Dynamic Data Assimilation
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture – 08
Optimization in Finite Dimensional Vector Spaces

The last three modules, sub modules in fact, we talked about financial vector space matrices, tools for multivariate calculus; there is yet another topic which is basic fundamental to pursuing data assimilation that is called optimization. Optimization in finite dimensional vector spaces.

I keep referring to finite dimensional vector space, because all the computational problems in numerical analysis and all the applications where we use computers to compute the solutions, the basic mathematical framework is finite dimensional, vector space in practice, we can do only finite things, we cannot do infinite things therefore, finite dimensional vector space is the appropriate background, on which the entire theory of computation has been built around and to emphasize that I keep referring to the importance of recognizing the role of finite dimensional vector space. Now, why optimization?

We earlier saw in module one a broad introduction to an overview of data assimilation, the data assimilation can be thought of as curve fitting. Data assimilation can be thought of as regression analysis, data assimilation can be thought of as identification. Data assimilation can be thought of as estimation. So, let us take the point of view of estimation, whenever we want to estimate, we want to be able to estimate optimally, what does it mean? We want to be able to get the best estimate, best optimum optimization. So, optimization theory is fundamental to pursuing estimation theory, estimation theory and optimization theory are interrelated in the fact, I use principles of optimization in estimation theory is a topic within statistics.

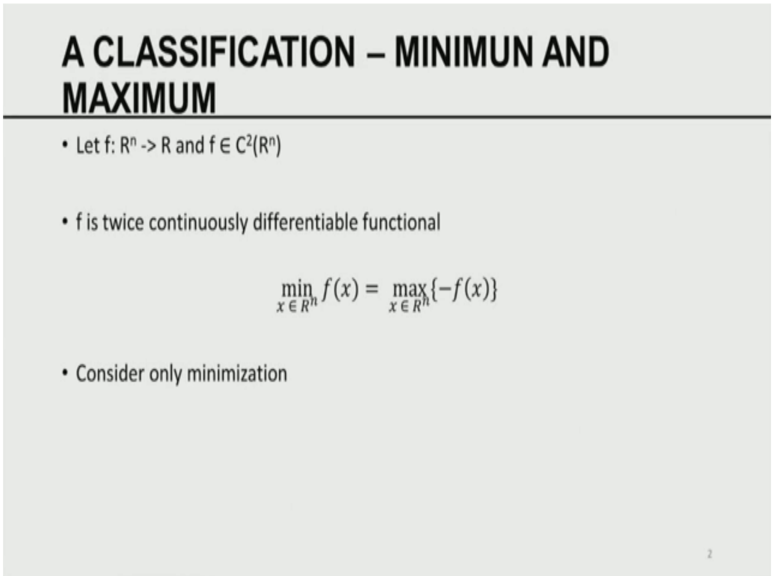
Optimization theory is a topic within multivariate calculus. So, it is the interaction between the two that provides the ability to optimally estimate. So, we would like whenever we do, whatever we do, we want to have the best prediction. We want to have the best estimate, we want to have, the best way to tell what the temperature; in

Bangalore would be tomorrow afternoon so on and so forth. So, we are always seeking for the best means optimum.

So, we need to be able to have a clearer understanding of the notion of optimality, when something optimum and the properties of optimum can be maximum or minimum, when you try to talk about cost functions. I would like to be able to minimize the cost, when you tried to talk about profitability we all want to maximize profits. So, in economics the aim is to be able to maximize profits.

So, in economics most of the problems are posed as maximization problem in engineering, sometimes we talk about minimization of the energy to be able to accomplish a particular task in the case of estimation, we would talk about minimizing certain magnitude of errors. So, maximization minimization are parts of the optimization theory. Maximization and minimization are truly interrelated with each other. So, in this sub module, we are going to be reviewing principles of maximization and minimization.

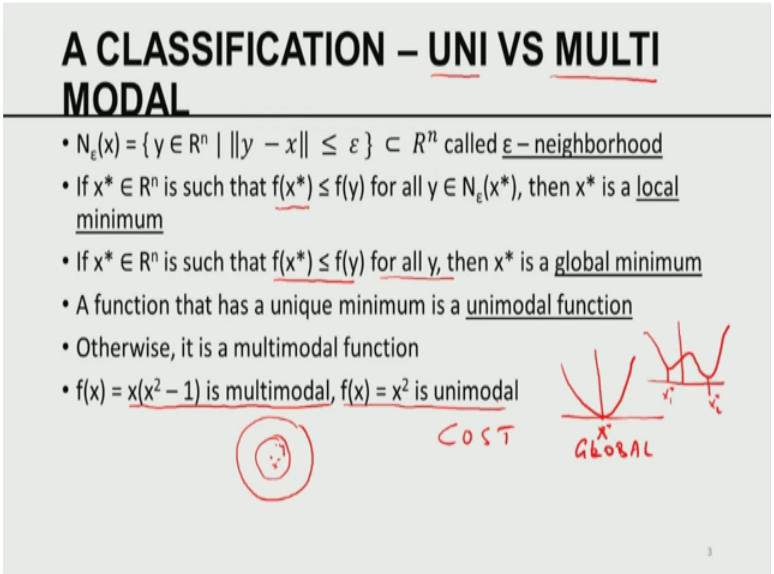
(Refer Slide Time: 04:02)



So, a classification let us be a scalar valid function. You can see the notion of a scalar valid function comes into play right away. Let us b c to R n; that means, f is a function which is continuously differentiable in R n at least twice continuously differentiable, if f is twice continuously differentiable functional. So, if it is a scalar value function it is called a functional. What is the relation? Minimum of x minimum of f, with respect to x is the same as maximum of minus f that is, what it is enough; I study either maximization or minimization, which are loss of generality.

We will consider minimization. So, that is the general idea in the start that we need to understand. So, you do not have to do maximization and minimization separately, because of this intrinsic relation between maximum and minimum. It is sufficient either to do maximum or minimum, we will do minimum.

(Refer Slide Time: 05:04)



A classification of minimization problems occurs in various shapes and forms, the first classification is with respect to modality a minimization problem can be unimodal or multimodal. In a unimodal minimization problem, the problem has only one unique minimum. For example, if you have a cost function, which is parabolic, which is a cardiac function. This is a unique minimum, but if you have some non-linear problems, you may have a function which may be like this; in this case there is a multiple minimum.

So, unimodal function versus multimodal function. Uni modality multi modality who creates this, the cost function that we use is the one that creates uni modality multi modality. It is easier to do, you need moral minimization as opposed to multimodal minimization.

But at the outset, we would like to be able to distinguish between the existence of two types of minimization problem. Now, let us describe mathematically, what is the minimum. Let f be a function, the next star be a point at which the function attains the minimum and why do I say the, how? What is the property of the function of the minimum, if I consider the value of the function at the minimum, if I consider any point which is in a small neighbourhood of the minimum.

So, let us draw a little picture, let this be the minimum point x star. If I consider a small neighbourhood around that, this is why the value of the function at y in a small neighbourhood of x star is always larger than x star. So, in this case x is called the local minimum. So, for example, in this case, this is the local minimum.

So, this is x 1 star, this is x 2 star, in this case, this is x star. So, x x 1 star is a local minimum, x 2 star is a local minimum in the sense that if you go in any direction away from the minimum, the value of the function is going to increase, that is the that the basis in here, if x star is such that f of x star is less than f of y for all y, then it is called a global minimum.

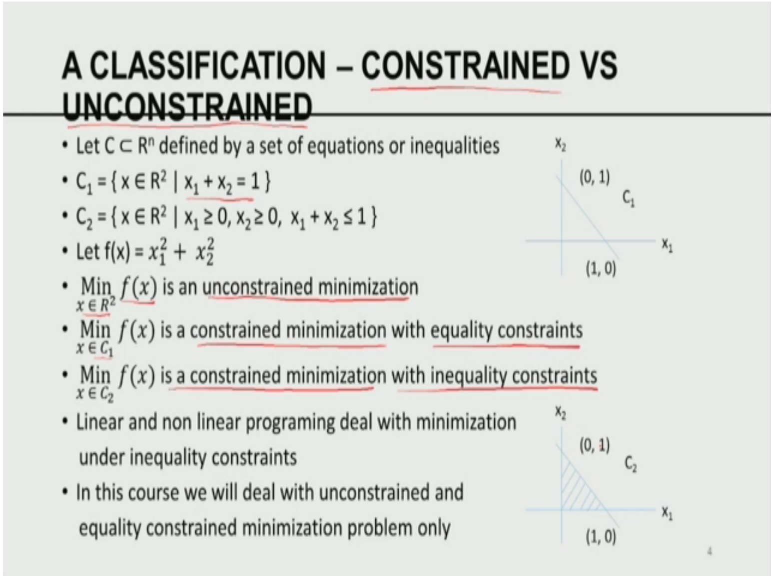
So, in this example; x star is a global minimum, this is global, here this point x 2 is a global minimum, x 1 is also a minimum, x 1 is a local minimum, x 2 is a global minimum. There are multiple minimum, uni minimum. Multiple minimum; your function is a unique minimum, is called unimodal function, otherwise called multimodal function. The function x times x square minus 1 is multi modal, the function x square is unimodal.

I would very strongly encourage to be able to plot these functions and see where the modality occurs, that where the minimum occurs where the maximum occurs. So, if there are going to be multimodal that it will be multiple minimum and multiple maximum to multi modality is always a headache. Optimization problems are manmade. So, when you are trying to create a particular problem do not create too much of trouble for yourself.

Choose the cost function such that it is endowed with unique global minimum, if you can arrange the problem such that it is endowed with a unique global minimum, your headache will be lot less, but if you formulate the problem, such that it happens to have multiple minimum, you are to scratch your head finding.

A global minimum will be a much more difficult problem of course, this is easily said that, than done for a given cost function, some problems may be endowed with one minimum, some problem may be endowed with multiple minimum, if the problem is endowed with my (Refer Time: 08:59) we have a duty to find all minimum to be able to pinned down, who is the global minimum, what is the best minimum.

(Refer Slide Time: 09:09)



So, unimodal verses multi more, the next classification is constrained verses unconstrained, again those of us who have done a little bit of meteorology or data simulation should be able to appreciate minimize f of x that is an unconstrained problem. I have no constraint on x, but I would like to be able to insert a constraint. Now, let us c be a subset of R n defined by a set of equations are inequalities for example, let us c 1 be the set of all x such that x 1 plus x 2 is equal to 1, let c 2 be set of all x in R 2, x 1 is greater than 0, x 2 is greater than 0, x 1 plus x 2 is less than or equal to 1.

So, let us understand those constraint. Now, the first constraint c 1 is given by this line, maybe I will use the arrow the will. So, in this problem x 1 plus x 2 is 1. So, what does it mean, even though there are very many points in the 2 dimensional plane. I am interested only in those points that lie on this line; the line is the constraint, this second one. On the other hand x 1 must be greater than 0, x 2 must be greater than 0, x 1 plus x 2 is less than or equal to 1 and that essentially tells you the region in inside the triangle.

So, these are the constraint, could be along a line, along a curve. It could be a sub region. So, what is an unconstrained minimization problem? An unconstrained minimization problem is started like this; given f of x minimize x with respect to all point in x square. In our square there is no constraint, you can go anywhere wherever it leads to, that is called unconstrained minimization problem.

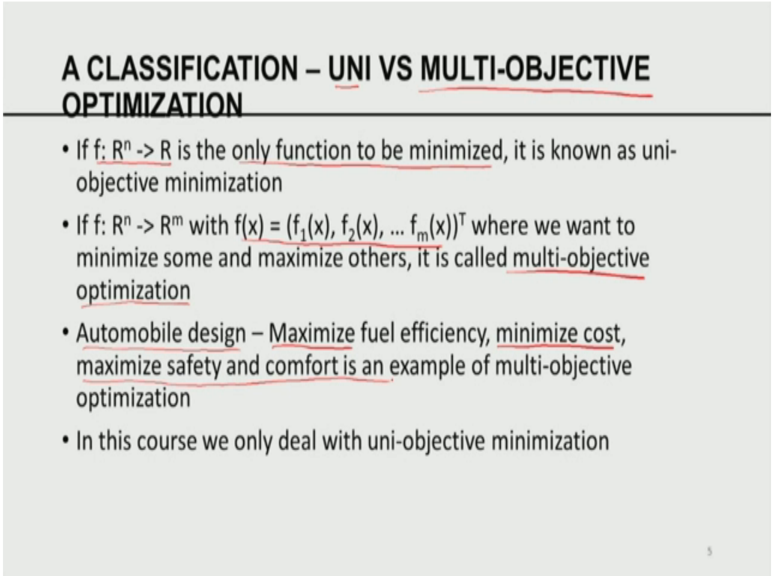
Minimize f of x not over all x, but over x belonging to c 1; that means, I am not interested in every x, but I am interested only those x that lie along the line say, that is a constrained minimization problem, this is a special form of constrained minimization problem, because the constrained set, which is, line is given by the equality constraint. So, this is an equality relation x 1 plus x 2 is equal to 1 is an equality relation, x 1 plus x 2 is equal to 1, there are infinitely many point that satisfy that equality. So, also on, I am interested in minimizing f of x not over the entire x, but along that particular line c 2.

I would like to be able to minimize f of x constrained minimization problem. This is called equality constraints, because excess have x 1 is greater than equal to 0, x 2 is greater than equal 0, x 1 plus x 2 is less than or equal to 1, that is called equality constraints, most of the problem in meteorology are either constraint, are other unconstraint or constraint with equality constraints.

In operations research they are often deal with constraint optimization with inequality, constraint of these three problem, constrained optimization inequality constraint is the most difficult one, but these problems are now very thoroughly understood, this body of literature represents one of the most thoroughly understood disciplines within optimization theory.

The theory of linear and non-linear programming deal with minimization, under equal inequality constraints, you may have heard of linear programming. What is linear programming? Linear programming deal with they deal with minimization and our inequality constraint in this course, we will deal only with unconstraint and equality constraint minimization problem. Why we try to formulate the problem in such a way?

(Refer Slide Time: 13:05)



So, equality constrained is easier than inequality constrained further classification. So, we talked about unimodal verses multi modal, constrained verses unconstrained. Now, I am going to talk about uni verses multi objective functions, there is only one objective, there are multiple objectives. Let f, if the f is R n to R n is the only function to be minimized then it is known as uni objective, when f is a vector valued function, if a m components where we want to minimize some component, maximize some other component, it is called multi objective minimization optimization.

Let me give an example, let us take an example of an automobile design, is one of the hard problems. I want to maximize the fuel efficiency, I want to minimize the price, the cost, I want to maximize the safety and comfort are you with me. So, an automobile design, if it tell him I want you to give me a car with the maximum efficiency, he will give you, but is of no safety, no comfort.

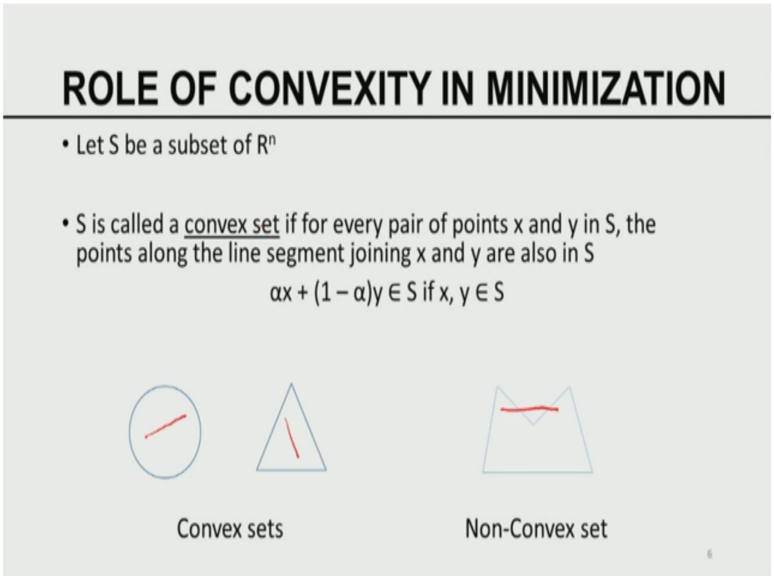
I want to give you a car which is best to comfort, but it is very few miles per gallon, the real automobile engine design, is there are multiple objective that do not, that are consistent with each other and automobile engineers have come up with strategies to be able to solve this multi objective optimization problems. Fortunately, in meteorology we are always dealing with one uni objective.

So, operation to such people dealing with multi objective optimization and the top of inequality constraint, wow that is some of the toughest minimization problem one can deal with the conceptually the optimization problem that often occurs within the context of meteorology oceanography. Dynamic data assimilation are the easiest of the problems, but we say is difficult, it is not become conceptually difficult is, because of the size of the problem in meteorology, the curse of dimensionality they are interested in solving large problem, but simple problems in operations is on the other hand they may be solving small problems, but are much more complex then compare to problems in meteorology.

So, that is a different, difference, I would like to be able to bring about. So, if you meet with them, automobile design engineer talk to him or her as to what kind of problem, how do they optimized. So, that we can learn from the kind of methodology they use. So, multi objective, you can think of all the things.

Now, uni modal, multimodal constraint equality inequality and then you need to multi objective, if you stack all of them together, you get a broad overview of this discipline called optimization theory and that is what people in operation research. There are two groups one develops the fundamental theory, others talk about applications of this theory to various problems of in does of engineering and industrial interest.

(Refer Slide Time: 16:15)

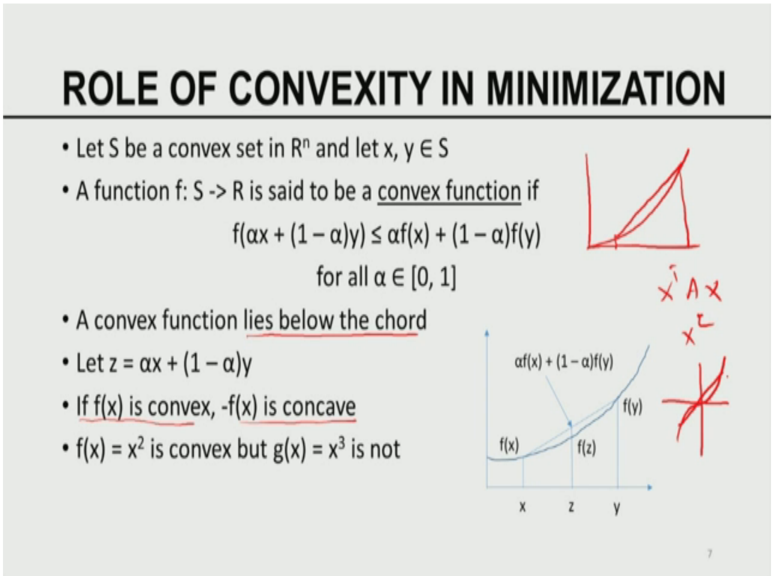


The next one is the notion of convexity and optimization are intertwined and I would like to bring about the beauty of the role of convexity, the notion of convexity in an optimization.

So, it takes a little bit of a definition. So, let s be a subset of R n, I call yes a convex set if for every pair of point x and y in s. If I join them by a line, the line segment joining any two x and y completely lies in acts for example, in the case of a circle, if I pick a point in the case of a circle, if I pick a point here, if I pick point here, if I strictly the entire line within this, if I get this the entire line within this, but if I pick a point here, if I pick a point here, part of the line is outside the same.

So, this is not convex these are convex sets, the notion of convexity essentially includes every point is such that line segment, joining them in include. So, you can think of that. Yes, sphere is a particular, a circle is a perfect example of convex set. A convex object, I have to define, what is called convex functions, but before I talk about convex functions, I am now talking about convex set.

(Refer Slide Time: 17:44)



When is a set convection that is the definition of a convex set, when do I say a function is convex, I will tell you pictorially, I am sorry, the picture is not perfect, but you get the idea, I have a function f of x. You take a point x f of x, y f f f of y, you draw f of x and f of y, join them by a line that is a chord, in this case the entire function lies below the chord.

So, what is an example of that, if you have x square, if you took two points any two points on a parabola, if you draw the parabola, the parabola is always lies below the chord, any function that has this property in general called convex function. x square has a unique minimum.

So, convex functions in general have unique minimum, that is whether also notion of a convex that it comes in, we are interested in quadratic functions; all the cost functions that are interested in data assimilation are all quadratic functions. Why quadratic functions has caused quadratic functions appropriately formulated is convex function have a unique minimum, I do not have to pull my hair trouble myself to be able to perform minimization. So, that is where the notion of the convexity comes in. So, yes it is a let s be a convex set. Let x and y be point in s your function from s to r.

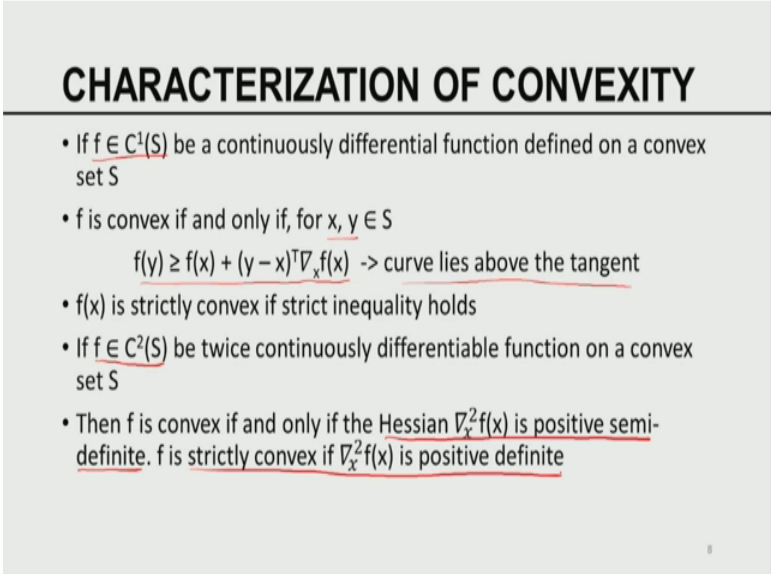
So, the underlying set, over which the function is defined, is a convex set are you all in place. So, there is an underlying set, which is convex, there is a function defined over that, if I take any two points in the convex set and evaluate the value the function and draw a chord, the function lies below the chord, what an example x square.

So, you have function f from s to RS, it to be a convex function, if f of alpha x plus 1 minus alpha x is less than or equal to alpha times f of x plus 1 minus alpha times f of y for all alpha 0 to 1. Now, look at this now when alpha is 0 this becomes f of y when alpha is 1 it becomes f of x for every point in between the function is below the chord. So, that is the definition, the function is lies below the chord. So, if f of x is convex minus f of x is called concave.

So, for you, how would I have imagine a convex function f of x is a typical example of a convex function. x transpose x x transpose, x is a typical example of a convex function, x square is a typical example of a convex function, concave functions, for maximum convex function for minimum maximum and minimum our dual of each other concave functions and convex functions are dual of each other.

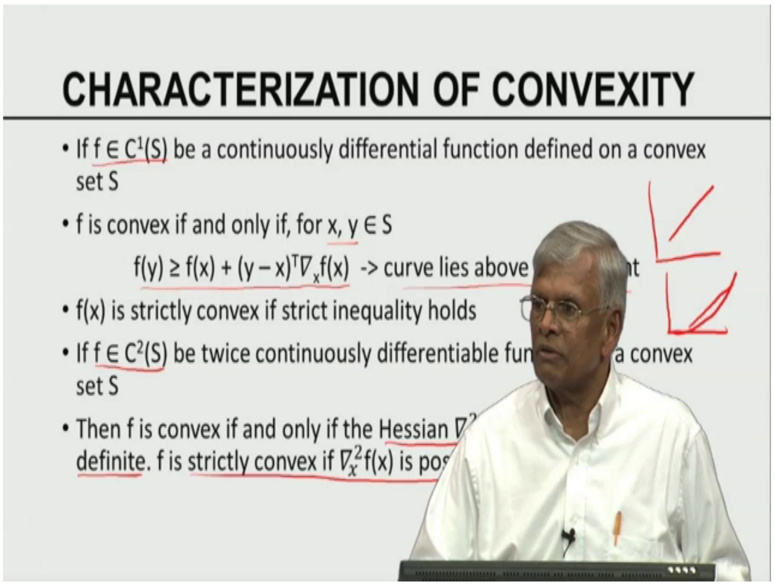
So, within the context of minimization we are in generally interested in convex function convex sets. So, x square is convex x cube is not convex. What is the plot of x cube that is x cube. So, if I took a point here, if I take a point here, some part is less, some part is above. So, x cube is not convex x square is convex. So, that is a model by which you can go with your notion of convexity convex function. So, we have defined convex set, we have defined convex functions.

(Refer Slide Time: 21:30)



So, there are many different ways of characterizing convexity, let me quickly run through them in the previous definition of convexity, I did, I simply assume that there is a function. I did not assume any differentiability or anything or f is a function, I defined it to be a convex.

(Refer Slide Time: 22:55)

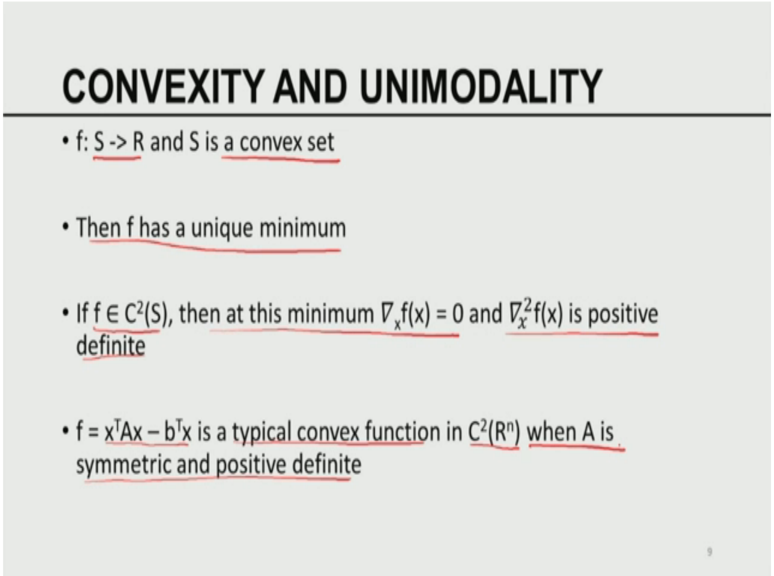


Now, suppose I know a little bit more, you say in addition f is c 1. a c 1 function is said to be convex, if for any 2 x and y f of y is greater than or equal to this the curve lies above the tangent a that is an equivalent definition, if on the other hand, if f is in c 2 twice difference continuous functions the that f is convex, if and only if the Hessian is positive definite positive.

In fact, it can be called positive semi definite. It just are strictly positive, definite do this is a distinction place. So, any f f in c 1 f in c 2. for a c 2 function, if the hessian is positive semi definite, it is convex, if it is strictly positive definite it is called convex, all of you might be for example, let us take a straight line. is it is convex. Second derivative is 0, are you with me please.

So, this is convex, but this is strictly convex, this is strictly convex, a straight line is simultaneously, is a convex and concave, because is the, is a separation between convex function and convex (Refer Time: 23:21) straight lines. So, you have examples of convex functions.

(Refer Slide Time: 23:26)

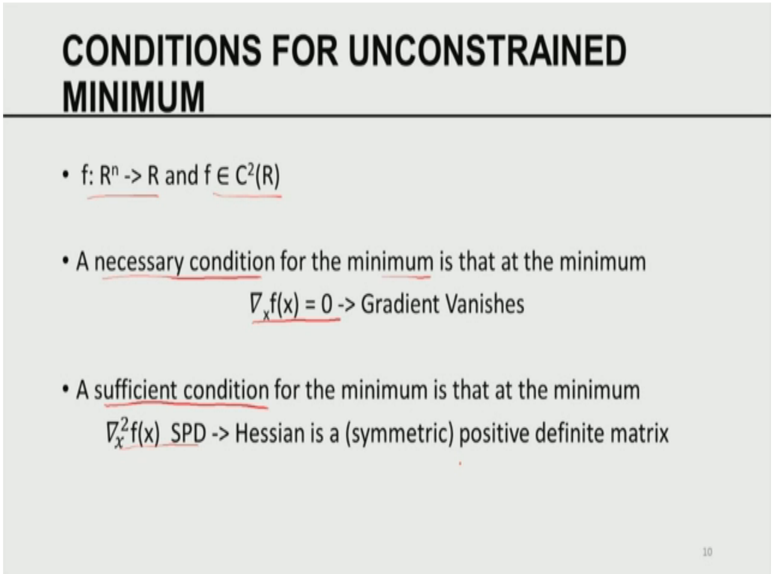


So, what is the, why are we interested in a convexity, you always talked about uni modality and convexity are close cousins of each other uni modality refers to functions with unique minimum global minimum unimodal problems are easier. So, uni modality and convexity are intimately associated with each other.

So, let us talk about this, now let f be s con. Let us be a convex set sorry, let f be a convex set f be a function from s to r. So, what does it mean f is a real valid function defined over a convex set, then f has a unique minimum that is a theorem. I am not going to prove this. Theorem is the theorem in convex analysis, if f is on the other hand is in c 2, then at the minimum the first derivative is 0, second derivative is strictly positive, definite at the minimum first derivative is 0 and second derivative simply positive, it means what that point is a very many minimum.

So, what is the typical example of a convex function, f is eq f is equal to x transpose ax minus b transpose x is a typical function that is in c 2, when a is symmetric and positive definite. Now, you can see all the tools that we have developed in optimization theory in matrix analysis, all comes into a hue now beautifully.

(Refer Slide Time: 25:56)



So, this is quadratic form this is a liner function, this is a general quadratic form. So, if you look at all the functions that 3 dvar, 4 dvar. All talk about, they are all functions of this type, they are typically convex, they are also typically in c 2, they have a as the Hessian is symmetric and positive definite. So, by definition by design, all these functions are uni model, convex functions everything is beautiful.

So, the role of convexity, in inducing uni modality is one of the fundamental beauty of the underlying mathematics, that one needs to have an appreciation to be able to see why? When we develop problems, when we develop objective functions we always think of objective functions as quadratic forms or quadratic objective functions and that is where the importance of convexity uni modality comes into play.

So, with that as a background now, I am going to run through conditions, for the existence of minimum, please understand algorithms tend to compute the point where the function attains the minimum, but before you start your computation somebody has to guarantee that there is one that exists there is unique.

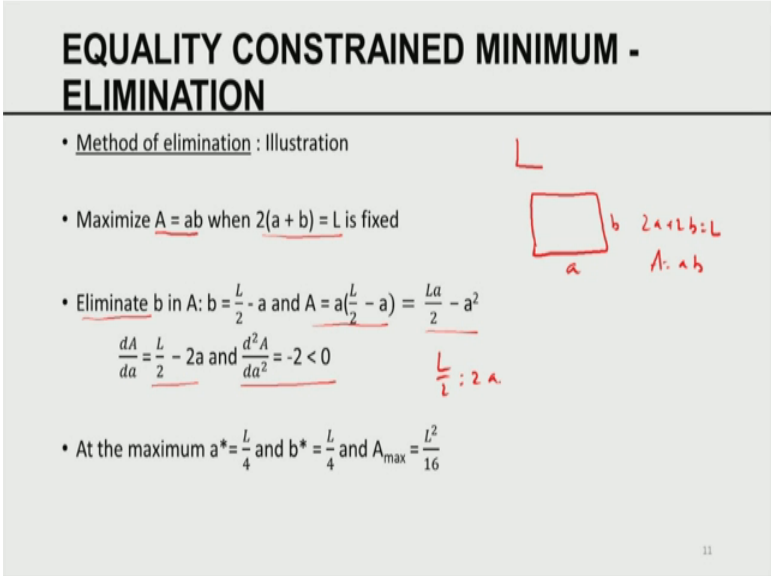
So, unless I know something, exists I cannot go and find it. So, mathematics helps the first level of mathematics helps you to prove existence and uniqueness of minima, once you establish the existence uniqueness of minima then you develop algorithms to be able to seek it. So, characterizing the properties of minima and maxima and then algorithms to see the minima and maxima as fast as possible, as efficiently as possible. These are two complementary aspects of the optimization area.

So, conditioned for the constraint min condition for the constraint minimum, again most of us know from basic calculus, but in calculus we talk about univariate function. Here, I am talking about the corresponding results for the multivariate function. A multivariate function means is defined over a vector, the valorous at the value scalar, f is in c 2 twice differentiable, twice continuously differentiable, a necessary condition for the minimum to exist is that at the minimum the gradient must vanish a sufficient condition for the minimum is that at the minimum, the second derivative must be positive definite; that means, the Hessian is a symmetric positive definite matrix Hessian is in general symmetric, it need not be positive definite. Every positive symmetry need not imply positive definiteness.

But when you consider positive matrices, you need to consider only symmetric matrices that comes essentially from the theory of quadratic forms. So, that is where the whole thing comes into play that is why Hessian is very important the tool from multivariate calculus that characterizes the second derivative.

So, what do I mean by saying the, this Hessian is positive definite. I do not the minimum the function, looks like a bowl, a punch bowl. So, the minimum is a valley. So, how do I characterize the valley at the minimum point in the valley? The function is convex, the function looks like a parabola. So, parabola x square a parabola is model for such minimization process, which is related to the quadratic function.

(Refer Slide Time: 28:45)



Now, I am going to talk about once we have talked about the conditions, for the minimum. I am now going to talk about equality constraint problems. This is simply an algorithmic process. I am going to illustrate by an example; let us suppose this is a problem we generally do in univariate calculus, I am sure every one of us would have done, I would, I am given a rope of length l feet. I am going to, I am as to enclose an area with this, let a be the length b be the width 2 a plus 2 b is equal to l.

So, I am given a rope of fixed length. What is the idea, here I am going to have to use the rope to enclose an area, the area a is ab a is the side b a and b are two sides of the rectangle. So, what is the idea, when a when 2 a plus 2 b is fixed, as l how do you maximize, the area a times b. Suppose, somebody says a, I am going to give you a rope, the area that you can enclose by that rope is going to be yours free.

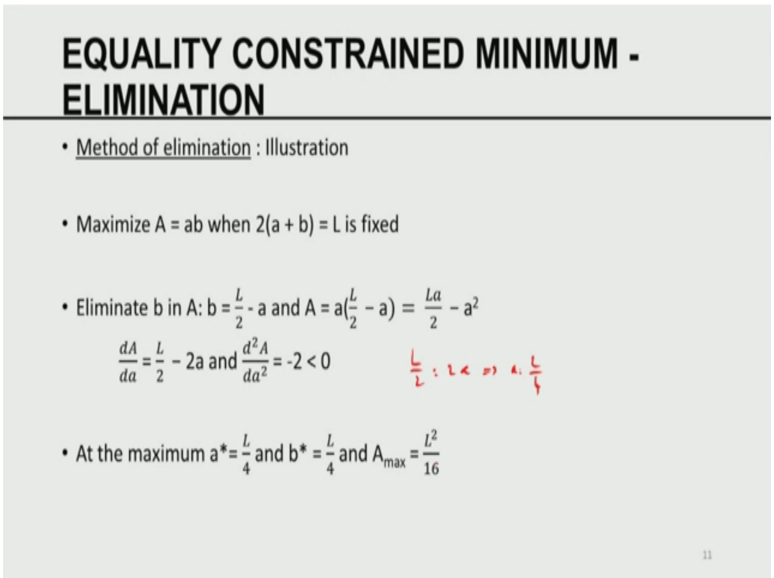
So, humans are built in greedy. So, you want to be able to say hey, with this rope I want to be able to enclose the maximum area that you have to, that I can. So, that is the problem now.

So, maximize it, a is equal to a b, when two times a plus b is l. So, now, you can see is a constraint problem is a variable b is a variable a and b are not independent, if a and b are independent variable, when is this maximum infinity a is infinity b is infinity, but a cannot be infinity b cannot be infinity, because a plus b is l by 2 or 2 times, a plus b is l. So, I have to solve a problem with the fixed length rope.

So, that is the constraint. So, this is the equality constraint. So, I have to maximize under equality constraint, what is a simple way, I am sure every one of us, if you cannot have garden, here b as degree until you have solved this problem, once in your life I am sure every one of us have solved.

So, what is that, we do, we first eliminate b from this constraint, you can simply say b is equal to l by 2 minus a. So, if you substitute be in the a becomes this. Now, a becomes, a quadratic function in a, you compute the derivative of capital A, with respect a little, a this is the derivative you compute the second derivative. It is negative, the second derivative positive means minimum second derivative, negative means maximum. So, the a obtained by solving da by da is 0. So, when l by 2 is equal to 2 a or a is equal to l by 4.

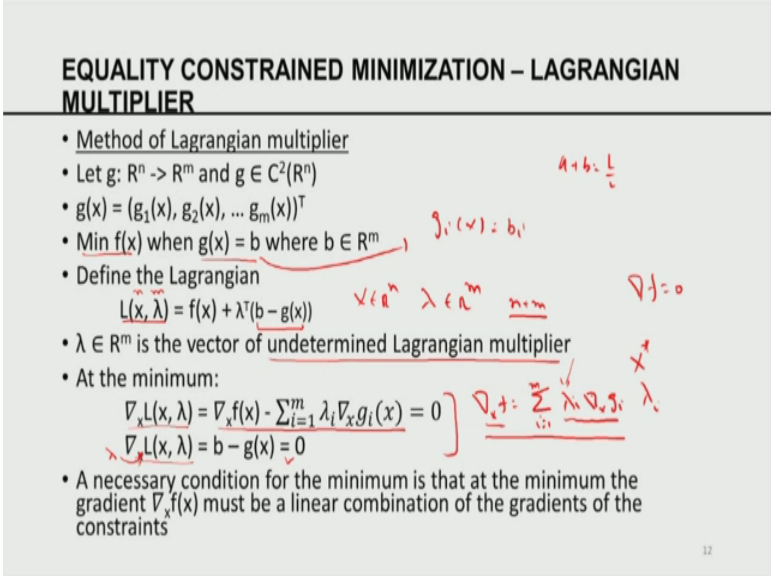
(Refer Slide Time: 32:04)



So, the first derivative is 0 means l by 2 is equal to 2, that implies the a is equal to l by 4, then a is equal to l by 4 b is also a l by 4. So, at which time the area's maximum the, maximum area is l square by 16.

So, this is the method for solving minimization problem under constraint. So, what is that you do use the constraint eliminate one of the variables. You convert that two variable problem into uni variable problem. Apply the principles of calculus, you solve the problem method of elimination, that is the illustration for solving equality constraint problem, this is easy for small number of variables, but in meteorology you have tens of thousands of variables you cannot do this, but again it is essentially an example.

(Refer Slide Time: 32:49)



So, what is the general method for solving equality constraint problem, the classical Lagrangian multiplier. So, I am going to quickly, run over the framework for classical a Lagrangian multiplier so, equality constraint minimization problem lagrangier multiple multiplier method.

Let g be a vector valued function I want to be able to minimize f of x under the constraint that g of x is equal to b. What is b b is a m vector g x is a function, is a vector valued function. So, g refers to g m g 1 g 2 g m b refers to b 1 b 2 b m. So, what does this constraint refers to g i of x is equal to b i for I running from 1 to n. So, each of this is going to constrain.

So, I would like to be able to minimize, this under equality constraint f of x is the non-linear function g is a non-linear equality constraint in the previous problem, whatever the equality constraint a plus b is equal to l by 2 that is a linear function a is a variable b is a variable. They occurred in the first variable, first degree here, g of x g 1 of x g 2 x. I do not know, what it is? It could be any function. So, in general is a non-linear function. So, define a Lagrangian, which is a sum of f of x b minus g of x that is the constraint, you lambda is a vector lambda transpose b minus g of x is a scalar add that scalar to f of x. So, that becomes a new Lagrangian function.

Now, x is a vector belonging to R m lambda is the vector belonging to R m, because g is a m vector b is the m vector lambda must be a m vector.

So, I have now defined a function, where this is n long, this is m long. So, the total number of variables is n plus m and expanding the space or which I need to do the minimization. This is the technique that lagrange are designed a number of years ago. What did he say? He said the following, what is the theorem, the theorem is as follows, the minimal of x, then g of x is equal to b, the constrained minimum of f of x is equal to the unconstrained minimal of x.

He converted the problem of constrained minimization to an unconstrained minimization we know how to solve. Constrained minimization are difficult. So, what is the method you convert a constrained minimization problem into unconstrained minimization problem, but what if the prize we are going to pay. The constrained minimization problem is a n dimensional problem, because x is m vector, but the resulting unconstrained minimization problem is a m plus n variable is a larger space.

So, by expanding the space over which I am going to minimize, I can convert a hard problem into an easier problem (Refer Time: 36:00). So, that is the fundamental idea. Lambda is called the undetermined Lagrangian multiplier. So, what is that an constraint minimization problem. How do I solve this? You compute the gradient of l with respect to x equated to 0, gradient of l with respect to lambda, the second one must be lambda, I am sorry, this must be lambda.

So, you compute the gradient with this x grading with this (Refer Time: 36:28) lambda equate them to 0. The gradient computation, I have already talked about, when I talked about multivariate calculus how to compute gradient of various types of functions. Now, you can see why we do all of you with me that is right.

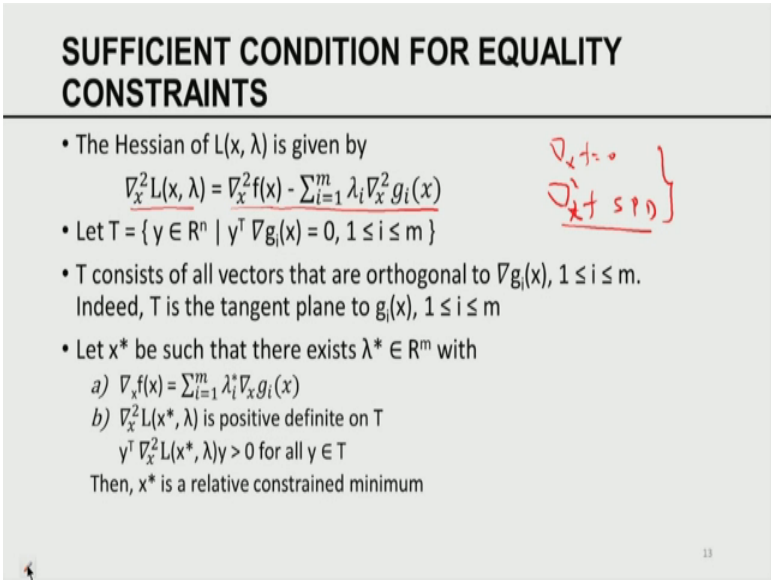
So, this gives you a set, these two give you a set of necessary condition, a necessary condition for the minimum is that at the minimum, the gradient of f of x. So, what does this essentially say, if b is equal to g of x. So, if this constraint is satisfied, this is trivially true. So, if the constraint is (Refer Time:37:02) satisfied, what does the first one say the first one, essentially tells you gradient of x with respect f is simply summation of lambda, I gradient with respect to x of g i i is equal to 1 to m.

So, each g has a gradient lambda. I is, are the constants, this is the linear combinations of the gradients of the components of g f of x is the gradient of f. So, at the minimum what does Lagrangian theory say, is one of the most beautiful results in applied mathematics. It says a necessary condition for the minimum is that, at the minimum, the gradient of f of x must be a linear combination, that gradient to the constraint functions.

In an unconstraint minimization, what must they gradient value delta f must be equal to 0, that is unconstraint in a constraint, the gradient must be equal to the linear combination of the constraint functions, the coefficient of the linear combinations are the lambdas. So, by solving these two equations simultaneously, you not only find the minimum x star, but also find the value of lambda, which are used in this linear combination.

So, you kill two birds in one stroke, you kill two birds in one stroke. So, either you solve by minimization or constraint minimization by elimination as we did in this simple example, which is feasible only for small dimensional problem, if the problem is a large dimensional, the only recourse to solving equality constraint minimization problem is Lagrangian multiplier method. Lagrangian multiplier method, the Taylor series expansion these are very fundamental tools in doing many things that we do in optimization theory.

(Refer Slide Time: 38:54)



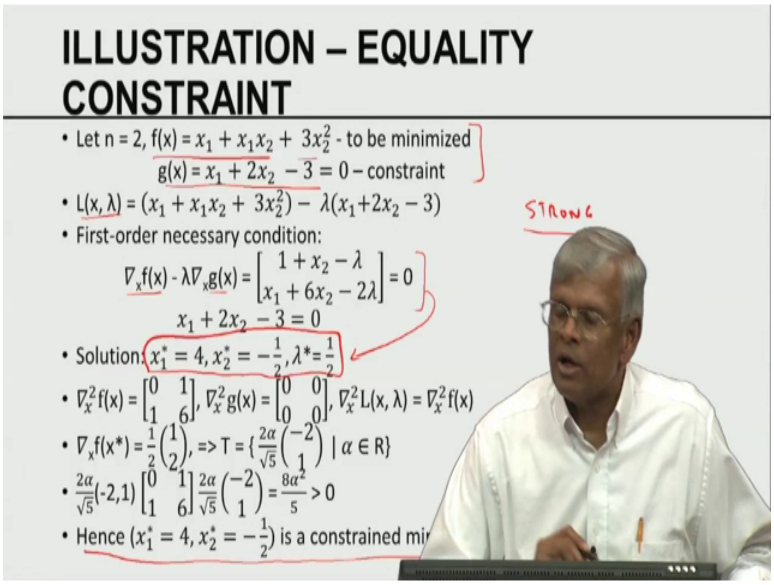
We can also talk about a set of sufficient conditions. I am not going to go over the details of this, but I want to, I want you to recognize the following, I am going to talk about the need for sufficient tradition in an unconstrained set up. In an unconstrained set up, what is that, we have talked about the gradient of f must be 0, the Hessian of with respect to x must be s p d, this is the unconstraint characterization of is that I am sorry, this is the let me I made a mistake, I have to erase this part, if the Hessian must be a s p d.

So, in the previous slide, we only talked about the necessary condition. First derivative, I have not talked about the second derivative condition, R is replaced it. There is a first derivative condition and second derivative condition, first derivative is necessary second derivative sufficient both were constrained, both were unconstraint. So, the sufficient condition for equality constraint problem is that the Hessian of l. So, that should not be surprising that is the reason, I went into this, in the case of unconsidered problem. The Hessian must be symmetric positive definite, what is the analog of that, if you consider the Hessian of l with respect to x, which is given by this, that must be positive definite in an appropriate set of space.

So, with the necessary condition and sufficient condition we have shown the conditions for existence and uniqueness of minima then there is equality constraint that is the fundamental talent of this. I am not proving the derivation of sufficient condition as I have not proved several of the claims, I want you to remember all the theory, that we have covered in matrix theory is the half course, in linear algebra all the topic, we are covered in finite open vector space is one third of a course. All the topic I am covering in optimization literature.

Optimization literature is about one third of a course in optimization theory. So, these are parts of several courses pulled together, in a huge to be able to develop an appreciation, for the underlying mathematical background needed in doing what we do.

(Refer Slide Time: 41:23)

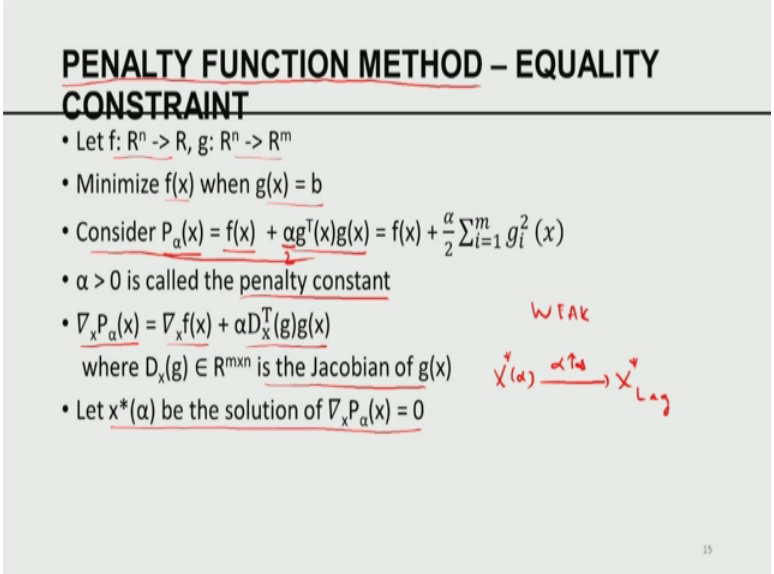


Now, I am going to illustrate by a simple example, let n be 2, let f be given by this that is a non-linear function. You can readily, see I want to be minimize, this I am considering inequality, constraint problem. I am considering a Lagrangian multi multiplier Lagrangian function. I compute the first order necessary condition. So, first order necessary condition gives rise to these two equations, I solve these two equations and I get this optimal solution to be this.

So, that is the optimal solution, I am going to leave the method of solving, this to get, this as a homework problem. You can readily see, I also compute the Hessian and I show the Hessian and an appropriate definition is also positive definite and hence I demonstrate that x 1 is equal to 4 and x 2 is equal to minus half is the constrained minimization problem for this. It is a very typical home work, problem that is, that one has to do why I would like you to emphasize the following.

The function is a quadratic function look at that now, first degree, second degree, the constraint is a linear function. So, objective function is quadratic the constraint is linear, simplest possible case, you have to, have an, here you have to be able to do this, to be able to do anything else in life in this area. So, this is a very nice example that illustrates the power of the Lagrangian multiplier by expanding. So, in here there is only one lambda, because there is m, is 1 the n is 2 m is 1. So, I find both the lambda x 1 x 2. We solve the problem and that is a constraint minimum.

(Refer Slide Time: 43:13)



Now, I am going to talk about an another class of function, which is called penalty functions, which is again I have, I have very well used in, it is very much in vogue, in data assimilation literature, in data assimilation literature, Lagrangian multiplier technique is called strong constraint formulation.

So, let me go back and. So, they do that to tell you that Lagrangian multiplier technique within the context of data assimilation, is called strong constraint problem, because I am, I want to be able to solve the satisfy, the constraint at any and every cost because, the constraint is sacred I cannot afford to, not to satisfy a constraint. So, the constraint is very strong, there is no way out of it, but in some cases I have constraint, but I do not want to enforce the constraint strictly, but you should be vary constraint, you cannot, you can deviate for the constraint, but not too much. So, that is what is called week constraint formulation. This week constraint formulation is done by a class of method called penalty function method.

See people in geosciences, they are very clever people. If they call penalty function they will say oh yeah, it is already existent in operation the literature. So, they give it a different name by whatever name you call, the rose smells the same. So, the same idea comes in different areas by different names, it is very easy to get lost. That is why I am trying to develop this bridge between the terminologies that are used in different disciplines, strong constraint Lagrangian weak constraint. Penalty weak constraint means, what. I want to respect, but not to the (Refer Time: 45:11) verbal verb inverted the law, it is like speed limit. You say somebody says the speed limit is 60 miles. If you go 60.1, are they going to give you a ticket? No, if you go 70, definitely they will give you a ticket. So, when does the police catches somebody? When there is when the speed limit in this highway 60 miles or 60 kilometers well plus or minus 5 percent.

If you go too slow they will come and ask you why are you going too slow? If you go too fast they will say hey I will give you a ticket you are. So, you are allowed to break the rule, but within certain limits. So, speed limit is an example of a weak constraint example of a weak constraint. So, f is a function. So, let us look at this now again f is a function g is another function, I want to be able to minimize f, I want to be able to minimize under equality constraint. So, I am trying to solve equality constraint problems that is where, that is interesting. So, instead of a Lagrangian multiplier of fun Lagrangian function now, I am going to talk about a penalty function what is the penalty function, this is fof x plus alpha times g transpose g x.

So, this is the alpha is called the penalty parameter alpha is a fixed. Large number alpha is a fixed. Large number, because I have a 2, here I think, I must add a 2, here 2 otherwise it is not necessary have it, you need to have a 2, if I have a 2 here, we should have a to belong. So, what is g transpose g x, it is simply g i square sum of g i square alpha is called the penalty constant or a penalty parameter alpha is chosen, alpha is not a free variable, there is only one free variable which is x.

What is the difference between this and lagrange multiplier? In the Lagrangian multiplier lambda as a free variable, x is a free variable. Here, it looks as though alpha plays the role of lambda, it looks as though alpha plays the role of lambda, but the difference is alpha is fixed. It is not a free, but x is free. So, I had two free variables, here one free variable and another para another free parameter, but I have to choose and fix it this is the difference place now that is right.

So, alpha is called a penalty parameter. Generally, penalty parameters are supposed to be large. So, the gradient. So, what is this? You solve this constrained minimization, unconstrained minimization problem. By solving this unconstrained minimization problem p alpha, what is the gradient of p alpha? It is gradient of this plus that. Now, you remember, when we talked to a multivariate calculus h transpose h g transpose g, the gradient of this is Jacobian transpose times g. In order to be able to do that you must have known the knowledge from multivariate calculus that we did in the last lecture. That is why we had tried to make it as complete as possible, where d is the Jacobian. So, x star the optimal solution is obtained by solving the gradient is 0.

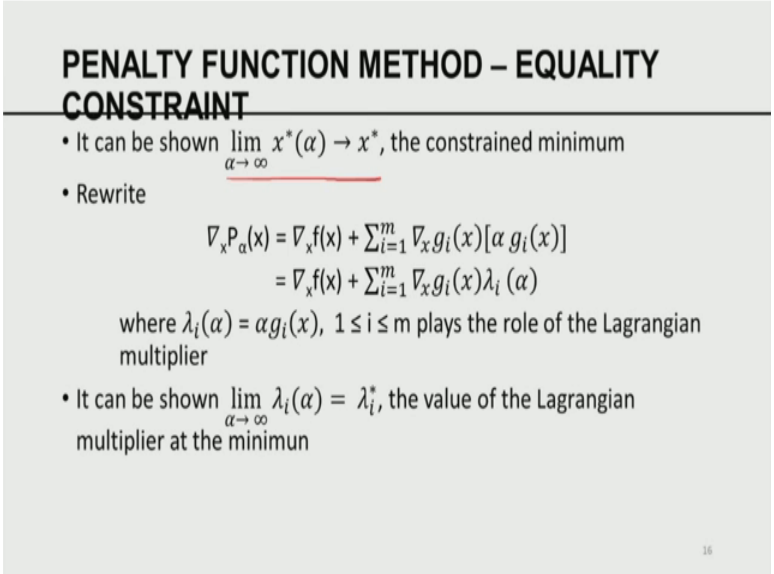
This optimal solution x will always depend on alpha. So, the optimal solution is a function of the penalty parameter. All of you (Refer Time: 48:29) place. So, give it a problem, I can solve the problem in a strong constraint fashion, given a problem I can solve the problem in a weak constraint problem, that things of the following question, how are these solutions are related? It can be shown, I am going to show in a minute, if I took x star alpha and let alpha go to infinity, it reduces to x star alpha of the Lagrangian multiplier technique.

In other words, the weak constrained solution converges to the strong constraint solution, when the penalty parameter grows and bounded alpha goes to infinity you see in other words a the problem is the same problem, you solve it, do different ways, you followed by two different ways. So, which solution do I take, I want to be able to talk about the difference or relation between the two solutions and that is the final relation, the weak solution converges to a strong solution as the penalty parameter goes to infinity. So, what does this tell you? If you believe strong constraint problation is difficult, weak constraint problem you can solve it and you can simply change alpha or plug alpha to a large value then you will know you got a solution, which is as close as possible to a strong solution.

So, what is the alpha in terms of speed limit in some police department? They are very strict if you, they will not allow more than 2 miles, above speed limit in some police department they will allow 5 percent, above the speed limit that is the value of alpha in different departments and why do they do a strict, a speed limit? They want to make money if I give you a ticket; the city gets money for out of you.

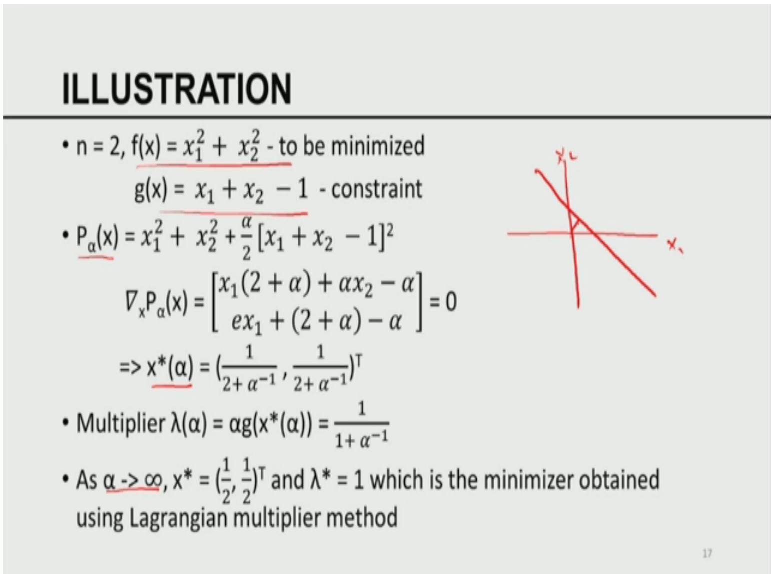
So, what is one cheap way for cities to be able to get good money from public put speed limit, if you put speed limit that will people, who will test it and if you catch them when they test the speed limit, you can find them, you get money, are you with me please. So, alpha is the parameter by which you allow you vary the allowance above speed limit.

(Refer Slide Time: 50:59)



So, that is where the role of the parameter alpha comes to being. So, it can be shown that the solution, the weak solution tends to a strong solution as alpha becomes large I. So, there are other properties I am sure you can follow this, but that js the crux the main result of this page is, this result. The convergence and is the fundamental result.

(Refer Slide Time: 51:24)

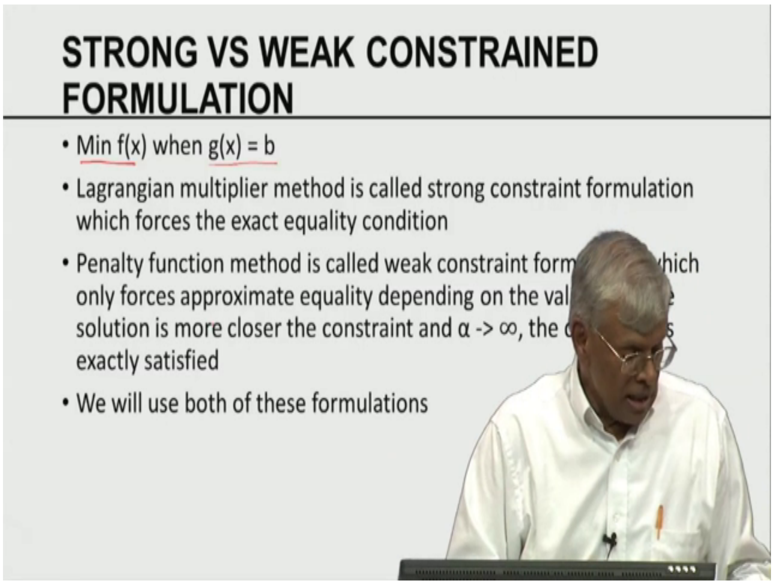


I am going to illustrate the weak solution and a strong solution again by example. So, I got an f of x, I got an f of x, I got a g of x. It is a chronic problem, with the linear constraint, you can easily see this problem. Let us look at this problem. Now, this is x 1, this is x 2, f of x is the parabola. You can see a bowl sitting on that, but I am interested that line. So, I now have to define the parabola over this line, all of you with me please this. So, instead of the parabola being located at the origin now, I have to think of a parabola defined over the line, where will the parabola be a minimum, a little reflection tells you it must be a half; are you with me.

So, the constraint solutions have (Refer Time: 52:19) is unconstraint solution is 0. So, that is what we have. We have shown, we have formulated this problem, as a penalty function problem p alpha alpha by 2. We have computed the optimal solution by penalty, the optimal solution by penalty is given by this.

Now, when alpha goes to 0 1 over alpha, goes to, when alpha goes to infinity 1 over alpha goes to 0 therefore, 1 over 2 plus alpha inverse over 2 plus alpha inverse tends to 1 over half, 1 over half everybody with me please therefore, you will be able to see and if you solve the same problem. We have already solved half, have is the strong solution the weak solution converges, a strong solution as the penalty parameter grows and now.

(Refer Slide Time: 53:17)



It is a beautiful illustration of the relation between the 2, strong versus the constraint formulation again there are two choices, mathematicians have provided us. So, which one? We are going to choose in our analysis that is the question, minimize f of x with respect to g of x is equal to b. So, what is the one standard condition that constraint condition that comes in meteorology. Barotropic is that, must be barotropic, there are constraints geostrophic constraints.

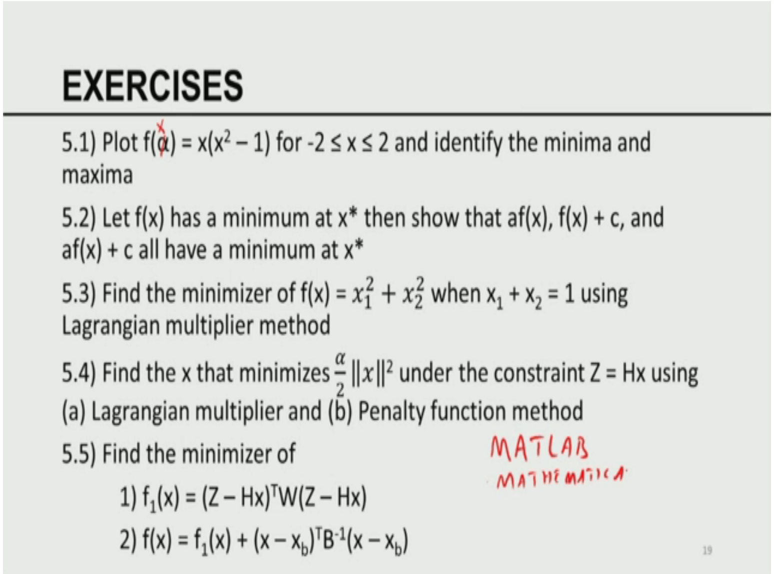
So, then you, when you recover the u v velocity it cannot be anything, it must can satisfy geostrophy. So, geostrophy is a constraint that comes from physics u and v are the velocities that you may recover from other means. There are covered values of the velocity, if they do not can satisfy the geostrophic constrain, the retrieved value of the velocities have no use. So, what is that you would like to be able to say, I recover the velocity with a constraint, the recovered velocity must satisfy geostrophic constraint. Do you want exact geostrophy or you want approximate geostrophy?

If you want exact geostrophy strong constraint, if you want approximate geostrophy weak constrain and in meteorology all the equations are approximation from navier stokes, from primitive equation or it may. So, when you are model, is an approximation. There is no point in requiring something to be stronger, are you with me; please. So, everything has to be consistent therefore, weak constraint formulation is a beautiful formulation, where you allow for variations from the equilibrium or conditions, but only by a very small percentage. So, it fits and. So, this mathematical concept of strong versus weak jives value with our concept of geostrophy, is that atmosphere which is perfectly geostrophic, no.

Is that atmosphere is always benetropic no is always benetronic, no in some cases benetropic in some cases, benetronic there are different situations. So, we would like to make different approximations. So, depending on the nature of the approximations, we can impose different kind of constraints to be able to handle problems that is. So, mathematics provides you the facility to be able to handle different parts of analysis and different kinds of assumptions. You want to make and that is where the fit between physics and mathematics comes into beautiful hue and that is what I would like you to appreciate when we do this.

So, Lagrangian multiplier method is a strong constraint formulation which process exact equality constraint penalty function method is called a weak constraint formulation, which only process approximate, equality depending on the values of alpha, the solution is most closer to the constraint. When alpha goes to infinity, when the, when alpha goes to infinity concerned is exactly satisfied, we will use both the formulations in most of data assimilation problems.

(Refer Slide Time: 56:34)



With this we come to a set of exercise problems. Again I want you to go over these exercises, these are very simple exercise problems, I also would like to recommend that you do these exercises, both the pencil, paper also on a computer. One particular medium of my choice, my favorite choice is MATLAB for example, when you say plot the function, it should be f of x, I am sorry not the f of alpha. This should be f of x not f of alpha.

I would like you to plot this function where the quick way to plot this function MATLAB in two lines, you can do. So, I would like to recommend use of either MATLAB or mathematica, if you have good facility in programming, in either and mathematica or MATLAB, you can do these exercises, pencil, paper and you can verify them by doing it, a computer and that will make your understanding rather complete. So, I have given exercises covering most of the topics that we have covered in this arena.

(Refer Slide Time: 57:46)

Some of the standard textbooks on optimization, these are my favourite textbooks. I have copies of these things in my library, my personal library, luenberger the optimization vector space 1969 is a classic book, luenberger introduction to linear and non-linear programming in 1973 is another classic book on constraint minimization nation. So, for linear and non-linear programming problem, published in 1996 is again another classic.

So, any one of these, all these books are very similar, you can read these books based on the notes to further expand on the proves and deeper understanding and I hope with this you come to realize how? What are the, What is the extent of the mathematical ability, that one needs to have to be able to pursue and do good work in data simulation, this involves conception finite dimension, vector space concepts from matrices concepts from multivariate calculus and concepts from basic optimization theory.

There is another part which I am not included in this discussion is probability theory, when you go to stochastic aspects of estimation, you need to understand reasonably good, you need to have a good background in probability theory and basic statistics.

We will try to fill in some of these things, as we go by, through the lectures. So, with this we will conclude our overview, of mathematical preliminaries for doing data analysis.

Thank you.