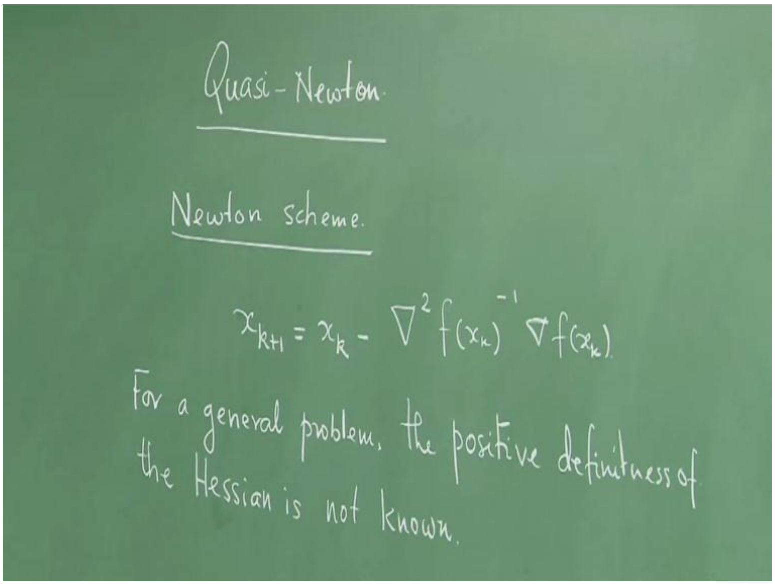
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 14

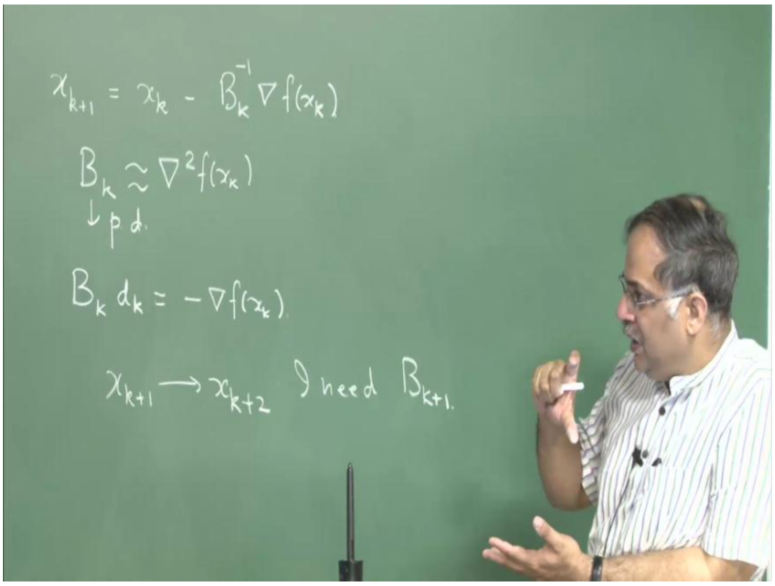
Let us start today on constant optimization. You know, in the last lecture, we stopped while making a progress on Newton to Quasi-Newton methods. We stopped telling that this issue of Quasi-Newton method cannot be understood better, cannot be understood nicely, if you do not have a good knowledge about constant optimizations, specially the Karush Kuhn Tucker conditions. It is very important to know the analysis of Karush Kuhn Tucker conditions or analysis of constant non-linear optimization had actually began much before these methods have like this Quasi-Newton method had been developed. So in 1951, people knew about what are the necessary and sufficient optimality conditions for a non-linear programming problem, while 1970s and late 60s these developments have taken place a Quasi-Newton methods.

(Refer Slide Time: 01:14)



So I would just give a brief idea of why we want to do the Quasi-Newton method and why we need and when where we need to use a constant optimization. So a Newton iteration, the Newton scheme, the Newton scheme is to write the k plus one eth iterate in terms of the k th iterate as. Now the important part is I either see I do not really know whether my hessian matrix at every point is iterates that I will get is positive definite. So for a general problem, the positive definiteness of the hessian is not known. So, if it is a strongly convex problem, there is no problem at all. So for a strongly convex problem, there is no problem often using Newton's method, you should use Newton method. If the hessian can be simply computed, otherwise a problem is that you now if you have a general problem, you may end up with an x k where the hessian matrix is not positive definite and then you lose that descent direction. So, this may no lager be this whole thing may no longer be the descent direction. Now how do I get about go about solving this problem.

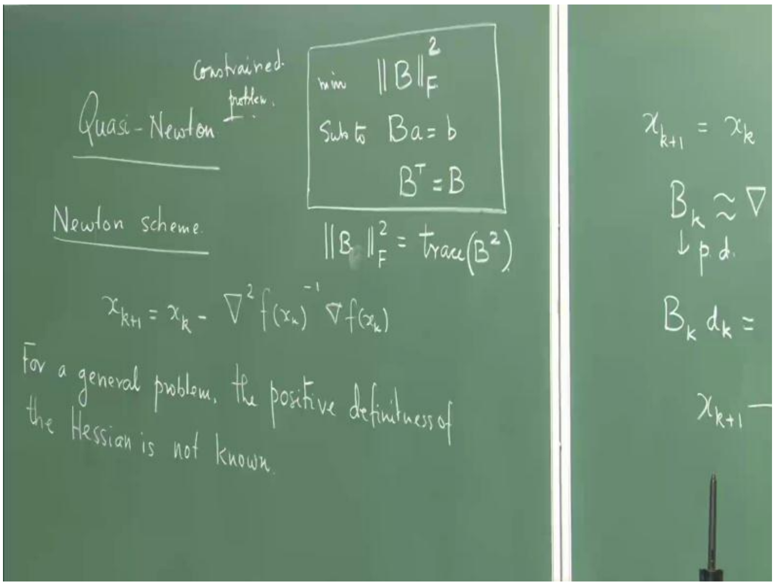
(Refer Slide Time: 04:01)



So let me think for a while that I want to whatever be by hessian matrix, I want to replace the hessian matrix by some other positive definite matrix B k. So, what I am making at this movement is that I am trying to has B k as some sort of approximation to where B k itself is a p d matrix or positive definite matrix. So I am trying to make this sort of approach. Now if this is the case then what I have to do is the following that you want to do this you want a p d matrix which is an approximate of this. So, I can now write this whole thing the equation as B k d k is equal to minus grade f x k that is what it turns around turns out to be. Now when I have B k and I have d k everything, I have got B k, now when I have to go from x k plus 1 to x k plus 2, so I need the next approximation which is B k plus 1.

So the question is, how will I get b k plus one that is a major question. Here in order to do so, people have said ok, I will not make a major change in the matrix B k. So the difference between B k and B k plus 1 should be same means, the distance between when they the the B and the B k plus 1 and B k they might should not have a huge difference between them. And to do it in such a way so that it maintains the positive definite. So I want that new B k plus 1 would also satisfy a similar type of condition. So if I fix some, so if these are fixed.

(Refer Slide Time: 06:58)

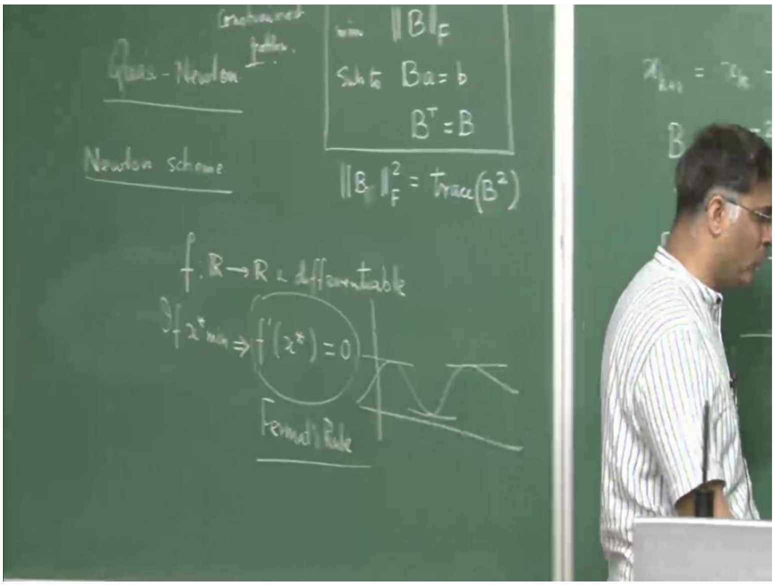


So basically what I want is to find B, where I want to minimize it is frobenius norm to be square, subject to frobenius norm is same as the I will tell you what the norm is is an norm of matrices, because this is space of matrices. And subject to say B of a is equal to small b; a and b are fixed, basically d k and minus grade of f x k. And also you have to have this additional constraint B transpose equal to B that is it is symmetric. So, you want to have a matrix which satisfies this, and so that is the new matrix that you will get, and you will be able to represent B in terms of a and b. So, once you know your grade f x k and d k you can find then new matrix B k plus 1 which will be given in terms of this., because, we want a matrix which is not very different from B k, the B k plus 1. So we want basically to minimize the difference between B k and B k plus 1. I am just I am not taking that I am taking that matrix out for the movement, and essentially the model of the problem that we have to solve is the problem like this to find B k plus 1 that is called them principle of least change.

So here you see the frobenius norm, sorry B square f is the trace of sorry do I have a duster here the trace of the matrix B B, B into B - B square, B square - B into B. So this is the meaning of this. Now you see, this is so this problem is a constraint optimization problem. See, if we do not, so this is one of the type of updating that you want to do there are some others also. We will come to them later on as a example of application of constraint optimization ideas. So this is the constraint problem. See, if we do not have idea about constraint optimization itself, we cannot really make any progress in our understanding of Quasi-Newton method.

Basically, we can just mug up certain rules of updating but, that does not give you the true feeling of what is really happening. So when you learn a subject - a mathematical subject, it is very important to know what the hell is actually going on. So, let us come to the story of constraint optimization. Let me tell you, the story of optimization has a very check out history. Optimization is a very ancient subject; it is not a subject that has just evolved in some 20 years or 30 years or 40 years or even 50 years. It is a subject which at least dates back to more than 300 years or so, 300 years when it started as really been pursued as a mathematical subject.

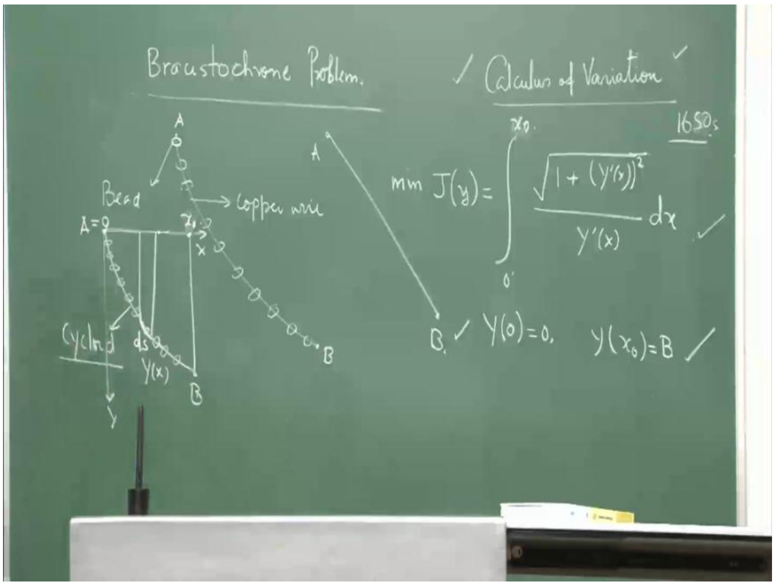
(Refer Slide Time: 10:54)



Now I want to stress on the following fact. Where do I write, may be I will write it here that one of the basic facts that you know that if you have a function from R to R which is a differentiable function. Then if you want to find an x star minimizer of this function of course, then you have to first attempt to find an x star which is equal to zero, f dash x star is equal to zero. This is what you have to first do. Of course, any x star, which satisfies this, need not be a minimum, but if x star is already known to be a minima, it must satisfy this. So, this is the necessary condition and in optimization, one of the major things is the study of necessary conditions, because it tells you how to compute at least a point which you can start suspecting of being your minimum or maximum whatever you want to do.

So here, so if x star is a min, min or local minimizer or whatever I am just writing very loosely so this idea was known to Farma, but during Fermat's time Fermat of the famous Fermat's last theorem, during Fermat's time you really did not have any idea about derivatives. It was done slightly later by Newton and definition developed by our people like Euler and the Bernoulli's. What he proved was a following he is what he tried to demonstrate that if you a polynomial equation, of course, in those days more function algebraic functions are taken as polynomials. So if you have a polynomial equation and then if you polynomial function and then you want to minimize it then at the point of minima or maxima, the point here and here that is wherever there is a hill that is at hill top and wherever there is a valley the tangent at those points would actually become parallel to the x axis that is what Fermat showed. And this is that is why this result is local minima, global minima I have noted in this result is known as Fermat's rule, but however the story of constraint optimization began 300 years ago with a very interesting problem called the Braustochrone problem.

(Refer Slide Time: 13:23)



Let us see what is that problem, I hope my spelling is right, so what is the it was a problem post by John Bernoulli. So, the problem is as follows, it says that I have taken a wire, some wire like this, I do not know some wire copper wire. And I have put a bead here, and I allow this bead to fall freely under its own way that is fall under freely under gravity I just put a bead there and just leave it. So its starts traveling, now this point my staring point A and my ending point B is fixed, because that is the end points of the wire. Now this starts traveling and slipping down. Now the question is so this is a copper wire if you have forgotten and this is a bead and that bead is now running down this wire. The question of course, is what should be the shape of the wire so that these bead will take minimum time to come from A to B a natural instinct is to say it will be just hold as make this copper hold this copper wire straight make the make them into a straight line that is hold the copper wire like this from A to B. Because you say the straight line is a shortest distance, but of course, who told you it is the shortest distance.

Of course, you can say from geometry you know that is the shortest distance between two points and so on the shortest distance it will have the shortest path because the gravity will act possibly in a similar way that you might it might appear to you. But, the answer to this problem is no along the straight wire it does not take the least time it is the least time is taken in somewhere of this shape which is called a cycloid. This problem gave rise to what is called the calculus of variations, so in calculus of variations, you are expected to find here so what I am suppose to do here. If I look at it like this problem A to B then so basically this my x axis, this is my y axis, I am looking at the point on the y axis. And so I have to find the time over particle running down this, so it will be the distance by the velocity, so if this is x and this is y so y is the function of x, so y dash x is the velocity. And basically I have to find the curve y; y is the curve that I have to find.

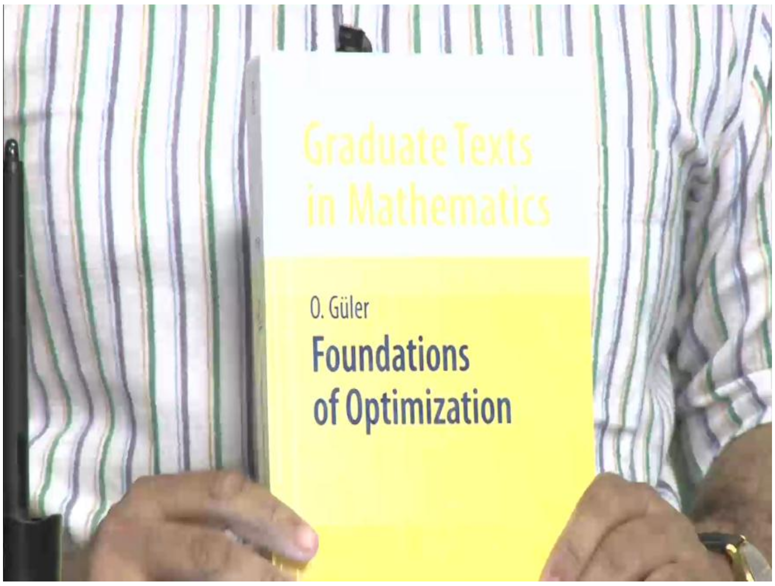
So basically my length is root over 1 plus d y d x whole square, so basically it is y dash x whole square divided by the velocity which is y dash x. So if this because this is the length of d s - elemental arc here d s and this is d s by instantaneous velocity at x which is y dash x. So this so if we integrate it over the whole wire from A to B this is what you get, so x is equal to so from here if you came so here it is something say x is equal to maybe I should write this as my corresponding B point, A is this point. So if this is say a is when a is my zero point and the corresponding point here is say some x naught. So basically I have to now integrate from zero to x naught, but I have to remember that my y of x naught, y of zero has to be zero; at the same time, my y of x naught has to be b; this is b this distance, so this end points are fixed. And now I have to find y, so this is actually a function of y; I need to find a y which will give me this, which will satisfy this as well as minimize this integral. So basically, I have to minimize this integral the distance, sorry the time taken by the bead and it also has to also satisfy these two, so called n point conditions.

But, these end point conditions are actually constraints on the problem, these end point conditions are actually constraints on the problem. And hence what you get is not a standard problem on minimizing f over R n, you get very mathematically involved an existing problem. Because here you have constraints and here you really have to found find a function y a function so of course, there is a question of what sort of function whether it is differentiable, how many times differentiable, how is it is continuity what are its continuity properties etcetera. So in of course, in those days nobody bothered about those continuity properties or differentiability properties or why they just said obviously, it has to be differentiable nicely. They took as good functions as they wanted nice very nice functions for which for every nice good things happened

So of course, now we in a modern days this calculus of variation two three hundred year old this was this problem was possibly given in somewhere in the sixteen fifties I guess if am not mistaken, so it is three hundred plus, three hundred more than three hundred years old problem. But this more than three hundred years old problem is still continuing to give us new insides and has lot of new things and lot of new applications. Calculus of variations is still a growing subject it is a very important area of research in optimization. Now this y from modern point of view has to be chosen from a function space, because it is a function and it besides in a function space those who have any idea about the subject of function, and analysis would know that this function spaces are not finite dimensions like our R 3 or R 2.

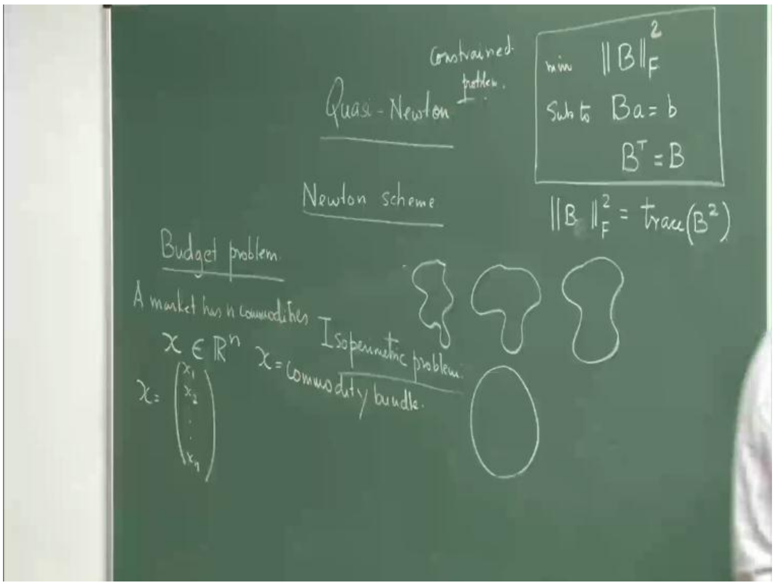
So this function spaces one has to understand are infinite dimensional, so in effect this is an infinite dimensional optimization problem with this sort of constraints. So what you have got here is a constraint optimizations problem, so the initial problem of the this Braustochrone problem gives you a very interesting constraint optimization problem. And in order to solve this problem, not only that the subject of calculus of variations had develop, this term variations is due to the you the technique that was developed by Lagrange. But it is important to note that this problem has given rise to lot of new mathematics in order to solve the major issues here. It is not a very trivial thing by the way, it is not so easy to just do things here, but it is one of the most exiting areas of mathematics and possibly I will do a little bit of a very basic things about it once we have some time we at the end. Maybe it is a good idea to bring in some of calculus of variations, but we will keep on our final dimensional approach and do it, so that is not a very big issue.

(Refer Slide Time: 23:14)



Now what is important is that I would like to also show you the book that I have been mentioning which I will follow while studying the Karush Kuhn tucker condition necessary optimality condition, it is called foundations of optimization by Osman Guler. It is a publication by Springer and it is in the series the graduate text in mathematics. I think those who are doing some advanced work in optimization like either PhDs or even very young researchers should have this book with them. Now this problem also has another story, the calculus of variations was that which which is actual a part of calculus of variations, but there is another old ancient problem that has been given that princess Dido fleeing from the persecution of her brother came to a land which is now Carthage the city of Carthage. And he asked the local leader there Yakub that she needs some land. So Yakub asked how much you need, she said cut a bulls hide that skin of a dead bull and make thin pieces and just sew up the pieces, stitch up the pieces and then see how much area you can enclosed. So the problem is as follows given a curve a closed curve of a fixed length what is the curve which will enclose the maximum area.

(Refer Slide Time: 25:02)

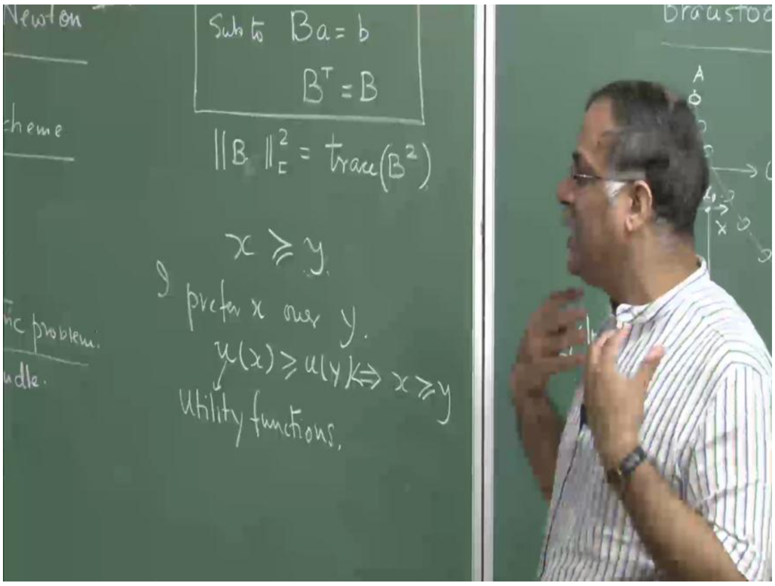


That is suppose I have two curves this is so I have taken a taken a thread like this, and this tread I can put it like, this the same thread I can put it like that, this the same thread, so that lengths are same perimeter is same. Now the question is which one of them will enclose the maximum area? The answer is surprisingly simple and that is where the beauty of mathematics lies answer is a circle. So there Dido took that land and established the modern city of the current city of Carthage, which is there and this problem is called a isoperimetric problem that is the parameter is same, but the one which would enclose the maximum area.

Now largely optimization problems were relegated to this to physical sciences, natural sciences and the constraints which appeared appeared in the form of equalities. Come back 250 year more when we are in the or 300 years and we are in the twentieth century, where during the Second World War and later on it was realized that they are lot of issues in optimization and lot of issues in business, engineering, economics specially. Where you cannot just have a equality constraints you have to impose inequality constraints. Let me give you a simple example, which comes from economics, which tells us how inequalities has become the hall mark of modern optimization. And now the Lagrange multiplier rule which Lagrange had thought to solve the calculus of variations has to be modified to generate a rule, which can handle also inequality constraints. And from that is where the subject of mathematical programming or finite dimension of optimization starts up.

So let us now look at this problem of budget in economics budget problem. Suppose a market has n commodities. So there is a market, and this a market has n commodities. Now if my market has n commodities, how do I represent that market from the point of view of modern mathematical economics, we would represent the market that any commodity bundle must have n commodities. And that would be an element in R n that is x is a commodity bundle, so this has n commodities - x 1, x 2 dot dot x n, so there is say first is rice, atta, dal this that and so on. Now this is the quantity of rice, this is the quantity of atta this is a quantity of so on, so there could be infinite such possibilities of quantities you can choose, so theoretically of course, not in really practical life practical life is somewhat little different. So in order to model it, so I can now say every commodity can be viewed as a vector and this vector x is called a commodity bundle. Now how do I chose a commodity, how do I know that, I want there are two commodity bundles are given to me, and how do I know how to chose it, which one I require, how do I know that.

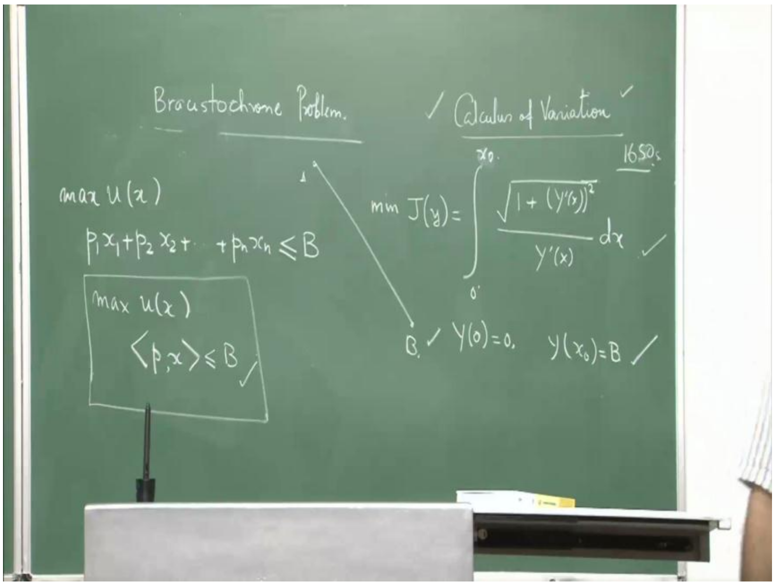
(Refer Slide Time: 29:44)



Now so there the question comes of preference so if I am given two commodity bundles x and y. So if I am given two commodity bundles x and y, how do I know whether I prefer x over y or y over x so I am indifferent I can choose any one. So if I choose x over y or I am possibly even indifferent, there is a symbol which is used in economics, which is called the preference symbol that is I want to chose x over y that is I prefer x over y. This thing means I prefer x over y, now suppose the unit price of the first quantity is p 1, unit price of the second quantity is p 1, unit price of third quantity is p nth quantity is p n. So price of these quantities are given, so this is a fixed vector in R n, price is fixed. Now what happens is that how do I numerically decide whether I prefer the bundle x over the bundle y. This can only be done if I have some functions which will tell me that whenever I want whenever I prefer x over y that function should be such a function u say such that u x would be bigger than u y.

So these sort of functions u are called utility functions in economics. So what you utility function does is the following it tells take x and y and find the values of u x and u y. If u x is bigger than u y, it would imply that I am prefer I would I should prefer x over y. And if I prefer x over y, I should have this. Now in a strict philosophical point of view if I am a utilitarian in the sense that I want to maximize my own life maximize my happiness, so what I have to do is to maximize my utility.

(Refer Slide Time: 32:28)



So I have to choose an x, choose a commodity bundle which will give me the maximum value of u, but this choice cannot be just arbitrary, because I have some fixed amount of money with me that is my budget, say I have b amount of money with me capital b. So if I buy a commodity bundle x my price that I will pay for it is of course, if you do not want to if you want to go step by step, it will be p 1 x 1 plus p 2 x 2 plus p n x n. So if I buy a commodity bundle x 1, x 2, x n this is what price I should pay, but this price cannot exceed the budget I have or at least equal to the budget I have, because I have limited amount of money. Now which means in general, I have to maximize u x subject to this constraint, so it is maximizing u x subject to p dot product x or inner product x u those who know very basic vector calculus just very basic liner algebra would know that. This is the inner product and so this is what I intend to do of course. Now can impose condition that x 1, x 2, x n has to be greater that equal to 0; of course, if I have x 1, x 2 negative means I just do not negative or basically 0, I do not buy, you can also put in that restriction x has to be greater than equal to 0.

So what you have here is actually a minimization of a function over a set of in terms of certain linear constraint but, these are inequalities. So the inequalities is very much real in modern day applications, and so here how would you handle and try to solve a problem with inequality constraints and that is one of the major hall marks inequality is remained to be the major hall marks of modern optimization. So our goal would be to first study problems with inequality constraints. You might ask why not study this, this involves lot of techniques from modern analysis which might not be known to all the students or all the viewers of this course. So we would go into something which is more manageable and done through a very beautiful mathematics of convexity of convex sets and functions, so we will try to first start understanding the constraint optimization problem with inequalities. Then our next step would be add equalities to it and and get and see the Lagrange's part, I mean, it is all its beauty; of course, as I said that we will go into this problem later on at a certain stage when we have some time. I cannot promise it but, I will try my best to do that, you need it is very important to have some information about this kind of problem this problems. And anybody who wants to be a optimization optimizer in the feature should really know this problems.

Now how do I go about it, what is the first question that I should ask about it. The first question I should know about it is whether this problem has a solution. When we will such a problem inequality constraint problem will have a solution, so this is a very general question when will this what when will this kind of problems have a solution. It says it is not so easy to immediately tell that how I looking at a problem whether this will have a solution or not but it is very important to know know when will a particular class of problems - optimization problems will have a solution, so that is one thing we will need to know. We will give a brief out line when a general optimization problem or constraint optimization problem will have a solution, which depend on the nature of the feasible set. So these are the set of all these are the constraints the set of all access which satisfies this in R n would be the feasible set associated with this for example, this utility problem. Now the important question that lies ahead is that if I possibly know that this problem has a solution, how do I go about and find it, so there must be some way to first get a point, which I can start suspecting as my minimum.

So this is the question of asking what are the conditions, which are necessarily followed by a local minimizer or a global minimize. So if I have a local minimizer that the local minimizer must satisfy this condition, and hence if I find a x star which satisfies those condition then I can start suspecting that might be chosen the my local minimizer. So the first step is to know after learning a bit about the existence of solutions, what are the necessary conditions for optimality of the existence of a local minimizer - that is if x star is a local minimizer the constraint optimization problem under inequality only for the time being. What are the necessary conditions for optimality and how they are helpful to us. So we will stop with this and we will take it up from the next lecture.