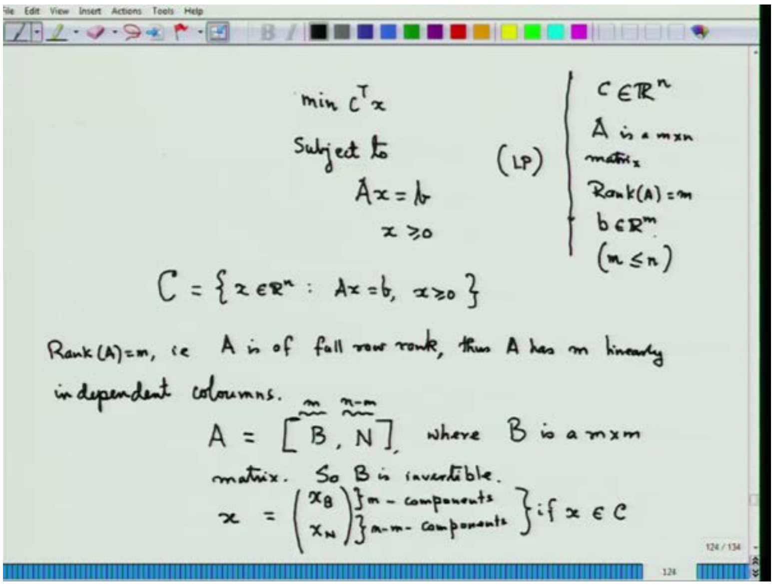
Convex Optimization

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

Module No. # 01
Lecture No. # 21

(Refer Slide Time: 00:24)



So, let us again indulge ourselves in the pleasures of linear programming. So, here is the problem that we are going to study, minimize C transpose x subject to A x equal to b greater than equal to 0. I will just list down everything, where C is in R n could be anything it could be 2,3,4,5 it does not matter. A is m cross n matrix and, the rank of A is m, b is of course element of R m. x the decision vector naturally is in R n which I do not have to specify, because you take a inner product.

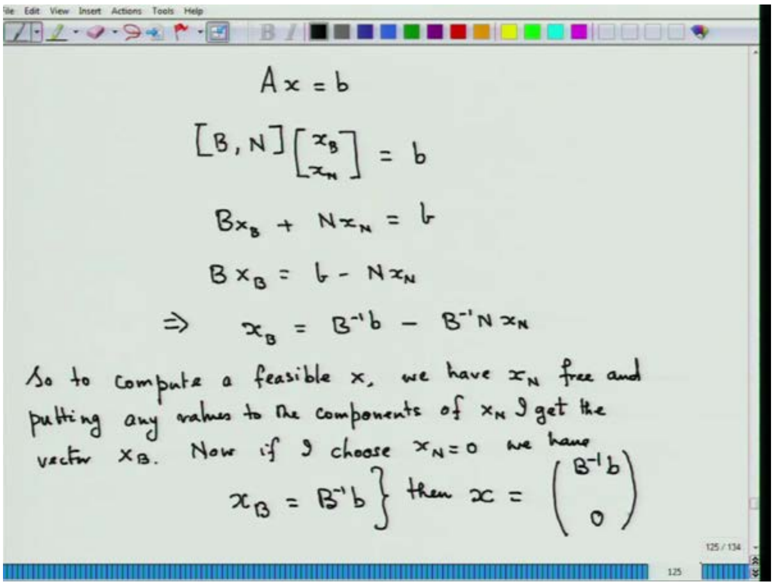
Now, if I want to solve this problem, it is very important to know what is the feasible set, there is do I or can I find points of the following set, now how do I go about doing this what is the easiest way of finding point? Now, this thing can be done through a little trick. The trick is as follows that, because you know that rank of A is m that is a is of full row rank, what we can do is the following, we know that thus A has m, linearly independent columns of course, m is less than equal to n. This is quite a standard thing, because the maximum number of linearly independent vectors that you can have in n space R n, is n. So, thus m has m linearly independent columns.

Now, suppose by a stroke of luck or by what we call multiplication with permutation matrices, I find that the first m columns that I have of the matrix A is actually linearly independent. So, it may not be so easy to determine which vectors are linearly independent if the matrix is very large then there is certain trick by which we do something we can do something, but let us just assume for the time being that I can partition the matrix into B and N, where B is m cross m matrix

So, B is a matrix whose rows as well as columns. So, this first m rows and these are N minus m rows m columns. So, the first m columns of away a linearly independent, and taking those columns and the rows themselves I form a matrix B. So, B is an m cross m matrix, and you know this I have partitioned the columns. So, this is m cross m matrix B is invertible, because it has m linearly independent rows, and m linearly independent columns. So, B is invertible. Now, because I portioned the matrix into B, and N any feasible vector can be portioned into two parts x B and x N.

So, take any vector x, let us write it down as x B and x N. So, x B corresponds to the indices the vector, corresponding to indices or the indices of the first m rows, the next one corresponding to the indices of the first, last n minus m rows. So, here I have m components, and here I have m minus m components, and let us assume that x is in C, if x is in C, then I must have A of x is equal to b.

(Refer Slide Time: 05:35)

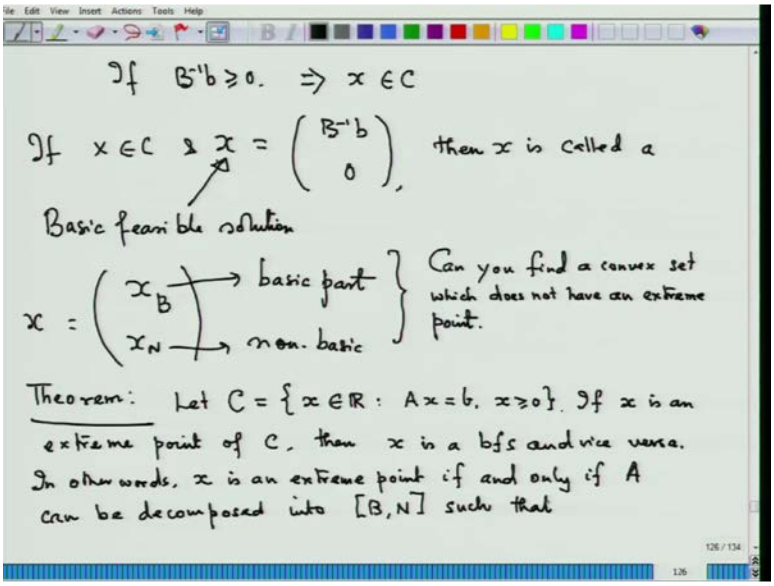


If I have A x of equal A of x equal to b, I can write down this as B, N. So, this is a partition matrix, but you know you can think of this itself as a matrix with two components, and then you can go on doing whatever you want in the same policy, in the same way you do matrix multiplication.

So, now what you have is B times x B, matrix B multiplied with the vector x B, and N times x N is B. So, x of B is B minus N x N. So, this would imply sorry B of x B is b minus N of x N. So, x of B using the invertibility of B is B inverse b, minus B inverse N x N. So, what do I get from here. So, x n is free, see if I can choose whatever x N, I want I can put any value to x N, then I can get x and hence I get the vector x.

So, to compute a feasible x, we have x N free and putting in any values to the components of x N, as I did desire whatever I want, the components of x N I get I get the vector x B, this is clear to everyone. Now, if I choose x N is equal to 0, Now you might ask me, why you have written as x B and x N we will very soon come to the point, just for the time being just take on this symbols, and I will come to the point very soon why x B and x N, if we choose x N equal to 0, we have x of B is equal to B inverse b and then x, this vector is called B is B inverse b, 0.

(Refer Slide Time: 08:50)



Now, if B inverse b is greater than equal to 0, it implies that x is a member of the feasible set C, because it satisfies A x equal to B, and x greater than equal to 0. So, any feasible solution so, if x is element of C, and x is written as B inverse b, 0, then x is called a basic feasible solution, and that is why this x B, x N equal to x, this is called the basic part and this is called the non basic part. So, a non basic components are all 0 feasible solution that we get is called the basic feasible solution. So, this is a basic feasible solution.

The most fundamental result in linear programming is that a basic feasible solution corresponds to an extreme points of the convex feasible polyhedral. So, if you take the polyhedral set which is the feasible the set C, then every extreme point of C is a basic feasible solution to the linear programming problem, and every basic feasible an every extreme point is the basic feasible solution, and every basic feasible solution is an extreme point, and in fact, it can be shown that any optimum solution is a basic feasible solution, and hence is attained at an extreme point is itself as an extreme point.

So, what we are doing in linear programming is computing the functional values over the extreme points and then trying to find which is minimum, but there could be huge number of extreme points if you have problem data is large that is there is lot of decision variables could be millions, and trillions of extreme points. I think which you cannot view nobody can view it so, difficult to think about it.

Now, if that is the scenario then you cannot geometrically view it, neither can you do enumeration of a huge number of points, the function value it will simply slow down the whole process. So, the whole question is that if I know that, I have a I am at an extreme point which is a bfs, but it is not a solution to the original problem, we will show under what conditions you can check that it is not a solution to the original problem.

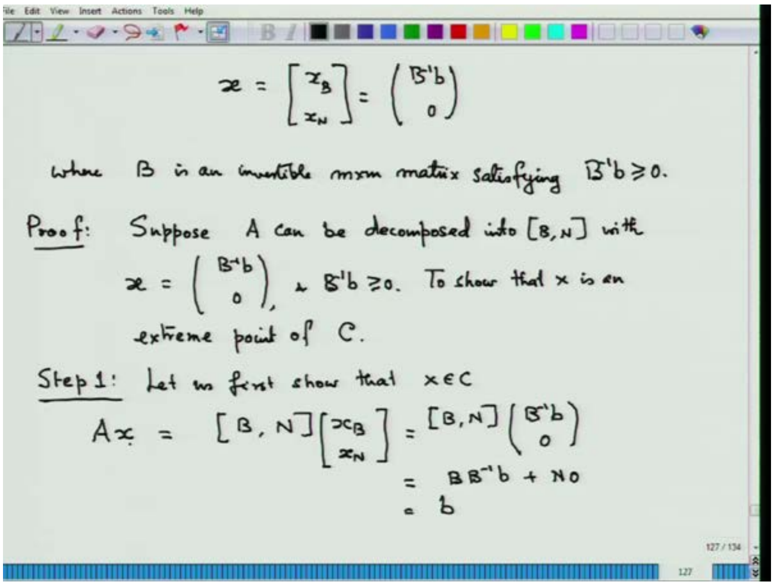
Then what we can do is by a clever way, which is called the simplest method is we can move from one vertex to another vertex. So, that the functional value the value of C transpose x actually goes down as I go to a new extreme point. So, I have to find clever way to keep on moving from one extreme point to the other, but at the same time keep on decreasing the function value as I keep on moving over the extreme points.

So, I do not cover all extreme points only cover some few of them, but I will I reach the solution. So, it saves an enormous amount of computing time, and enormous amount of effort enormous amount of mental stress and so thing, so much that. This process which is achieved by the simplest method in quite a simple way is as a result very very popular, and is one of the most elegant algorithms in optimization theory.

So, what is the idea? So, let us write down the theorem, we have not written a theorems for long time. So, you write down the theorem need not give it a number, but mathematicians likes to state very very state of that important result, I would say not set of the very important fundamental results are usually given as theorems. So, now, you might ask me, how do I know, that there will be an extreme point of this convex set, could there be a convex set, which does not have an extreme point, could there be a convex set which does not have an extreme point, I am asking with this question think of an example.

Can you find convex set which does not have an extreme point. Now consider this feasible set, if x is an extreme point of C, actually there is a major result in convex geometry we say that every convex polyhedral set has an extreme point. If x is an extreme point, then x is a basic feasible solution whose short hand throughout the world in any optimization book is always bfs, and bfs an x is a bfs and vice versa.

(Refer Slide Time: 16:01)



In other words x is an extreme point if and only if the matrix A can be decomposed into the partition B and N, such that x equal to x B, x N equal to B inverse b, 0. Where, B is an invertible matrix. So, if x is an extreme point it can be represented like this, and if x can be represented like this then x is an extreme point. Where B is an invertible m cross m matrix satisfying B inverse b greater than equal to 0. So, this is what? It is called the bfs.

So, anything or any made thing will give some slightly more general versions of this very different way of defining things. So, what we have is this following result, which is very very fundamental found in any linear programming book. Now given any point x in the set C, you know x can be represented as any vector which has some point 0 some point non 0, you can always represent them, may be the number of non 0 variables, non 0 components are not there all are positive it could be like that, because there could be something internal x must be strictly greater than equal to 0, but in general I can always write a point with some 0, some non 0.

So, let me just tell you, how to go about proving this very very important result. Now let me do the proof. Suppose A can be decomposed, I want to show that, if I have a bfs it is actually an extreme point. A can be decomposed into B, N with x equal to this whatever I have assumed in the result.

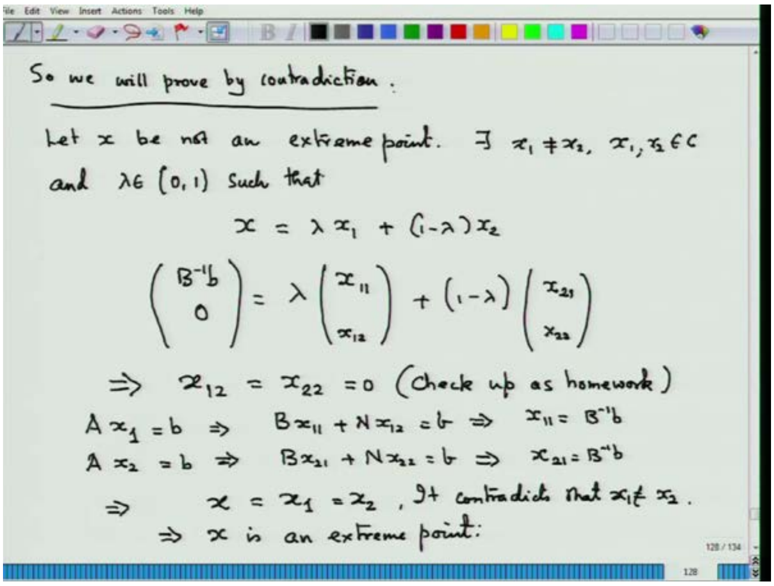
Now, to show that x is an extreme point of C. Now, x is of course if I have this x is of course, feasible. So, step one that show that x is feasible. So, let us show first where I know just this decomposition, I know that x is greater than equal to 0, all the components are greater than equal to 0, but let you not know whether it satisfies A x equal to b, which is the major thing that we have to check.

Let us first show that that x is in C, to show that let, me compute A of x which is A, because by the hypothesis a is decomposed into two parts B and N. where, B is an invertible matrix m cross m, and x B and x N, which is B inverse B and N, which is B, N and B inverse b, 0 which would give me B inverse b plus N into 0. So, that will be nothing, but this would be identity, because this is invertible matrix being to be inverse b is identity so, its b. So, what I get is A x equal to b, showing that x is belonging to C. So, to first show the extreme x is an extreme point my first step is to show that x is in C, and then show that x is an extreme point of C.

Now, the next step we will do or prove by the method of contradiction or reduction (( )) as known to mix, but many mathematicians would not really like proves by contradiction, they would rather like straight forward proofs straight forward or may be constructive proof, but in certain cases it is much more easier to prove by contradiction, that is whatever we want to prove we take an hypothesis completely opposite to that the negation of that, and then we reach a contradiction that is some actually given hypothesis is contradicted, because we have assumed that our original claim is wrong.

So, it means if our original claim is not correct then there is a contradiction in or actual hypothesis, which means if our actual hypothesis is true it implies that our original claim is also true. So, it is P implies Q negation of Q implies the negation of P. So, this logical you know structure is used in the proof by contradiction.

(Refer Slide Time: 21:51)



So, let us do the same thing. So, we will prove by contradiction. So, let x be not an extreme point. Not a good English, any does not matter, if you understand what I am trying to say it is fine. Now, what Now, which means that there must be two distinct points in a C. So, when I take a convex combination of them with lambda between 0 and one, then x is one of those points. So, there exist x 1 not equal to x 2, x 1, x 2 in C, and lambda in 0,1 such that so, x is represented as B inverse b, 0. So, B inverse b consists of few vectors, few rows and columns.

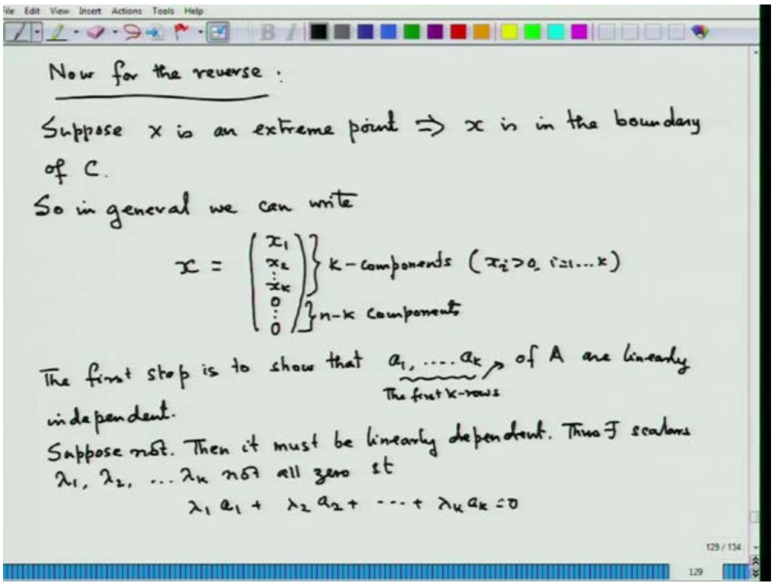
So, I am writing corresponding to the x B part, I am dividing x 1 into x 11, and x 12, and I am breaking up x 2 into x 21, x 22. You could as well as write x 1b, x 1m, x 2b, x 2m does not matter, what I would have is that this has the same number of components as x B, this has the same number of components as x N, this has the same number of component x B, and this has the same number of components as x N.

Now, what would immediately happen is that I equate with these vectors components, here I equate the components of these two vectors. Now, because this part in the non basic part everything is 0. So, here just by convex combination, because x 12 is greater than equal to 0, x 2 is also greater than equal to 0, it would imply which I will leave as the homework further the details.

Now, let us see what it means? It means the following so, some part is 0. Now, I have A of, because x 1 is feasible, I have A of x 1 is equal to b, which would imply again by the partition B x 11 plus N x 12 is equal to b, but x 12 is 0 from here, from the above line and this would imply x 11 is equal to B inverse b. The same goes for x 2, because that been an element of C, and because it is already greater than equal to 0, it also has to satisfy it is greater than equal to 0, and also has to satisfy x 2 is equal to b.

So, this would imply B x21plus N x 22 is equal to b, where again from the previous line we have x 2 equal to 0 implying that x 21 is equal to B inverse b. So, that would simply imply that x is equal to x 1 is equal to x 2. So, it contradicts that x 1, contradicts that and hence it is an extreme point. So, this implies that x is an extreme point.

(Refer Slide Time: 26:55)

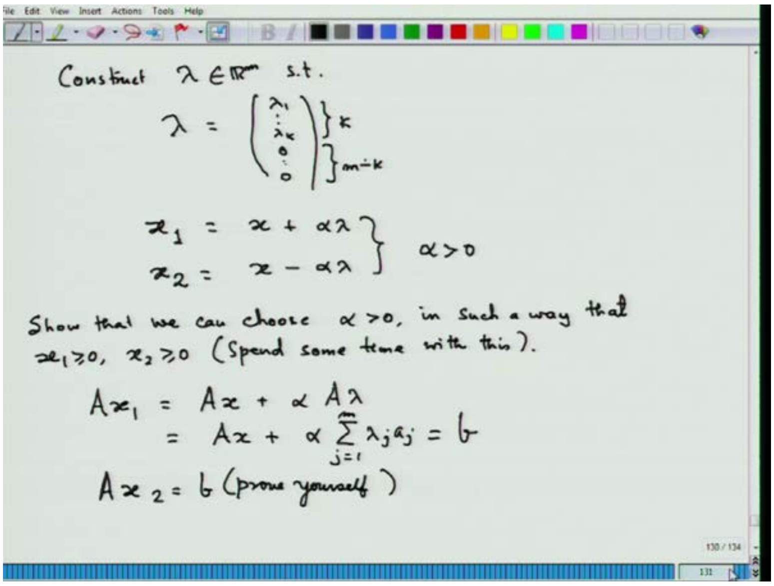


Other part the reverse is slightly tricky. Now, for the reverse: suppose x is an extreme point. So, x is not an interior point. So, this implies that x is in the boundary of C. So, in general so, without loss of generality basically in general we can write x is equal to. So, the first k-components are non 0. Where, x is an element of C, it is an extreme point, in minus k-components. So, it could be that in minus k is 0, that is n is equal to k, but in general you can always write extreme point in this way.

Now, corresponding to this k rows there are k columns in the matrix A. So, we are going to show that, the first k rows suppose these are 0, k-components x I strictly greater than 0 are equal to 1 to k. So, these are non 0 components, and these are 0 components. So, the first step is to show that, the first k rows a 1, a k. The first k rows of A are linearly independent. Again we will go by the method of contradiction suppose not.

Then it must be linearly dependent. So, this set of vectors must be linearly dependent. Thus there exist scalars as a real number, as we are in real field lambda 1, lambda 2, lambda k not all 0, such that lambda 1 a 1, lambda 2 this linear combination is actually 0.

(Refer Slide Time: 30:51)



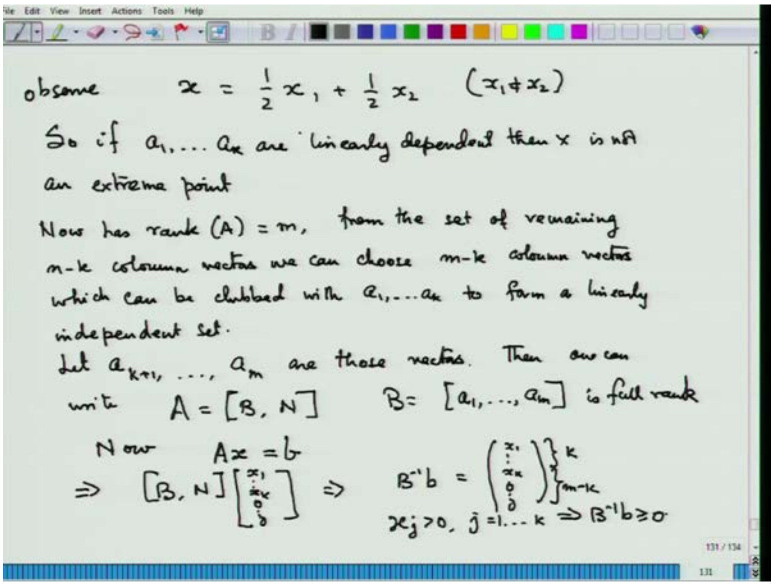
Now, construct the vector, construct a lambda in R n, such that lambda is lambda 1 lambda k 0 two dots and 0. k-components in minus k-components. Now I will construct from the given vector x, two more vector x 1, x 2. So, I will take an alpha, now the homework you prove show that, we can choose alpha greater than 0, in such a way that x 1 is greater than equal to 0, x 2 is greater than equal to 0, spend some time with this is a good exercise

Now, let us see what is A x1, x 1 I have got it to be greater than equal to 0, is A x plus alpha A lambda. Now this is giving me A x plus alpha times summation lambda j a j. Where, j is equal to 1 to m, because beyond m sorry this is not k m, this m in R n, because this corresponds to m components, m columns which are so maximum number of linearly independent columns is m.

So, if k is bigger than m we have to do something, if k is smaller than m then fine we already have it. So, if k is bigger than m, I can always reduce that number, I can always bring down and show that if once k is bigger than m you cannot have they cannot be linearly independent, and that is the whole idea of doing so.

Now, this is 0, because first k component is non 0, which gives you the linear independent thing the other part is 0. So, this is nothing, but other lambda is 0. So, this is b similarly A x 2 is b.

(Refer Slide Time: 34:16)



So, which you prove yourself. Now, once I know this fact observe that I can write x observe, write x as half x 1 plus half x 2, where x 1 is not equal to x 2.

So, if a 1, a 2, a k are so, if a one a two a k are linearly independent linearly dependent, that is what Then x is not an extreme point. So, what I want to show is that if my a 1, a 2 a k are now, linearly independent. I should know that, because I have full a has full rank m, I still have and if k is and of course, k has to be strictly less than m, k is not equal to m if k is strictly less than m. Then I can choose m minus k vectors from the remaining m minus k vectors, and just join them up with this a 1, a 2, a k and form a linearly independent set of columns.

Now, as rank of A from I can write rank as full or big letter or small letter is m from the set of remaining m minus k column vectors, we can choose m minus k column vectors which can be clubbed up clubbed with a 1, a 2, a k to form a linearly independent set. Because there are there are m in linearly independent columns.

Let a 1, a sorry a k plus 1 to an m are these vectors. For simplicity, they are or really just to make a multiplication with the permutation matrix. So, a k1 plus dot, a m are those vectors then one can write A is equal to B, N. With B formed of the first m column m columns is full rank, full row and column rank, full rank. And now, because x is in C we have A x equal to b implying B, N. So, our x is x 1, x 2, x k. So, B inverse b is x 1, x 2, x k assuming that k is strictly less than or equal to m 0. So, here I have k row, and here I have m minus k row.

So, since x j greater than equal to 0, and j is equal to 1 to k this implies that B inverse b is greater than equal to 0. And that is exactly what we wanted? If x is a extreme point then I can represent A in this form B, N. Where, B has all its rows and all its columns linearly independent, and x can be represented as B inverse b. So, anyway N the remaining part N into x N is anyway 0. So, and the B inverse b has to B greater than equal to 0 and that is our result.

Now, that a polyhedral set as an extreme point is something we are not going to prove, because that it will take us of what we want to do, but it is not a very difficult one, and we can possibly do that and the whole idea of bfs might come into play there again. But the whole idea of the this prove will come there again, but that is not a very very big issue, the proof idea of this extreme point business whether that the polyhedral set as an extreme point when comes from the idea proof or the theory or theorem. So, we will not get into this business, but I will just ask you to think that for a polyhedral set extreme points always finite just think about it why.

Today we end the or talk with this one, this proof which has taken quitter good amount of time, and we want to say that in the next class, tomorrow we are going to prove a very very fundamental result, this is something you have to remember, this is one of the most important results in optimization theory in convex optimization.

That if I know that a linear function has a lower bound, that which means that the dual problem is feasible. Once you know that the dual problem is feasible, you know that the linear program has a lower bound, and once it has a lower bound there exist a minimize for this problem. So, there is an extreme point where the solution will be attained where the minimam will be attained. So, this is exactly what are going to prove in the next class.

Thank you, very much, and I hope you have followed this proof, and you have enjoyed this fascinating fact that every extreme point corresponds to some special type of feasible point of C. And actually the optimal point, if the optima exist, it is in the in one of these basic feasible solution that is exactly what is we are going to show tomorrow.

And we will prove this very very important fact and since, you have already learnt about duality, and you know from weak duality that is very is if I just construct the dual problem and check it is feasibility, and we will tell you how to check feasibility of a system or linear equations very soon. And that can be done by one of the type of simplest methods, and we will tell you that once you know that there is a lower bound it is immediately cleared that there is a minimize, this is a very very important fact from the point of view of computation.

Because now a days there many standard algorithms which once you give in the input of the primal they know the inputs of the dual, because the dual inputs dual data is generated out of the primary data, there is no extra data anywhere. So, in optimization be where if you are studying duality, if you see any dual problem, which has some data which does not simply appear in the primal and suddenly come into, the dual then b where of such duals.

Any dual problem has to be created out of the data of the primal, because that is the only data you have, any other problem that you want to say is come running side by side with the same problem has to be generated with the data of the original problem, and that is something you have to keep in mind.

So, thank you, good night, and good bye.