (Show an example). Remember the notion of conic combination. A conic combination of vectors $x_1, \cdots, x_k \in \mathbb{R}^n$ is any vector of the form $\alpha_1 x_1 + \cdots + \alpha_k x_k$ for $\alpha_1, \cdots, \alpha_k \geq 0$.

The previous paragraph implies that a cone is not necessarily convex (Give example of a cone which is not convex). A set which is a cone and is convex is called a convex cone. In this course we will mostly be concerned with convex cones.

Mathematically, a convex cone C is a cone where,

$$\forall x_1, x_2 \in C \text{ and } \alpha_1, \alpha_2 \geq 0,\ \alpha_1 x_1 + \alpha_2 x_2 \in C.$$

So, a cone is convex iff it contains all the conic combinations of its elements. Convex hulls and cones are closely related.

Exercise 14. Take x_i's as row vectors. Prove,

$$x \in Conv(x_1, x_2, \cdots, x_k) \Leftrightarrow (x, 1) \in Cone((x_1, 1), \cdots, (x_k, 1))$$

Note 2. Some authors define cones as sets closed under *positive* scalar multiplication. We have defined cones as sets closed under non-negative scalar multiplication.

Clearly the set $Cone(x_1, x_2, \cdots, x_k) := \{\alpha_1 x_1 + \cdots + \alpha_k x_k :\ \forall i\ \alpha_i \geq 0,\ x_i \in \mathbb{R}^n\}$ is a cone. It is called a finitely generated cone because it is generated by finite number of vectors. A convex finitely generated cone is also a polytope. Next theorem gives a characterization of a finitely generated cone.

Theorem 4 (Weyl). *A non-empty finitely generated convex cone is a polytope.*

Proof. Suppose the set of generators for cone C are $x_1, \cdots, x_k$. We can define a matrix X which has x_i's as the columns. Then the cone C can be written as

$$C = \{x : x = X\alpha,\ \alpha \in \mathbb{R}_+^k\}$$

Converting equalities into inequalities

$$C = \{x : x - X\alpha \leq 0,\ X\alpha - x \leq 0,\ -\alpha \leq 0\}$$

Now α can be eliminated from these inequalities using something known as Fourier-Motzkin elimination.

Lemma 1. *Let $Ax \leq b$ be a system of m inequalities in n variables. This system can be converted into another equivalent system $A'x \leq b'$ with $n-1$ variables and polynomial in m many inequalities. Here equivalent means any solution x of old system will be a solution of the new system ignoring the removed variable. Also given any solution x of new system ($A'x \leq b'$), we can find a solution (x_0, x) of old system.*

Proof. Suppose the variable to be removed is x_0. We divide all the inequalities into three sets depending upon whether the coefficient of x_0 is positive (P), negative (N) or zero (Z). Divide the inequalities in P and N by the modulus of the coefficient of x_0. The inequalities in the new system are the inequalities from Z and every inequality of the form $p_i + n_j \leq 0, p_i \in P,\ n_j \in N$.

Exercise 15. Prove that this construction works.

□

With the α eliminated from the system of equations which define the cone, we get

$$C = \{x : Ax \leq 0\}$$

Hence it is a polytope. □

Note 3. A general polytope is $Ax \leq b$ and we will see that a finitely generated cone is $Ax \leq 0$.

3 Characterization of polytope

We saw that cones and bounded polytope has two representations. One in terms of linear inequalities and one in terms of convex combinations (conic combination). There is a similar theorem for polytope.

Theorem 5. *Let there be a polytope defined by a set of inequalities, $P = \{x \in \mathbb{R}^n : Ax \leq b\}$. Then there exist vectors $x_1, \cdots, x_k \in \mathbb{R}^n$ and $y_1, \cdots, y_l \in \mathbb{R}^n$, s.t.,*

$$P = Cone(x_1, \cdots, x_k) + Conv(y_1, \cdots, y_l)$$

Note 4. The sum defined here is Minkowski's sum. For intuition, $S_1 + S_2 = x_1 + x_2 :\ x_1 \in S_1,\ x_2 \in S_2$

This is known as the *affine Minkowski-Weyl* theorem. We will not do the proof of this theorem in this course. Notice that now there are equivalent characterizations of cone/polytope/bounded polytope in terms of convex/conic hulls or linear inequalities. It is instructive to remember the special forms of linear inequalities and hulls required to make these shapes.

	Linear inequalities	Convexity
Bounded polytope	Bounded and $Ax \leq b$	$Conv(y_1, \cdots, y_n)$
Finitely generated Cone	$Ax \leq 0$	$Cone(y_1, \cdots, y_n)$
Polytope	$Ax \leq b$	$Cone(y_1, \cdots, y_n) + Conv(z_1, \cdots, z_m)$

4 Properties of convex sets

Next, we will see properties of convex sets. These properties make convex sets special and are the reason why convex optimization problems can be solved much more easily as compared to other general optimization problems.

4.1 Intersections and unions of convex sets

Suppose we are given two convex sets S_1 and S_2. What happens when we take their intersection or union. Intersection of two convex sets is convex. Consider two points in the intersection $S_1 \cap S_2$. They are contained in the individual sets S_1, S_2. So the line segment connecting them is contained in both the sets S_1, S_2 and hence in the set $S_1 \cap S_2$.

The same argument can be extended to the intersection of finite number of sets and even infinite number of sets too. Since a polytope is an intersection of halfspaces and hyperplanes (linear inequalities and linear equalities), it gives an easier proof that a polytope is convex.

But the same property does not hold true for unions. In general, union of two convex sets is not convex. To obtain convex sets from the union, we can take convex hull of the union.

Exercise 16. Draw two convex sets, s.t., there union is not convex. Draw the convex hull of the union.

4.2 Convex functions

A function is called convex if the line segment connecting any two points on the graph lies above the graph. Formally, a function $f : \mathbb{R}^n \to \mathbb{R}$ is called convex iff

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y),\ \forall x, y \in \mathbb{R}^n,\ 0 < \theta < 1$$

We have assumed the domain to be the entire $\mathbb{R}^n$ space. In general, if the domain is a convex subset of $\mathbb{R}^n$, and satisfies the above mentioned property, it is called a convex function. The canonical example of a convex function is $f(x) = x^2$. You should check that this function is convex.

Note that since the domain is convex, if we restrict the function on any line passing through the domain, the restricted function will be convex. Conversely, if the function is convex on all the lines passing through the domain, then it is convex on the whole domain.

A function f is called concave if it satisfies the above mentioned inequality in opposite direction ($\geq$ instead of $\leq$). It is clear that if f is convex then $-f$ is concave.

Exercise 17. Characterize the functions which are both convex and concave.

4.3 Relation to convex sets

A convex set is the set which contains all possible convex combinations of its points. What is the relation between convex sets and convex functions? The *epigraph* of the function is all points which lie above the graph of the function.

Formally, given a function $f : \mathbb{R}^n \to \mathbb{R}$, the epigraph is the set of all points

$$\{(y, x) : y \geq f(x),\ x \in \mathbb{R}^n,\ y \in \mathbb{R}\}$$

So, if the function is convex, then its epigraph is a convex set and vice versa. This gives a geometric interpretation of a function being convex. Opposite of epigraph, points which lie below the graph of a function, is known as *hypograph.*

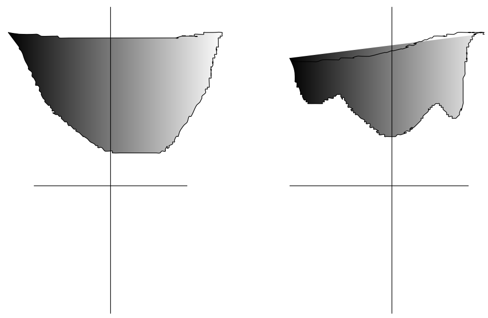

Fig. 6. Epigraph for functions

5 Optimization with convex functions

Due to the structure in the convex function, it is comparatively easier to optimize a convex function. There are essentially two properties which are very helpful for optimization.

- A local optimum is a global optimum.
- Maximum is found on the boundaries.

5.1 Global optimality and local optimality

For a general function there are two kind of optimal points. There is *global optimum*, which is the general notion of being optimal as compared to any other point on the domain. Hence a point x^* is globally optimal (say minimum) for a function f iff $f(x) \geq f(x^*)$ for any x in domain of f. In general it is very hard to find such points.

On the other hand, there is a concept of *local optimum*, where the point has minimum value among all values in the neighborhood. Hence, a point x^* is a local optimum point for a function f iff for some ϵ,

$$f(x^*) \leq f(x), \ \forall x \in B_\epsilon(x^*)$$

Here $B_\epsilon(x^*)$ is the ball of radius ϵ around x^*. As compared to global optimum, it is much easier to check for a local optimum. For instance, we can check that the derivative in every direction is positive and continuous.

The special property of convex functions is that a locally optimal point is globally optimal too. This again makes it much easier to optimize a convex function. The proof is given for a single variable function, but the theorem works for multi-variable function too.

Proof. We will use another characterization of a function being convex. A function is convex iff for any two points x, y in the domain

$$f(y) \geq f(x) + f'(x)(y - x)$$

This definition holds for functions of one variable. In the case of more variables, $f'(x)$ will be a vector and a dot product with $y - x$ can be taken.

Suppose point x^* is locally optimal. Since the function is convex, for any point y,

$$f(y) \geq f(x^*) + f'(x^*)(y - x^*).$$

This can be written as,

$$f(y) \geq f(x^*) + \left(\lim_{t \to 0} \frac{f(x^* + t(y - x^*)) - f(x^*)}{t(y - x^*)} \right) (y - x^*),$$

using the definition of first derivative. This can be simplified to,

$$f(y) \geq f(x^*) + \left(\lim_{t \to 0} \frac{f(x^* + t(y - x^*)) - f(x^*)}{t} \right).$$

Clearly the second term on the right is positive because x^* is a local optimum point. This implies, for any y in the domain,

$$f(y) \geq f(x^*)$$

So x^* is globally optimal too. □

5.2 Optimality at the boundary

For a convex function it can be shown that the maximum always lies at the extremal points of the feasible region. Similarly for a concave function the minimum is always attained at the extremal points of the feasible region.

Exercise 18. Prove the above statement.

Suppose that the feasible region is bounded. Given that a linear function is both convex and concave, the optimal for a linear program lies on its extremal points. From the previous discussion, *vertices* are the extremal points of the feasible region of LP. Hence, the optimal of an LP lies on its vertex.

Note 5. This does not mean all optimal solutions lie on vertex. Just that, there always exist at least one optimal solution on the vertex. In general, an entire edge or a face could give us the optimal solution.

One way to see this intuitively is, take the optimal point in the feasible region of an LP (say minimum). Pick any direction, at least one of the sides (out of two) will not increase the value of the objective function (because function is linear). We can reach a face using this side of the direction. Continuing this we can reach a vertex without increasing the objective value.

This is the essential idea of *Simplex method*. Suppose we are interested in finding the maximum. We start by finding a vertex, called a *basic solution*. From any vertex which is not optimal, there is a direction where the objective function will increase.

Exercise 19. Why?

Hence, we can find another vertex with greater objective value. This process can be continued till we attain a global maximum. We are not going to cover any LP/SDP solvers in this course. You are encouraged to read about them on your own.

In case the feasible region is not bounded, then there might be a direction in which we can travel (ray) and the optimum will keep increasing. In this case the optimum is unbounded.

6 Extra reading: Jensen's inequality

One of the most important properties of convex function is *Jensen's inequality.* Given a random variable X and a convex function f,

$$E(f(X)) \geq f(E(X))$$

This can be proved using the straightforward extension of the basic inequality for convex functions.

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

This innocent looking inequality can be used to prove other inequalities by using different convex functions.

– AM-GM: Since the function $\log x$ is concave (it satisfies Jensen's inequality in opposite direction). Applying Jensen's inequality on it for a random variable which takes values a and b with equal probability.

$$\begin{aligned} & \log \frac{a+b}{2} \geq \frac{\log a + \log b}{2} \\ \Rightarrow \quad & \frac{a+b}{2} \geq \sqrt{e^{\log a + \log b}} \\ \Rightarrow \quad & \frac{a+b}{2} \geq \sqrt{ab} \end{aligned}$$

This shows the arithmetic mean-geometric mean inequality.

– Suppose $x_1, \cdots, x_k$ are some positive numbers. Say $p_1, \cdots, p_k$ is a probability distribution ($\sum_i p_i = 1,\ p_i \geq 0$). Then we need to show,

$$\Pi_i x_i^{p_i} \leq \sum_i p_i x_i.$$

Proof. Assume that $x_i = e^{y_i}$ (why can we assume that?). Then the inequality turns into,

$$\Pi_i e^{\sum_i y_i p_i} \leq \sum_i p_i e^{y_i}.$$

This is Jensen's inequality for function $f(x) = e^x$. □

7 Assignment

Exercise 20. Show that the vector space associated with an affine set does not depend upon the point chosen.

Exercise 21. Prove that the line (in $\mathbb{R}^2$) between x_1 and x_2 is the set of all points p, s.t.,

$$a^T p = a^T x_1 = a^T x_2.$$

Here a is the unit vector perpendicular to $x_1 - x_2$.

Exercise 22. We have given combination interpretation (affine, convex, conic) of polytope, cones and bounded polytopes. Can you give similar interpretation for hyperplanes and halfspaces?

Exercise 23. Read about *simplex method* of solving linear programs.

Exercise 24. Given a convex set S, show that the set

$$T = \{x : Ax + b = y,\ y \in S\}$$

is also convex.

Exercise 25. A *Minkowski sum* of two sets S_1, S_2 is the set formed by taking all possible sums such that first vector is from S_1 and second vector is from S_2.

$$S_1 + S_2 = \{x : x = x_1 + x_2, x_1 \in S_1, x_2 \in S_2\}.$$

Prove that the Minkowski sum of two convex sets is convex. You can prove a stronger result too,

$$Conv(S_1 + S_2) = Conv(S_1) + Conv(S_2).$$

Where $Conv(S)$ takes the convex hull of set S.