Indian Institute of Science

Variational Methods in Mechanics and Design

Prof. G. K. Ananthasuresh
Department of Mechanical Engineering
Indian Institute of Science, Bangalore

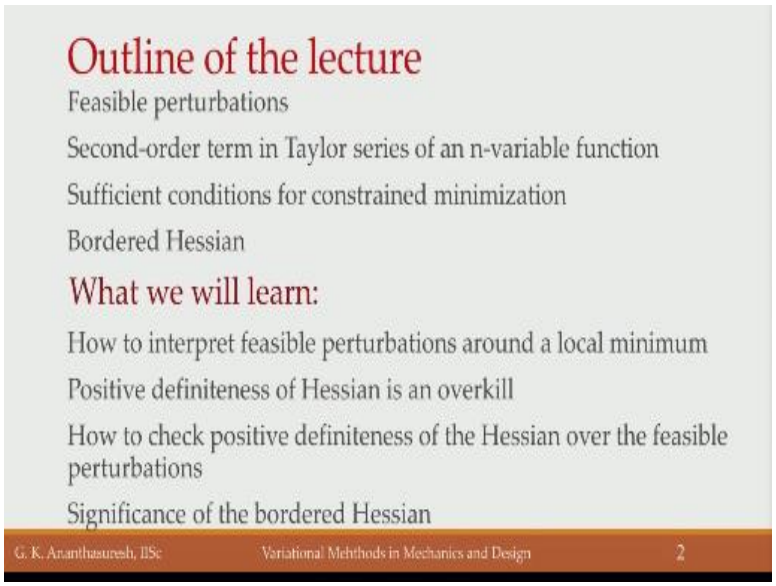
NPTEL Online Certification Course

(Refer Slide Time: 00:22)

Welcome to the sixth lecture in the course variational methods in mechanics and design in the last two lectures we said it was a detour and so is today's lecture it is also a small detour but it is important to appreciate calculus of variation. So we are talking about the finite variable optimization first we discussed unconstrained in lecture number four and then constrained where we discussed the KKT conditions constrained minimization lecture number five.

Today we are going to talk about the sufficient conditions for finite variable constrained minimization.

(Refer Slide Time: 00:59)

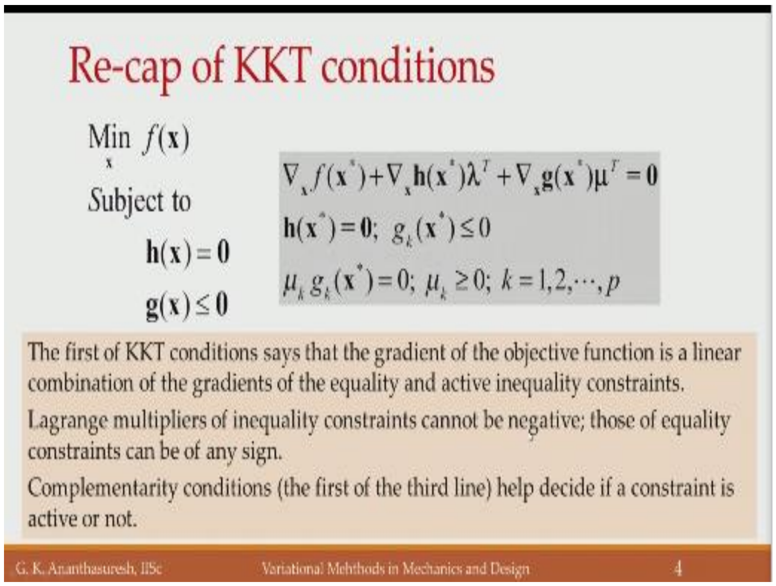


Here is the outline of the lecture we will first talk about what are called feasible perturbations when you have constraints you need to satisfy them when you put up to see if a point is a minimum or not for that for establish efficient conditions we have to take the second order term in the Taylor series approximation of the invariable function using that will establish sufficient conditions.

And then introduce a term that is not often found in optimization books called bordered Hessian some good books do talk about it many of them do not this concept of bordered Hessian which is very important to know if a given minimum is sufficient or not to do it easily okay. What we learnt in this lecture will be how to interpret the feasible perturbations around a local minimum okay.

Then we talked about this Hessians positive definiteness being more than what is required like over kill that is where this bordered Hessian concept comes and will understand its significance and we will also learn how to check if a given minimum for a constrained optimization satisfies sufficient conditions are not we will get a technique to verify that.

(Refer Slide Time: 02:24)



So first let us recapitulate this KKT conditions that is Karush-Kuhn-tucker conditions for constrained minimization so for a problem such as the one shown here minimize f of X subject to equality constraints there will be M of them this is a scalar f(x) is a scalar H(x) when we put in bold letters we mean that it is a vector an array so the h 1 h 2 hmr there that is equal to 0.

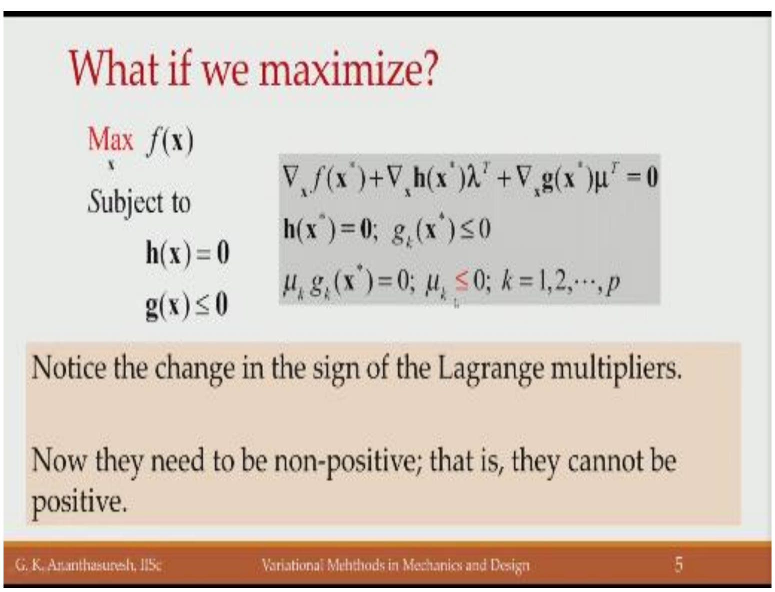
And then inequality constraint G(x) less than equal to 0 again we have g1 g2 up to P when you have that we wrote the KKT conditions here car wash Kuhn Tucker conditions the first one again to recap says that the gradient of the objective function is a linear combination of the gradients of the constraints both equality constraints and inequality constraints.

And then such a point X* which satisfies this first of KKT conditions should also be feasible meaning that it should satisfy the equality constraints and the inequality constraints here when I put k the k goes from 1 to P and then we have these mu KGK evaluated at that point X* which we believe is a minimum those are called complementarily conditions.

And then we also discussed that this μ corresponding multiplier corresponding to the inequality constraints should be non-negative when it is 0 corresponding constraint is inactive when it is not 0 the corresponding constraint is active meaning is strictly equal to 0 we noted that there could be a special condition where mu k + gk both can be 0 but what it says is that it is greater than or equal to 0 it cannot be negative okay. These are the things that one should note that KKT conditions say that the gradient of the objective function can be expressed a linear combination of the gradients of the inequality constraints okay.

And the Lagrange of the inequality constraints cannot be negative they have to be positive or zero whereas those correspond to the equality constraints can be of either sign they can be plus or minus okay that is why we use a different symbol lambda and μ μ has to be strictly nonnegative whereas λ can be negative or positive or zero complementarily conditions these are the ones which tell you how to decide whether a given constraint is active or inactive okay.

(Refer Slide Time: 05:15)



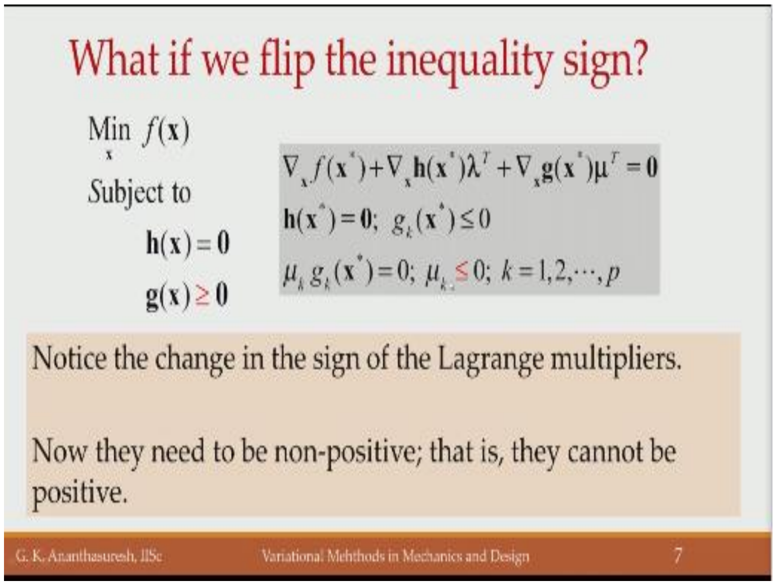
Now to understand KKT conditions let us flip it if you see in the previous slide we were saying minimize with respect to X this objective function now let us say we want to maximize it maxima is f of X with respect to X how do the KKT conditions change the change is shown here right. So here previously we said μ K should be non-negative. Now we are saying this should be non-positive because they can this shows that less than or equal to 0 it can be less than 0 which is negative or equal to 0 okay.

That is a change it comes about why does the change comes about because we are maximizing if you look back at the previous lecture where we argued why there is a restriction on the sign of Lagrange multiplier correspond to the inequality constraint you would recall that if you take the first order term that is gradient of F times ΔX that perturbation around the minimum X* that we wanted to be positive when f of X* to be a minimum whereas when you are maximizing we want this change that is brilliant of F times ΔX* that has to be less than that is negative.

In order to have f(x)* to be the local maximum so that change brings about this chain because again we had said inequality constraints when it part of it that additional chain should be negative this is negative that is negative when you look at this grant of objective function and this gradient the inequality constraint then both are positive then this if I multiply this by Δx* right without μ of course.

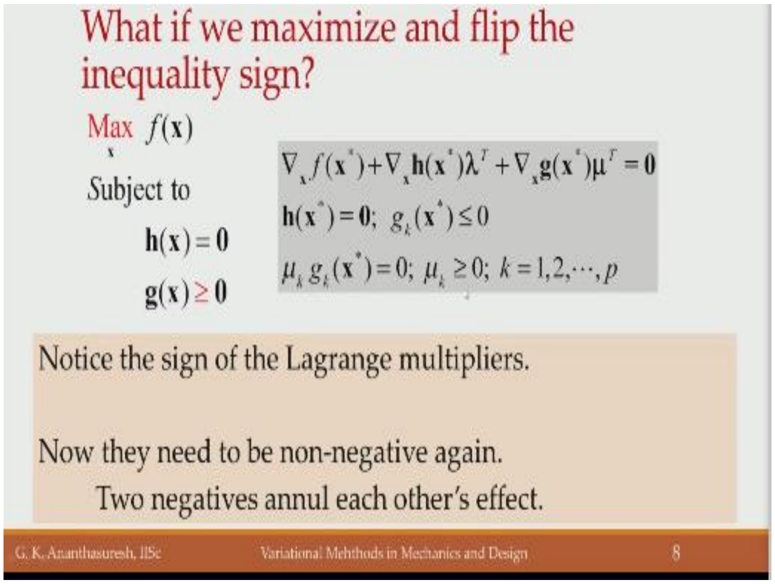
Then we see that these two should be equal to zero in that case this μ multiplies this one that has to be less than equal to 0 right just recall the same argument when I flip to maximum should become less than or equal to 0 because different books have different conventions some of them minimize some of them maximize and some of them may put this inequality constraints instead of less than or equal to the way put greater than or equal to if we do that what happens. We will see in the next slide so notice a sign change of the Lagrange multiplier correspond to the inequality constraints okay, now this should be non-positive.

(Refer Slide Time: 07:42)



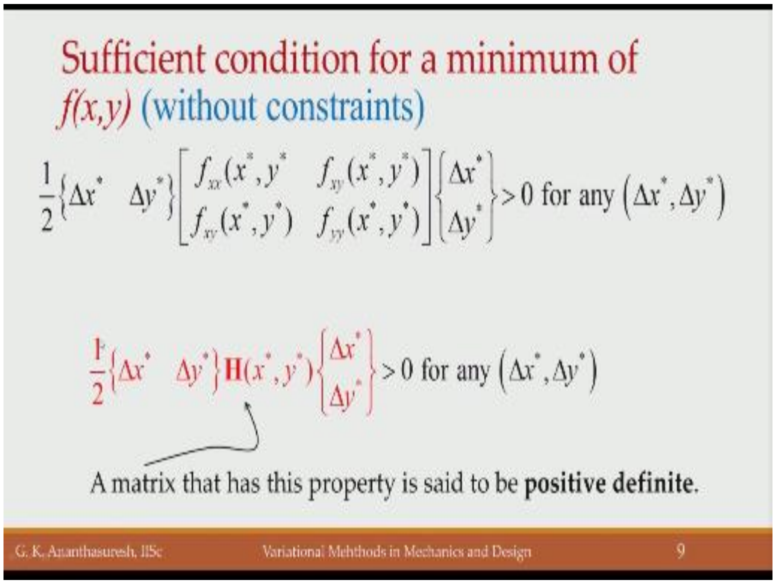
Now if I say max this we have okay now what if i change the sign of the inequality constraint here greater than or equal to zero instead of less than equal to 0 that also flips it to be less than equal to 0 because now after perturbation this should be positive with the perturbed value that is gradient of G times ΔX* should be positive the same thing with the corresponding first order term to f(x) then both of them should sum together 20 so the negativity of the Lagrange multiplier will take care of it okay.

(Refer Slide Time: 08:19)



Now what if you make both changes that is we are maximizing and inequality sign is greater than equal to then it will remain to be again non-negative. So once you understand these four variants original form that we wrote and then max the third one greater than equal to 1equality constraint and the fourth one which is on this slide where you are maximizing and inequality constraint is posed a greater than or equal to okay we want to be careful about how the problem is written accordingly you write the restriction on the Lagrange multiplier correspond to the inequality constraints.

(Refer Slide Time: 08:55)

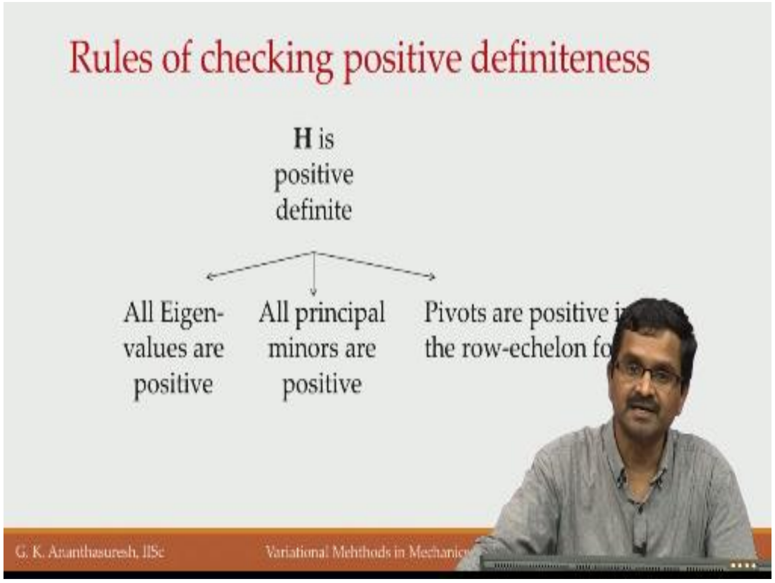


Now let us move to sufficient conditions which is the main point of today's lecture and for train problem there are no constraints these what we said should be true this is a second order term in the Taylor series approximation because for local minimum first order term is 0 right and now second order term has to be positive in order for f(x)* to be a local minimum these are second order term written in the matrix form ΔX* ΔY*.

If I take a two variable problem these are the perturbations from X* and Y* and the row vector form of that column vector here in the two-by-two matrix which we call the hessian denoted by this HX*Y* that is evaluated at the active minimum X* Y* this should be greater than zero for any perturbation if that is what this H has as its property then we say that particular point that is X* Y* is a local minimum.

And that is a sufficient condition okay, so this H is the Hessian to be positive definite that is what we want that is the nature of postural matrices whatever if you take any vector you post multiplied by the column and then pre multiplied by the row whatever you get should be greater than equal to 0 this half is just there because quadratic term have has half otherwise half has no significance okay.

(Refer Slide Time: 10:26)

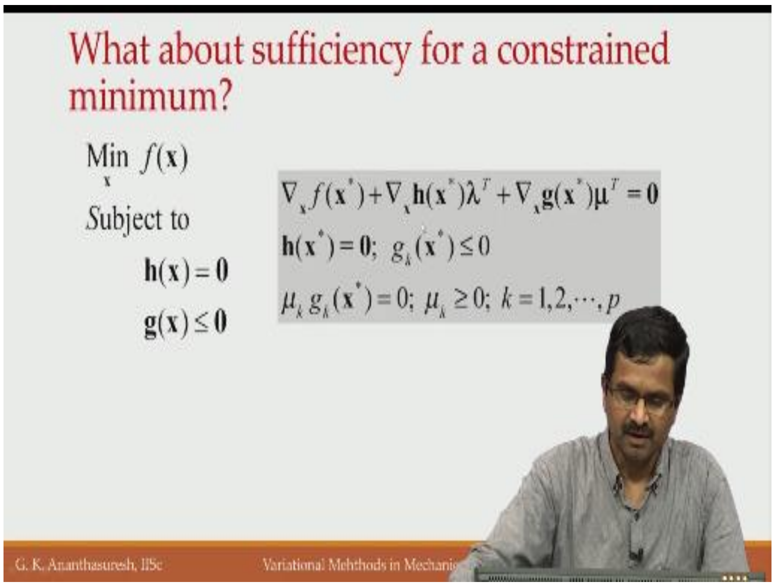


Now how do you check this for the unconstrained problems that H is positive definite there are many ways there are there are three ways these is these eigenvalues if all eigenvalues are positive such a matrix is said to be positive definite and that is that will satisfy this property other way of doing this is to check the principal minus of the matrix and ensure that they are all positive principal minus are the largest principle minor is the determinant itself.

That is you take the entire matrix take determinate the one before that will have the first row and first column removed and it goes like that okay. So first row second row first row first column second column if you keep on removing it we get principal minus all of them have to be positive or the pivots so called pivot that come about when you take the matrix and reduce to this row reduced echelon form if all papers are positive that also is a test for positive definiteness.

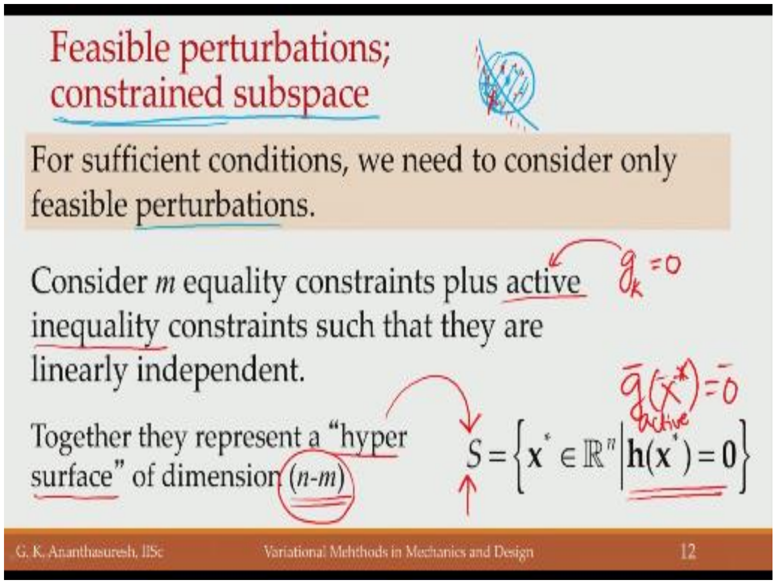
The letter to the principal minors and papers are not used that often, because now we have very powerful numeric analysis software which we are the eigenvalue. So that is the best one if all eigenvalues are positive then you have positive definite matrix which is sufficient condition for a problem without constraint that is unconstrained minimization.

(Refer Slide Time: 11:55)



Now if you consider a problem which is constrained whose necessary condition the so called Karush-Kuhn-tucker conditions that we discussed in the last lecture a little bit today what about the sufficiency of such a problem okay.

(Refer Slide Time: 12:13)



First of all if you talk about sufficiency we have a point which is told to us to be a minimum now we want to check so we want to move round perturb in the vicinity of that point and then show that any other point in the vicinity has a higher value of objective function compared to value at that point but then when you have constraints which is the case with constrained optimization the search the vicinity that you look for if there is any point that has higher value than lower value than the function that the point that where you are which is supposed to a minimum.

Then you cannot search everywhere in the vicinity but you have to search in the constrained subspace so let us let me take it a pen okay. So the constraint subspace right, so that is the focus here we have to talk about constrained subspace that is if I have a point let us say I take a two variable all given problem if I have a point somewhere I look around that point little space around it when you do not have constraints you can search everywhere.

But now when there are constraints a part of it may not be allowed right that may not be alone then you are to search only in the here that is what we constrain subspace okay. So for sufficiency conditions we need to consider only what are called feasible perturbations feasible point is the point that is wise the constraints both equal to any qualities. Now when you say feasible perturbations those perturbation this is the point I can check here if those are allowed right.

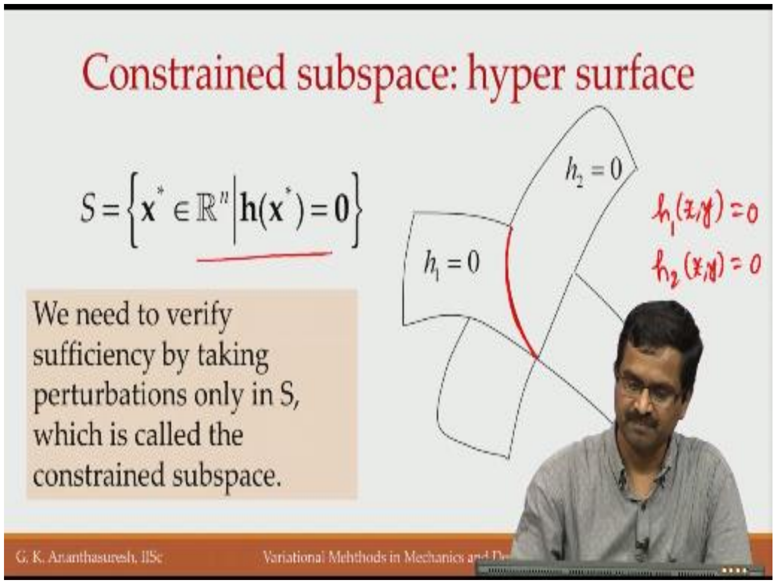
So let me change the color yeah if this is the point I can take here but not here because that is not feasible because that does not satisfy the constraints we say that this constraint says that he cannot be below that right. So that is what we feasible perturbations so if you have inequality constraints and some active inequalities inactive inequalities are inactive so you can throw them away.

But there could be some active in equals meaning those where GK if I call k is an index for this should be strictly equal to 0 rather than being less than or equal to so there are active inequality constraints so once you have an active inequality constraint it is same as inequality constraints and if these constraints are linearly independent which is a requirement that we call constraint qualification ok if you have such a thing then the perturbations should be chosen such a way that they lie on a hyper surface okay, of dimension n –m.

Because you haven variables and there are M equality constraints are active inequality constraints so the portable that we make will satisfy these little perturbations of the active inequalities and the equality constraints and hence will have freedom to choose only n -m perturbations okay. So that is mathematically shown like this where the constraints of space which is this S here is such that all X* R ΔX* whichever way you take they are to satisfy the constraints HX*.

And I must also add G active X* equal to 0 and putting our bar because there could be more than one inequality constraints a here denotes active okay active inequalities are equal to equality constraints such a hyper surface hyperspace or constraint subspace is where we need to check in order to see if the point has the lowest value in the vicinity because we only talk about local constraint minimum okay how do we do that.

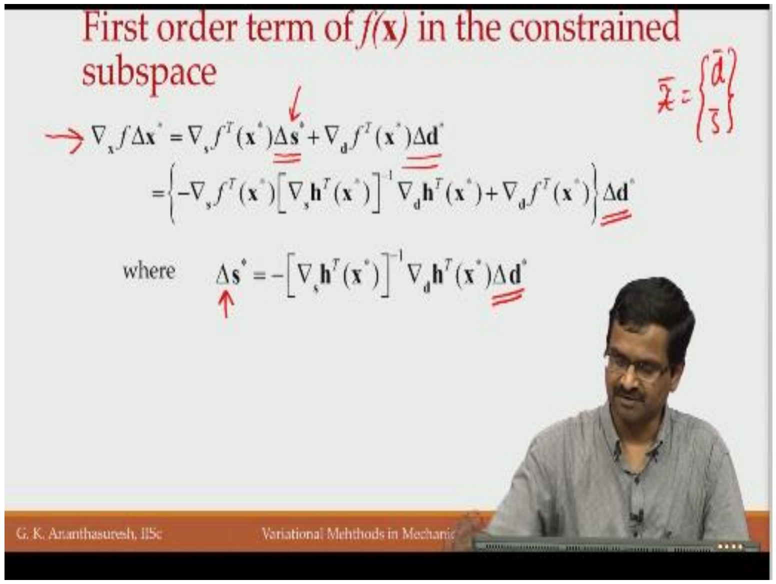
(Refer Slide Time: 16:41)



So if you look at it graphically consider that a two variable problem where you have h 1 is a function of X Y and then aged 2 is a function of XY=0. So we have to equality constraints that means that there is a surface h1 if I put xy as a plane and z I take this h1 then I get a surface correspond to h1, I get a surface as to h 2 I get a surface right so these things will be there at some value.

So you have h 1 = h 2 equal to 0, then since both constraints have to be satisfied then we get this thing which is intersection of those two that will be the hyper surface we talk we call it a surface here I am showing it like a curve. So if you have in let us say four dimensions h 1 = 0 will be like a surface okay and HT will also be a surface their intersection will be a curve only there I can search right that is what we are showing that is the hyper surface which satisfies this again when I say H it also includes the active inequalities.

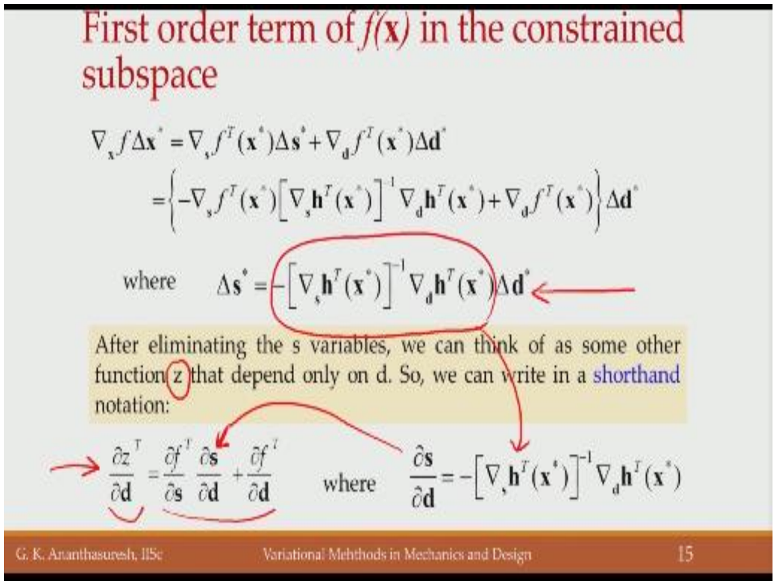
(Refer Slide Time: 17:59)



Which means that when we perturb and look at this term which is what we are seeing here on the slide for the objective function okay. So we have to consider the perturbation first order perturbation which is this first order term in the expansion of the function Delta X Delta F there is gradient of f x perturbation ΔX* which includes the independent perturbation sand dependent perturbations.

In the last lecture we were looking at this X being subdivided into D and S decision variables you are free to choose and s are the solution variables which you cannot choose because the constraints determine that. So we have the first term where we are taking with respect to the solution variables and then second term with respect to the decision variables since equality constraints have to be satisfied after perturbation we can express this ΔS* that is solution variable perturbations in terms of decision variable perturbations okay, when you substitute this ΔS* for this Δ* for this over there we get something like this where we have ΔD* or the decision variables.

(Refer Slide Time: 19:30)

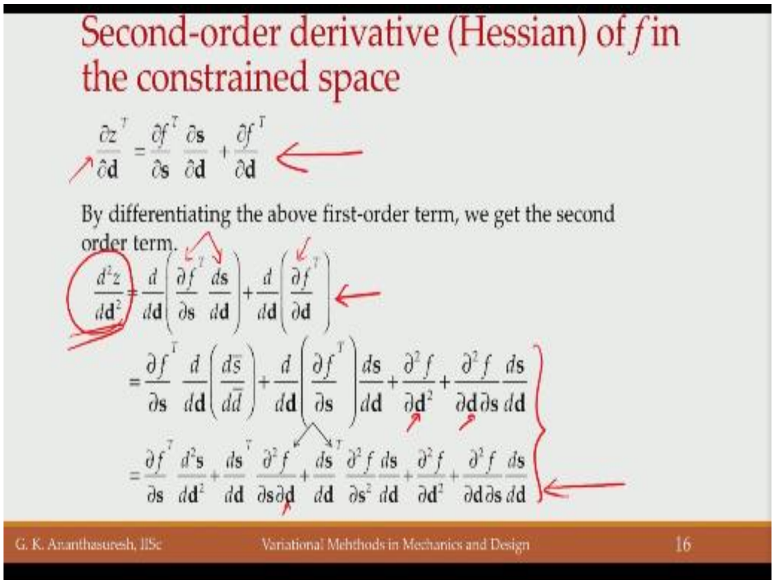


That is all we can change all right so now we want to see how to establish sufficiency for that we had thought of this function Z think of that as a new function that depends on n –m variables because original function was dependent on n variables. But then we have M equalities are active inequalities. Now those M variables are taken out we are left with n –m variables those n-m variables we can talk about this as a gradient of the new function Z okay.

That ZIdou Z/dou if I take that you have basically Z, but Fdou f/dou Sdou/ dou + dou f/dou D which is what we have is two terms because the Z essentially depends on D and S, but S is dependent on D. And that is where this comes about okay what is dou S/dou T it comes from here right. So whatever is seen here is exactly what we have there okay, I put a ΔS* equal to something time ΔT* that essentially means that douS/douD that the partial derivative of these solution variables where speculation variables is given by this quantity.

And that is what we need to put in here okay, then we get this gradient and we can take another derivative of this quantity to get the SC an equivalent of the constrained problem see if you remember for the unconstrained problem we are taking the Hessian or the F directly. But now we are not supposed to do that we need to look at the constraint one after taking out the dependent variables based on the equality constraints and active inequalities okay. So we start with this douZ/douD the new function which is n -m variables okay, we substitute this over here and compute this dou/dou and then take another derivative of that okay.

(Refer Slide Time: 21:48)



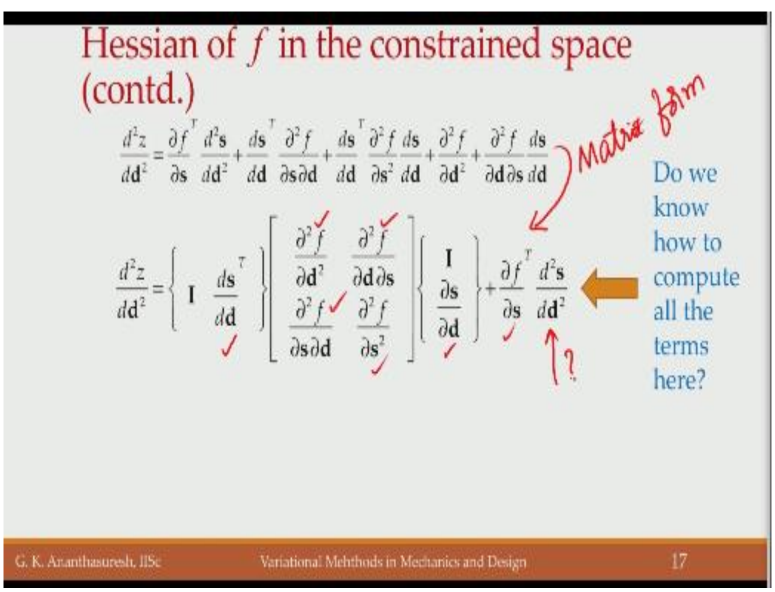
So we are taking this which comes from the previous slide now we need to differentiate this first order term of this new function Z one more time to get the second one okay that is the form we are looking at that should be positive right this should be positive definiteness positive definite you when you say positive for matrices it is not like a matrix is positive when all its entries are positive that is not at all correct what we say is a positive matrix ax plus definite matrix is if you get a quadratic form with any vector you should be greater than 0.

That is what we had in the previous lights that is what we need to compute for this quantity so now we are taking derivative of this dou Z.dou and do one more time so again we have the chain rule here because some variables depend on the others d/d of the quantity here plus d/d of that quantity over there okay well take derivative of this thing with respect to D that is what we have done the first line ok that is here.

Now we need to expand so over here dou f/dou S^T is there d/d of this and then D/D, D of this basically product rule we have a product that is two terms here one and two and we are taking derivative one at a time to get this and then same thing here. So we have only one term over here but then this douf /douT with respect to D first. So it becomes second derivative with respect to X the same thing we have to do here when it with respect to D we take with respect to S.

And then we have also with respect to D no this is first with respect to D here and then second time with respect to S and that part process we also need to put D s by D Das well because the solution dependent on the decision variables okay. This is an expansion even if you do not see it right away if you sit down and think about it you will know how these steps come about it is just a plain old differentiation no tricks involved here okay. Now we have basically d square Z/DD^2 from the previous slide this lengthy one that is this lengthy one.

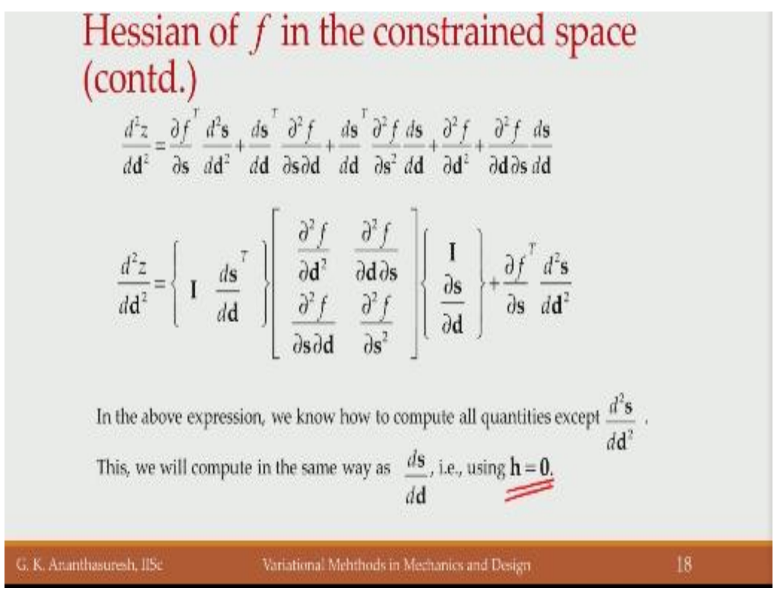
(Refer Slide Time: 24:20)



Is produced here and we are writing in the matrix form it is only rearranging these things there is no other trick here it is just a matrix form of the scalar expression to this form right if you look at this now what are all the things that we can compute we already know this because we had used that earlier and we know how to get this that is just differentiating objective function toys r expectation variables we can do this we can do this we can do this.

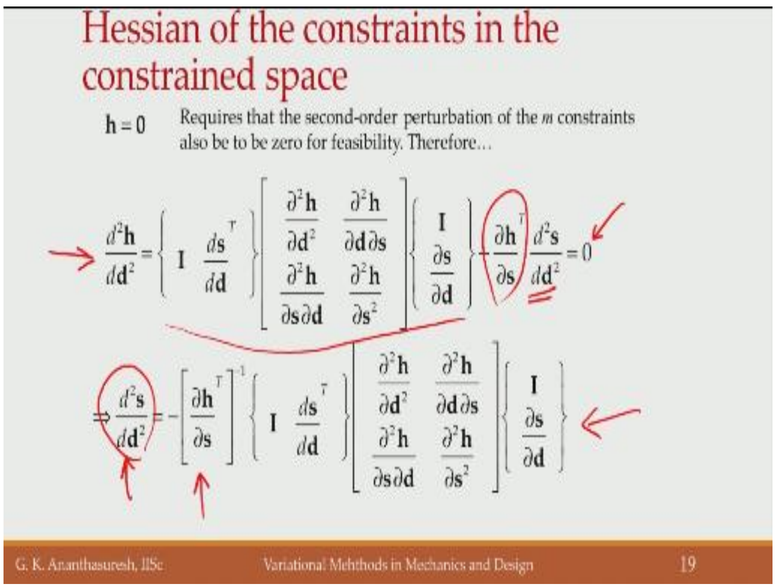
And as I said we already have this and we have this what we do not have is this we do not know dou square D^2/DD^2 we do not know how the solution variable second derivative with respect to decision variables is in order to compute this we do exactly the same thing for the constraints right. So we take second order term of the constraint we say that should be equal to 0 then we get this quantity that we do not know right that is what we will do next okay.

(Refer Slide Time: 25:34)



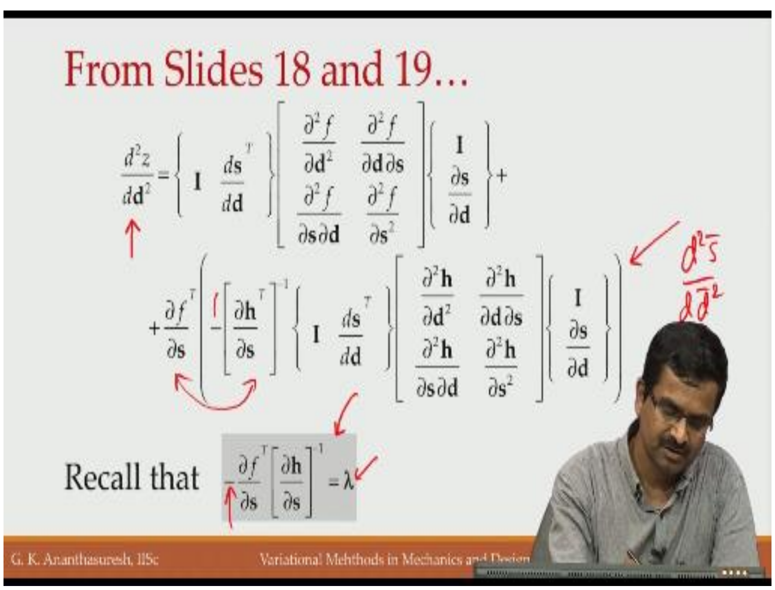
Using the fact that H = 0 to second order can you talk about sufficiency they are to consider the term in taylor series up to second order. So you have constraints you want to be zero after perturbation constrain should be zero up to second order anyway first 0^{th} order term is satisfied there is a feasible point and the first order term it will be part of KKT condition that will be there now when you perturb you have to make sure that the edge remains 0 after the perturbation.

(Refer Slide Time: 26:06)



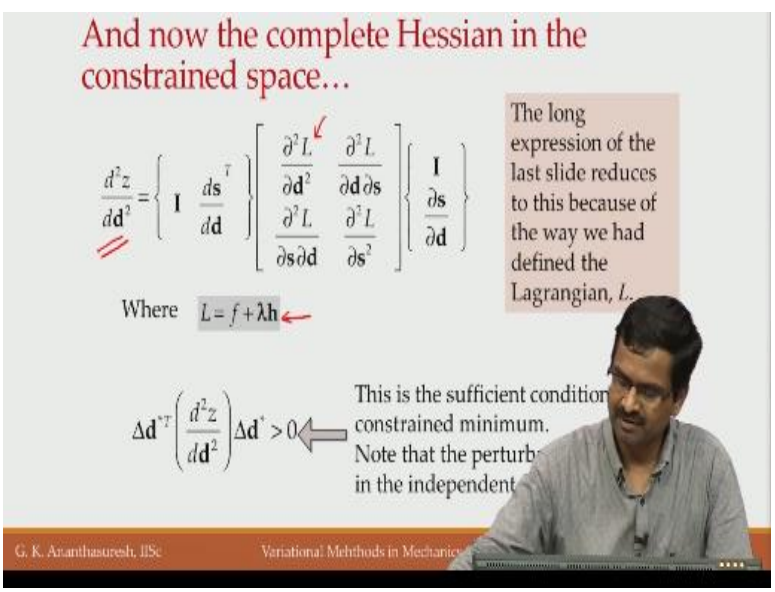
So whatever we did for the objective function now we are showing it for the equality constraints as well okay. Now this should be equal to 0 okay that gives us a way to compute our unknown thing D^2 S/DD^2 will become inverse of this quantity which is over here and then the rest of it that is all that is here is seen over there okay. Now we got this now we go back and substitute this thing that is substitute for this in the second order expansion of our ZZ is basically the new function that we imagine which takes care of the active inequalities and equality constraints okay.

(Refer Slide Time: 27:00)



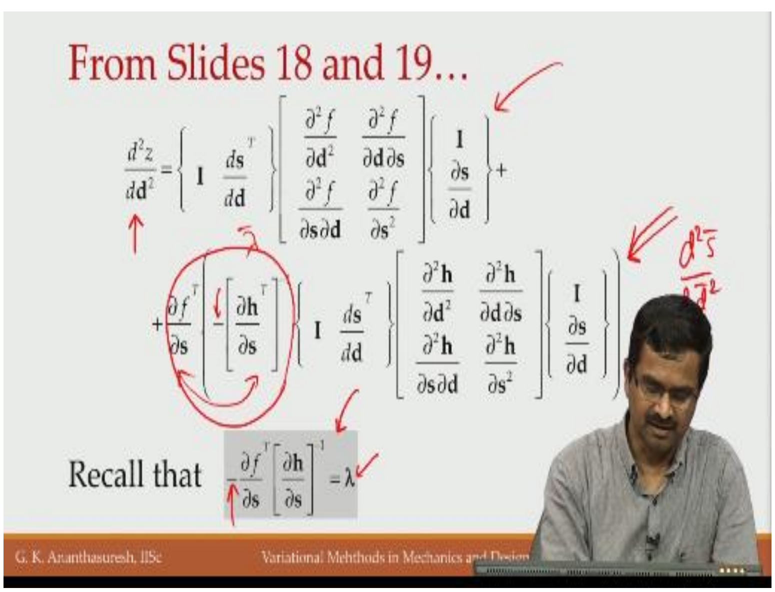
Now we substitute and go back look at this d square z by DD^2 then we get something like this right, so we have plus this whole thing that we have as a second term is that D^2 S/dt^2 okay. That we have this long one now we also recall this fact how we define the Lagrange multiplier it is basically the sensitive negative of the sensitivity of the objective function to this entity in the constraint okay that is what is λ when recognize this and this together for these two terms you become λ that include in the negative sign. So negative sign is also here okay.

(Refer Slide Time: 27:53)



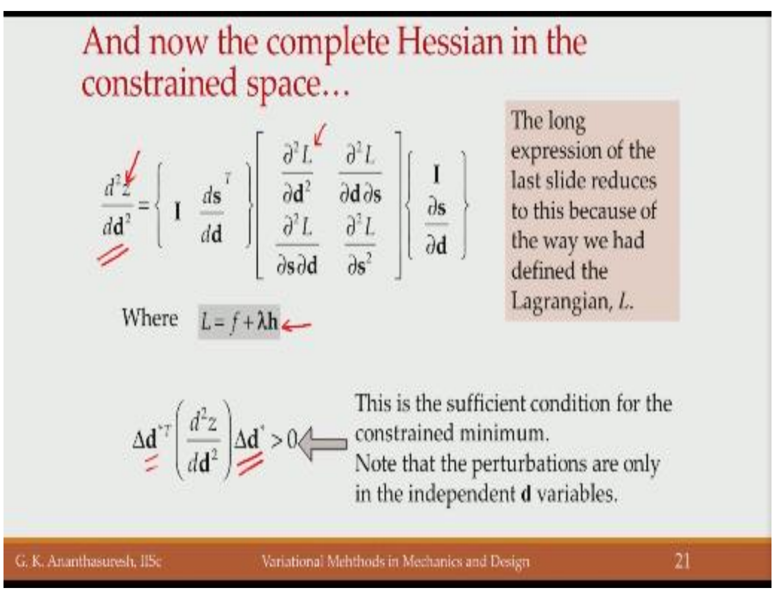
So we get that D^2 is that by DD^2 in this form where we note that this Lagrange is nothing but f+λ.

(Refer Slide Time: 28:10)



So whatever if you go back to the previous slide whatever was happening to F in the first term is also happening to H in the second term because this whole thing is lambda okay so we can say rest of it is similar only f here becomes λ times H okay.

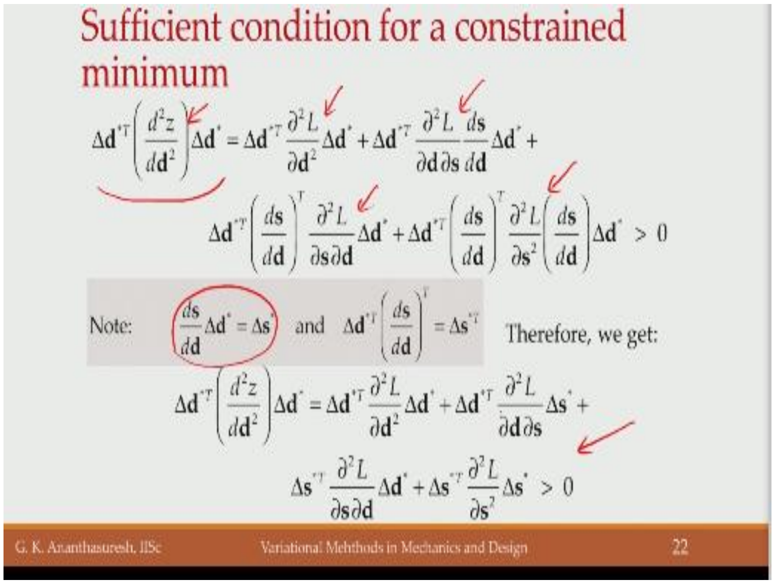
(Refer Slide Time: 28:30)



So that is how we get f+λ edge in this in this manner right now again notice that this Z which is depend on n -m variables is like unconstrained function so for this efficiency for that such a thing is that should be positive definite in other words whatever ΔD* that I take that is decision variable perturbations.

And a transpose of it get a scalar that has to be greater than 0 that is positive definite greater than a greater than 0 stick to the positive definite it is rather than equal to 0 positive semi-definite and all that we had discussed earlier in lecture number 5 a lecture number for all that applies to this Z as well so this is the sufficiency condition but then which ΔD* do you take the question will arise right.

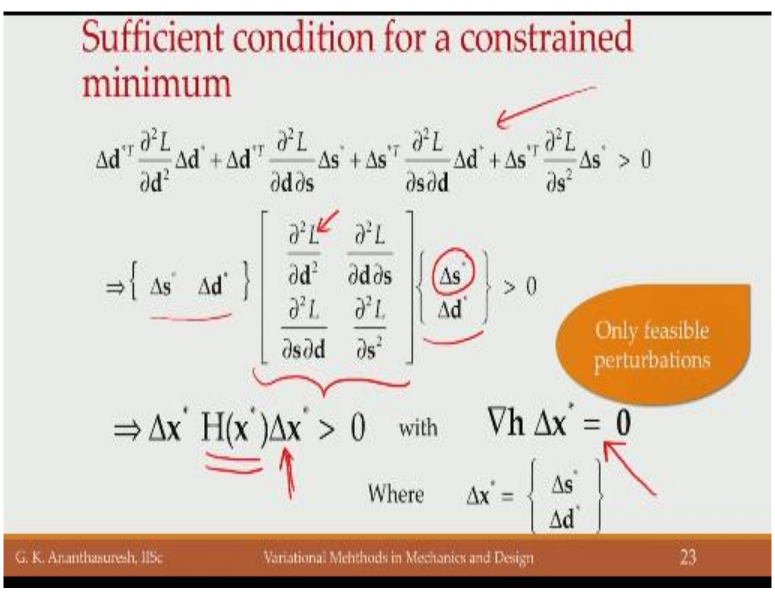
(Refer Slide Time: 29:19)



So for that we have to go back and look at the fact that whatever we have taken ΔS* ΔD* partitioning that does not really matter if you now look at this Δ if you take this quantity which we say should be greater than zero we expand that now that the Z part of it what we have this L that we derived earlier okay in terms of DS.

As well as this SS because that is expression we had for this thing d square Z/DT^2 in terms of L we substituting everywhere long term right first second and third term and fourth term and then note a few things that ΔS* is nothing but DS/DD times ΔD* and you know transpose we put all of that and do a little bit more of the substitutions into this okay again you have to pass and look at every term to understand if you do not see it right away.

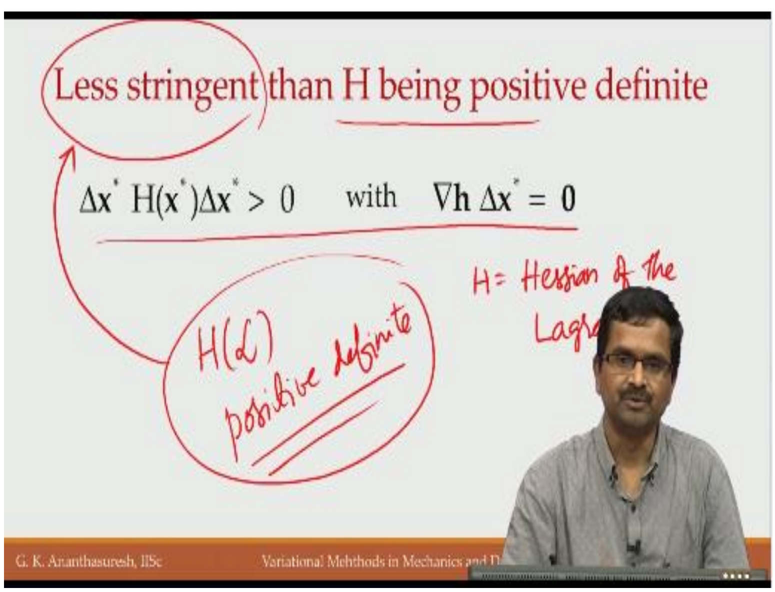
(Refer Slide Time: 30:23)



So if you arrange this long expression where we substitute in terms of ΔD* ΔS* you get something like this okay so we have the perturbations ΔS* ΔD* which is nothing but ΔX*. Now which way you partition does not matter no because they came together and what we should notice is this matrix that is dou square l by d^2 l/dou and all that okay the nothing but basically Hessian with all of them because D and s which every part does not matter it basically becomes the Hessian of the Lagrange.

So notice that it is l not object function but Lagrange that should be greater than 0because that should be satisfied but again remember that these cannot be arbitrary only ΔD* that is only n - m of those can be arbitrary other ones get depend on those. So we have this constraint that equality constraint gradient and active inequality constraint gradient should be equal to 0 we do not want this ΔX towards to be arbitral which was the case in the with the minimization un cashed minimization but now minimization these cannot be arbitrarily chosen there to satisfy this okay how do we do that is a question.

(Refer Slide Time: 31:51)



So when we look at H what H this is the Hessian of the Lagrange when you say H here this H is sign of the Lagrange okay that being positive definite is whatever we are saying along with this constraint okay. That is those perturbation that satisfies the Equality constraints first order term being equal to 0 and active inequalities this is less stringent than asking generally this Hessian of the Lagrange to be positive definite.

So if you say this should be positive definite you are asking for more that is an overkill this is why okay but we can go or something that is less string than that which we discussed in the continuing lecture thank you.