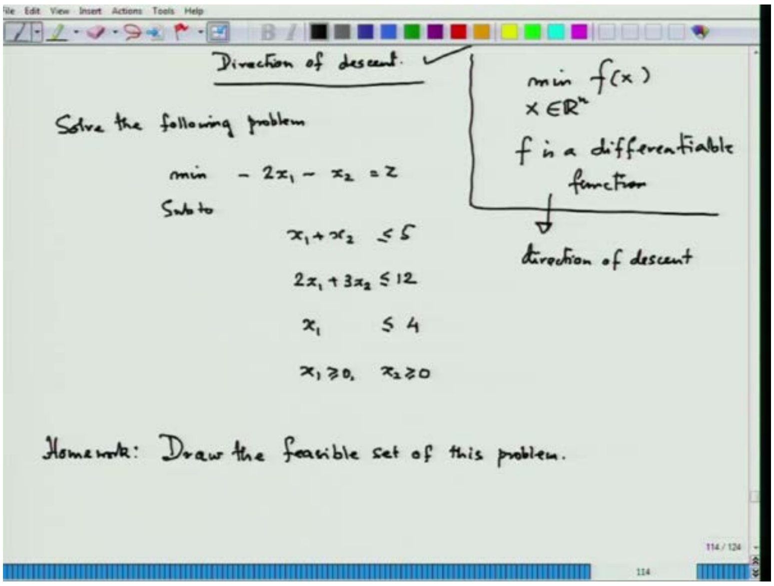
Convex Optimization
Prof. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

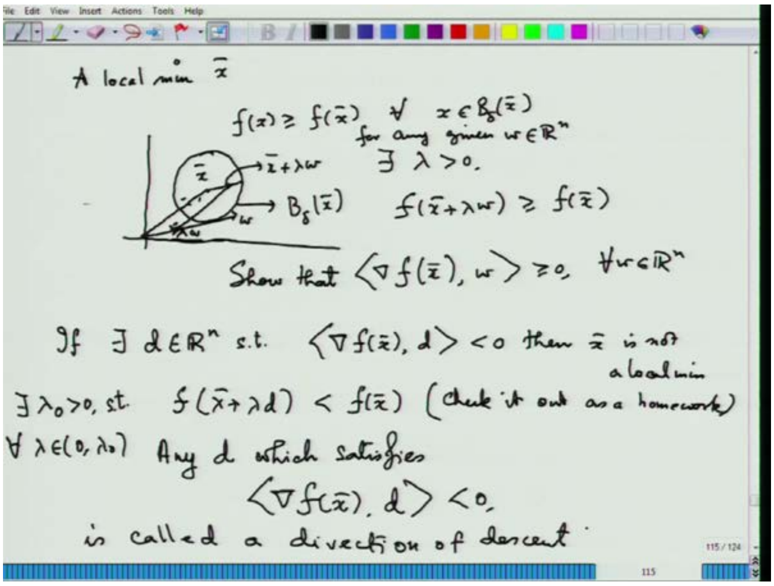
Module No. # 01
Lecture No. # 20

(Refer Slide Time: 00:31)



I began by reminding you that we were supposed to talk about the direction of descent of our differential function over the whole R n. We are trying to minimize the differential function of our whole R N and we are going to talk about the direction of descent. But (No audio from 00:36 to 00:44) means if I move along the direction, my functional value decrease, that are minimizing. The important part here is this that I have also given you home work to take this optimization problem or linear programming problem in two variables, but three constraints. And I told you to draw the feasible set of this problem. I hope you have tried it out, but after we do the direction of descent we will try to solve this problem.

(Refer Slide Time: 01:20)



So, if I have a local minima, what would I have? (No audio from 01:17 to 01:24) See, you have a local minima at a point x bar say in R 2, all the drawings are in R 2 as you know then again find the ball of radius say delta (No audio from 01:49 to 01:56) so, radius delta. Such that so, if you have x bar as a local mean then you have f of x bigger than f of x bar for all x element of b delta x bar and you know what is the ball means. So, any x in this disk would satisfy this. Now, what it means that, this is my vector x bar, if I take, I want to remove this stuff. So, here we see the drawing where x bar is a point and there be a drawn a disk around it.

So, take any direction w, w is actually I would say slightly larger possibilities. It is a direction w and this is this vector up to here is lambda times w and this when we add them x bar plus lambda w, they get of affected here, a point here which is inside the ball. So, there exists lambda greater than 0 for any w, for take for any given w. Such that f of x bar plus lambda w is bigger than f of x bar. A Taylor's expansion I would again leave you a home work. Show that f of grad f of x bar times w is greater than equal to 0 for all w are in this is exactly what you have as an optimality condition in fact, when you are talking about a local minima our differential function.

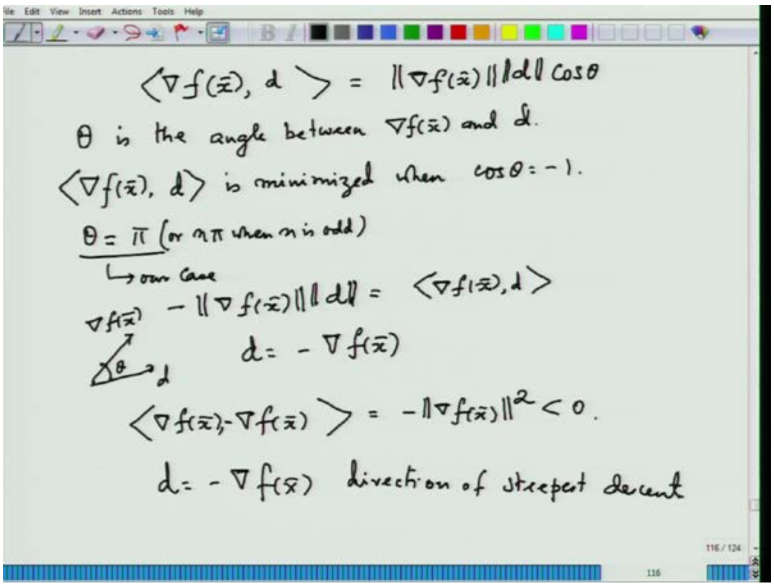
In fact, here of course, n w you will you will get grad f x bar equal to 0 because you will put w equal to minus grad f x bar so and so. But this is the basic optimality condition, what does it mean? If x bar is a local minima, then this is happening. Now, if there exists that w or say d may be is better to write d because talking about direction, if there exists d in R n, such that grad f of x bar d is strictly less than 0 then x bar is not a local mean. Of course, because if P implies q local mean implies this, then naught p which is this naught implies naught q implies naught P so, this just a logical rewriting of this thing.

So, see if I write d like this, what it gives me? So, it gives me the following, you can show b, such that for all lambda between 0 and lambda naught f of x bars plus lambda d is strictly less than f of x bar. So, how do I check as a home work? Check it out as a home work. (No audio from 05:38 to 05:46) Now, if I have done this so, which means what that I have gone to a point in R n which is x bar plus lambda d have moved along the direction d from x bar. And I have got a point, there is a lambda for which I will get a point, such that the functional value is strictly less than the functional value here.

So, again if I do not have a local minima at x bar, I am trying to move in a directions so that my function value decreases. So, any d which satisfies grad f x bar d strictly less than 0 is called a direction of descent. So, this d is called a direction of descent. (No audio from 06:37 to 06:44)

y just the using again Taylor's theorem that there exists a lambda naught greater than 0

(Refer Slide Time: 06:48)



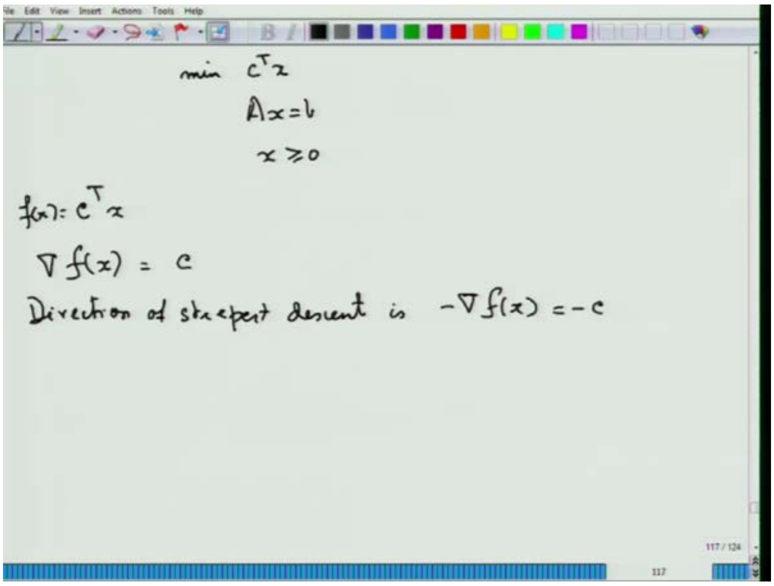
So, we will be now talking about the direction of steepest descent. By our standard notions of dot product, this is nothing but norm of f x bar this d into norm of h sorry norm of d to cosine of theta where cos theta, where theta is angle between grad f x bar and d. (No audio from 07:21 to 07:35) This is the following fact. Now, what is the minimum value, where how can I minimize this value? Suppose, this I get a d even if it is strictly less than 0, it does not matter. What is my minimum value? My minimum value of this would a be attained. So, the minimum value of this function is attained when cos theta is the most negative. So, grad f x bar d is minimized when cosine of theta is minus 1. So, when cosine of theta is minus one what I will get is theta is equal to pi, cos theta minus 1. Is theta is equal to pi or some multiple of pi, or pi, 2 pi 3 pi 4 pi whatever. Cos n theta is same as pi is minus 1 because cos theta will come down to pi and go up, pi will again come down to minus 1 at 2 pi.

So, cos n theta is minus 1 to the power n, sorry it is minus 1. So, in this case, the two vectors cos theta is minus 1, the angle between them is cos 0 is 1, cos pi is minus 1 cos 2 pi so, two pi is not is again 1 and cos 3 pi is again same as cos pi is minus 1 so, cos n theta is minus 1 to the power n. So, which means the angle between them is anyway theta is pi or any odd multiple of pi. So, if theta is pi or n pi where n is odd, but because we are talking about vectors that angles are and if you take a vector with this angle, does not matter.

So, which means that our case we require theta equal to pi, our case; because we are talking of vector same two dimensions so, theta is pi. Now, what would be then h? So, norm of grad f x bar norm of d so, this into minus 1 minus is equal to grad f x bar d. So, what should be d? You see those theta is equal to pi so, they are two vectors grad f x bar and d, angle is theta what this angle is pi grad f x bar and minor d are in opposite direction. So, d is the negative of grad f x bar, that is what is the required d for which this functional value would be minimized and this would be the value.

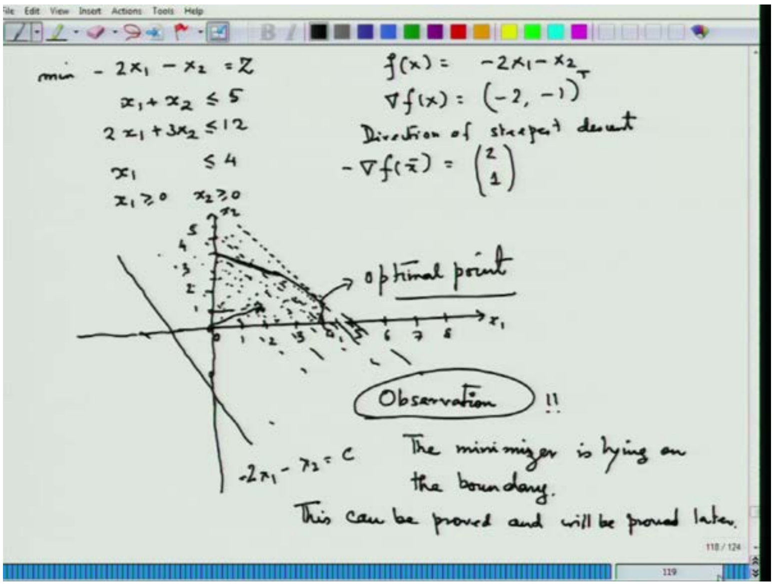
So, means now grad f x bar with a minus sign would only give me norm of grad f x bar, if f x grad f x bar is norm 0 which is why we are doing all this. This is strictly less than 0, so, it is a minus grad f x bar is a direction of this set and because it minimizes this value this particular value of d, d equal to minus grad f x bar is called the direction of steepest descent. (No audio from 11:30 to 11:40) So, once you know that which is the direction of steepest descent, you will ask the question in our setting, what is the direction of steepest descent?

(Refer Slide Time: 11:53)



So, if I have minimized c transpose x A x equal to b and x greater than equal to 0. So, if I just take this functions c transpose x and I find the gradient of f x which is c. So, direction of steepest descent of this function (No audio from 12:11 to 12:19) is minus grad f x equal to minus c. So, whichever direction is c, I will just have to move them I am the opposite direction to get a decrease in the function value. So, now we will try to solve the problem in a geometrical fashion that we had given yesterday. So, I will rewrite the problem because you might have forgotten about it. It was a few slides back.

(Refer Slide Time: 12:48)



So, minimize minus 2 x 1 minus x 2 which is my z value objective value some easy easily as a standard to writing. These are all things written in linear programming. I will give you reference, but just listen to me for the time being. (No audio from 13:06 to 13:20) So, in this case my f x is minus 2 x 1 minus x 2 so, grad of f x is minus 2 minus 1. So, direction of steepest descent.

(No audio from 13:42 to 14:03)

Now, let me draw the feasible set of this problem. (No audio from 14:07 to 14:14) So, my thing is minus 2 x 1 minus 2 x 2. So, this is my x 1 this x 2 so, 0 1 2 3 4 5 6 7 8 and so far. 1 2 3 and do not bother about this part because and so on. So, something x is less than equal to 4. So, all my x must pass through this point, no x can be chosen bigger than 4. Then there is a first thing then we look at x 1 plus x 2 less than equal to 5 so, it is passing through this, this line, anything inside this. Now, 2 x 1 plus 3 x 2 is less than 12. So, if I put here equal 2 then if I put say x 2 is equal to 0 then x 1 is 6, it is this one. And if I put x 1 is 0 and I get 4 so, it is connecting this 4 and 6.

(No audio from 15:42 to 15:51) Now, what is my feasible set? My feasible set is this one. (No audio from 15:56 to 16:06) So, now I will rub the parts which are not in the feasible set. So, you will see the complete feasible set (No audio from 16:14 to 16:27) x 1 x 2 greater than equal 0 we have chosen so, this is all feasible set.(No audio from 16:32 to 16:39) So, there is not a straight line there is a slight curve here. So, what is my direction of descent? What is my problem here? Minus 2 x 1 minus x 2 so, when x 2 is 0 so, here you see if I look at this curve so this is constant so I can put here something say this is equal to 2 3 whatever say something it will be a curve like this. And 2 1 is my direction of descent so, it is 2 and 1 2 and 1 so, not writing, this is my direction of descent.

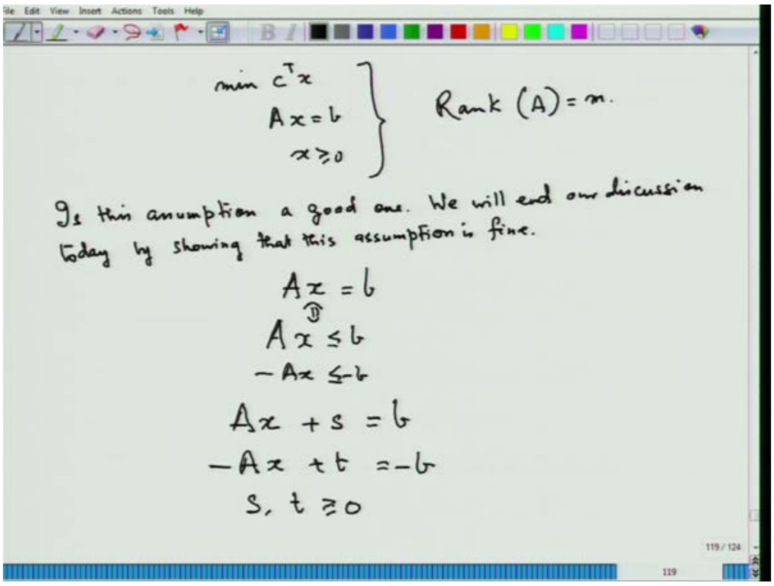
So, if I move this whole thing along it is not exactly correct here is slightly because that c of c transpose x has to be perpendicular so, it is something like this. This is my objective 2 minus 2 x 1 minus x 2 is some constant c. So, now I am moving it along this direction so I come and touch here when I put this c here. So, I am coming then I am inside the feasible zone, but I have to move along the direction of steepest descent. My if I has a move like this, my function value is decreasing. Here, you see I am still in the feasible set, but then there comes a point where I simply go out of the feasible set.

So, I have a point where here there is a bent. So, I am coming and touching at this point, then if I move a little bit off, I am outside the feasible set. So, remaining in the feasible set, this is the maximum drop in the value of ten. So, this point which waits now we will. (No audio from 18:58 to 19:06) So, I am coming this is this is my feasible set and I am coming like this. You see finally, I come in touch here this point, this bent you see here, this particular bent, this bent, this corner point. So, what happens is that as I am moving along this direction I am bringing it and I am once I am inside a feasible set I am fine, but I am that those are the values I am required to bother about the functional value of the objective for these points.

But as I move the push the line parallelly, you see the functional value of the objective the z keeps on decreasing because that is the direction of descent. And it comes and comes and comes here, here, here and then it is in a position when it passes through this point. Because I am moving it continuously, it occupies all the space here, not discretely it moves continuously; it comes to a point here when it is in this form. And at this point if I just move a little bit I am outside the feasible set and that is it.

So, then once it is in this situation, this is the only contact it has with the feasible point, there is a feasible value, feasible point which is touching it and because if you leaving this, you get the maximum descent. So, the maximum drop in the value comes, when it comes here. So, this point is my optimal point. So, which you can figure out what it is which I will not do, figure it out. It is for you to figure out, optimal point just you take see intersection of x less than 4 with this one. So, the interesting part is that this is how we are looking into the now using this direction of descent to find geometrically the minimum point, this is optimal point. And there is a review observation which you might say is too early to decide, but this observation is an important observation. (No audio from 21:18 to 21:30)

(Refer Slide Time: 22:25)



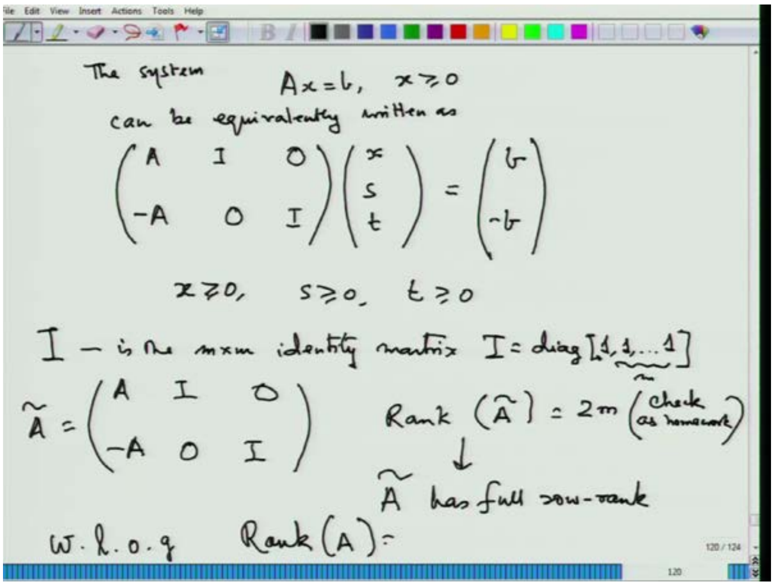
The minimizer.

(No audio from 21:32 to 21:48)

This can be proved and we will prove later. Now, let me go into certain aspects of the simplex method which is very very important. First thing to know that if you look into any book on the simplex method or any book on linear programming where simplex method is discussed, there are books on linear programming where simplex method is not discussed. So, one might say what a strange thing to say, but there are books in interior point methods where it is not discussed. For example, the book by Stephen J Wright which was published by Siam called primal dual interior point methods for linear programming.

So, if I take this my standard l p problem, a standing assumption is that rank of a is m. Now, the question is this assumption a good one, (No audio from 23:06 to 23:18) we will end our discussion today by proving this fact. That yes, it is indeed a good step, by showing that this assumption is fine. (No audio from 23:36 to 23:48) Now, look at the constraint system A x is equal to b. I can write this constraint system as follows so, this can be equivalently written as sorry, A x less than equal to b and minus a x less than equal to minus b. This is same as A is greater than equal to b so, these two combines to give you that. Now, in both the cases I can apply the law of adding a slack variable. So, I can write this as A x plus s is b is m vector and minus A x plus t is minus b where, s and t are greater than equal to 0.

(Refer Slide Time: 24:52)



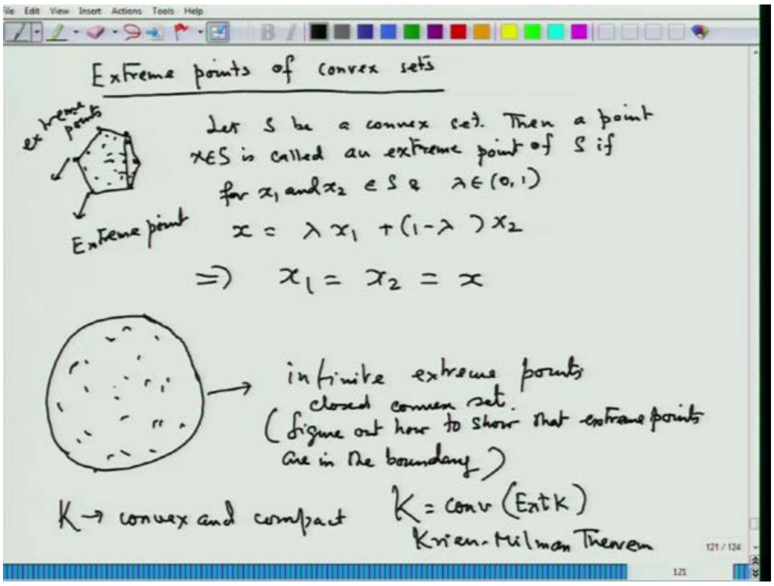
So, once you do that then the system A x equal to b can be equivalently written as.

(No audio from 24:57 to 25:32)

And add the x greater than equal to 0 so, this is the system. So, I here is the identity matrix, I is the m cross m identity matrix (No audio from 25:45 to 25:55). So, I is nothing but diagonal 1 1 m times. Now, if I take this matrix and I write this as a tilde. (No audio from 26:17 to 26:25) It is part of your home work to check up the rank of A is twice of m, check as home work. So, give me any system of A, I can always convert it into a system where using slack variables into a system where I have a matrix which has full rank. m rows m rows 2 m rows so, rank is 2 m so, it is full row ranks so, a tilde has full row rank.

So, give me a system like this I do not bother about rank of A, I can convert it into a system where I have a matrix which equivalently expresses the constraint system, but has full rank. So, without loss of generality with that is usually written like this, without loss of generality we can assume that rank of A is m. Once I know this thing it will it is very important to know about certain aspects of convex sets called extreme points because these are the things that will come up very soon. Because the solutions of the linear programming problem would lie in some of this corner points or extreme points you see these points this point, this point, this point, this point, this point, these are corner points of this nice looking convex set or a polyhedral convex set. These actually are bounded polyhedral convex set which is called a polypore. So, I just remind you once again the by the name bounded polyhedral sets. (No audio from 28:19 to 28:26)

(Refer Slide Time: 28:40)



Now, once I know this, these points are called extreme points and why they are called extreme points will soon come up with the definition. (No audio from 28:36 to 28:47) It is something new and I understand that this are corner points, like you should have a convex set like this, even I understand this is. These are the extreme points, very bad corner raises this is my whole set, but how do I mathematically talk about this point, what is the feature of this point? The feature is that I cannot take two distinct points on this set and express this point as a linear combination or a convex combination of these two points.

Suppose I take this point and this point two distinct points on the given set and I join them. I will get some point here; it will never become a point on the corner. A point on the corner can never be expressed as a convex combination of two distinct points of the set, these are the characteristics of the extreme points. So, I can define an extreme point as a follow. Let s be a convex set, then a point x is called an extreme point that is you cannot express it as a convex combination of two points when lambda is strictly between

0 and 1. That is excluding these two points if you take all the points all such points. For example, on this line this is not an extreme point, this is not this point. So, you cannot find any two distinct points on this set so that if you join those two distinct points by a line segment. This extreme point would be one among those points, a point on the line segment other than those two points.

So, this is a point is point x; obviously in s if you are not happy then I can write like this that point x element of s is called an extreme point of s. If for x 1 and x 2 element of s and lambda element of 0 1, x equal to lambda x 1 plus 1 minus lambda x 2 implies that x 1 is equal to x 2 equal to x. The only possible convex combination is when x 1 x 2 is equal to x. It cannot be any either cannot be two distinct points whose lines segment other than them. The two points itself an extreme points lies no such extreme point can will have the property of been a proper convex combination not the extreme, the x those points themselves the proper convex combination of the two distinct points.

For example, here I have finite number of extreme points such sets are called polyhedral, polytopes, actually bounded polyhedral sets. But this some sets which have an infinite number of extreme points consider a circular sorry circular disk which is a convex set. So, you take a circle circular disk the boundary, this line and curve and then everything inside. Now, any point on this curve cannot be expressed as a linear combination proper linear combination of two distinct points x 1 and x 2 on this set with lambda b belonging to 0 and one that is lambda is neither 0 nor one. So, I am excluding two rows to extreme points it is in the interior of the line segment joining the two points, it is not possible.

So, this is an example of infinite extreme points. (No audio from 32:56 to 33:07) Of course, it goes without saying that the extreme points are actually in the boundary. If it is not in the boundary, then I can do this. If it is not in the boundary, then I can always get two points where who such that this expression is true so, it has to be in the boundary because this is not true. Now, I will try to give you a home work figure out how to show that extreme points are in the boundary. So, take a close convex set. So, if there is famous theorem called the Krein Mailman theorem which says every compact convex set is the convex hull of it is extreme point. So, if k is convex and compact and k is the convex hull of it is extreme points which I say mark as a extreme points of k. This is the (( )) famous Krein Milman theorem.

So, what we essentially want to show that if you give me a minimize, I will show our linear programming problem I show that minimizer corresponds to an extreme point on the feasible set. And this is the idea that we are going to explore in the next class and which will lead us to the simple method. So, we will not do all the proofs in detail, but some of the proofs would be done to show you some ideas, but not all the proof because we cannot spend the huge number of time on linear programming problem. But just to give you a basic idea of this very particular class of convex optimization problems. Thank you very much.