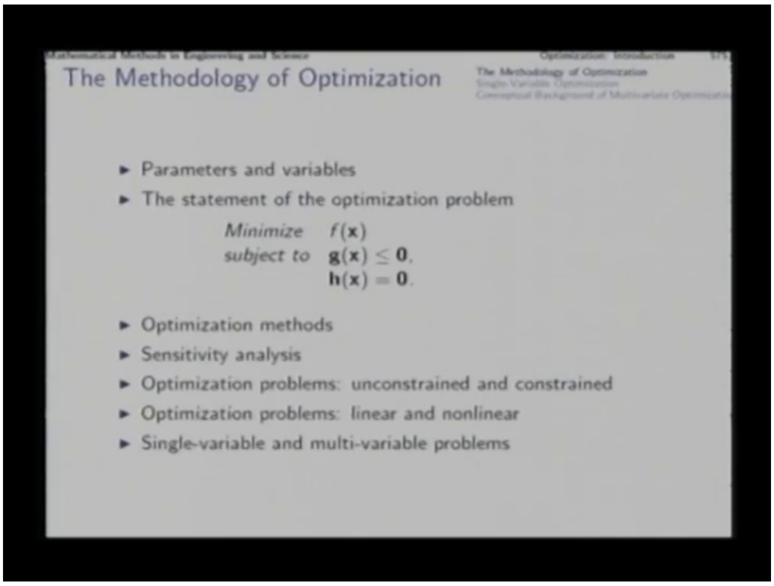
Mathematical Methods in Engineering and Science
Prof. Bhaskar Dasgupta
Department of Mechanical Engineering
Indian Institute of Technology, Kanpur

Module - IV
An Introductory Outline of Optimization Techniques
Lecture - 02
Introduction to Optimization

Welcome. In this lecture we will start our study of non-linear Optimization Techniques. And in this topic we will devote 3 to 4 lectures. And in the current lecture I will first summarize the general methodology of optimization and briefly recapitulate the topic of single variable optimization which you are already conversant with, and continue into developing the conceptual background of Multivariate Optimization. The actual Multivariate Optimization methods will be taken up in the next lesson.

(Refer Slide Time: 01:03)



First, in any typical situation where you encounter an optimization problem: to begin with you will have a number of variables which you can choose in order to minimise or maximize something some function that function which you try to minimise or maximize is called the objective function. And among the variables which you can choose in order to have the minimum or maximum value of the objective function those underlying variables are separated into 2 parts: one of them that is sum of the variables you can see it as parameters which are kept constant for one particular study. In a particular problem you may choose to keep all variables as designed variables in hand and process all of them together.

On the other hand some of the variables at some situations are kept constant as parameters. After fixing the values of the parameters, the remaining variables which in a particular study you want to explore in, to get the best possible value for the objective function those only are treated as the variables of the optimization problem. And a typical statement of an optimization problem goes like this in which you say you want to minimise a function of x subject to certain constraints.

So, there may be an optimization problem in which there are no constraints or there may be one in which there are constraints. Constraints are again of 2 kinds: one is in inequality constraint, the other is equality constraint. Now after formulating the problem now this formulation part comes from the domain in which you are going to apply the optimization methods. Now after studying the optimization methods, you would notice that in almost every branch of science and engineering and even humanities quite often you come across situations where many problems can be solved through an optimization formulation.

The problem may be one of explicit optimization where you actually want to minimise or maximize something or many times it happens that the actual problem is something else, but you can reformulate it in the form of an optimization problem. For example, in the last lecture we formulated an equation solving problem in the form of an optimization problem.

So, there are many such problems which can be formulated in the form of an optimization problem and then optimization methods can be used in those problems with advantage. Now after the formulation is made from any given field then you look for a suitable optimization method or algorithm to find a solution of this problem. Now any point x any variable value any set of values for the variables x that satisfies these constraints g x is less than equal to 0 and hx equal to 0, the given constraints for any such point is called a feasible solution that is it is allowable it is permissible by the constraints in both on the problem definition itself.

And out of those feasible solutions you try to find the one in which the function value is minimum, if it is a minimization problem. The optimization problems can be of minimization or maximization type, but in most of the theory that we discussed you will find that most of the time we are talking about minimization.

Fixing our attention to minimization problem helps to keep the entire theory in one standard form. If a problem is of maximization then we can always try to minimise the negative of the objective function. So, this is typically done in order to avoid the hassle in the notation sectors.

Now after you apply some optimization method a good number of them we will be studying in this course. So, after you apply that method you get a solution which is the solution of this optimization problem that means, it minimises the function objective function and satisfies all these constraints. Now after getting that solution in hand quite often you want to find out whether the parameters that is that subset of variables which you kept fixed whether it was a nice idea whether it was wise idea to keep those values fixed.

So, in that case after you have got the solution in hand then you conduct a sensitivity analysis you try to find out that how sensitive is this solution that we have got in hand how sensitive it is to the values of those parameters if they are found to be very sensitive then you try to see whether those parameters can be also changed in order to get a much better solution.

On the other hand if you find that the solution is quite insensitive to the parameters that you have fixed as parameters, then you say that fixing that is a vice idea unnecessarily we need not conduct the optimization process with too many variables. Now optimization problems as I told you just now can be unconstrained without these constraints in which case the entire space of x is feasible or they may be constrained. And in that way we classify optimization problems and for that correct optimization methods into unconstrained optimization and constrained optimization. Obviously, unconstrained optimization problem is easier to solve compared to constrained ones.

Then you also classify optimization problems as linear and non-linear problems, if both the objective function and the constraint functions g x and h x are all linear functions then you call the optimization problem a linear optimization problem or linear programming problem l p problem. On the other hand if either the objective function or any of the constraint functions is non-linear, then it is a non-linear optimization problem or non-linear programming problem n l p problem.

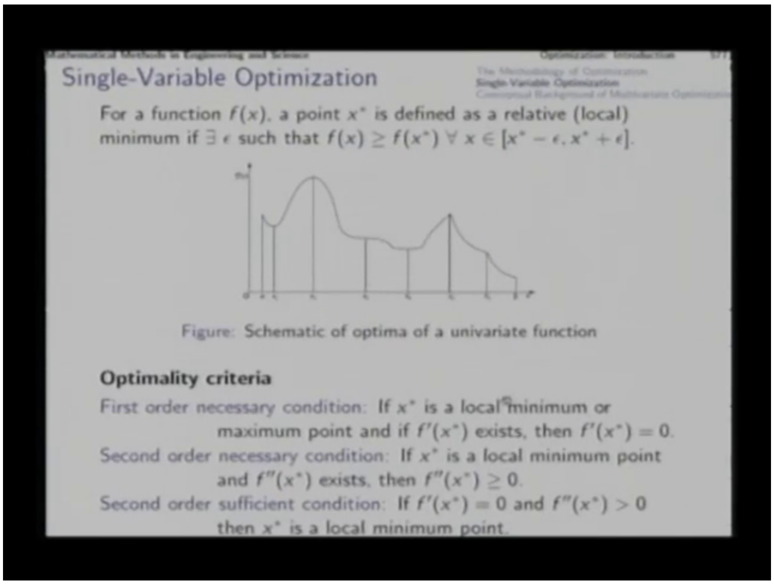
Then you could also classify the optimization problem as single variable and multivariable problems. Single variable problem has a single variable and in multivariate problems you have several variables in hand with which to play around to minimise the objective functioning.

Now you will notice that when you classify optimization problems as unconstrained and constrained and in another way we classify them as linear and non-linear in total we do not get 4 kinds of problems that is any of the 2 above and any of the 2 below we cannot combine. Because in the case of a linear optimization problem you cannot have it unconstrained, because linear functions go on reducing in a certain direction in certain directions.

So, if there is no constraint on the variables in x then they are unbounded on the lower side as well as on the upper side. So, linear unconstrained optimization problem does not exist. So, if it is a linear programming problem then constraints will be there in any case. So, you get linear programming problem which are linear constrained problems or unconstrained non-linear problems and constrained non-linear problems which is the most difficult.

Now before, going to the methods for multivariate optimization which is going to be our main focus.

(Refer Slide Time: 09:11)



Let us quickly recapitulate the ideas of single variable optimization which with you are conversant already to a good extent, say for a function f of x a single variable a point x star is defined as a local minimum point. If there is some epsilon such that the value of f x at all other points that is all points in that neighbourhood in this epsilon neighbourhood that is epsilon on this side and epsilon on that side in at all other points the function value is greater than or equal to the function value at the current point x star then you will define the point x star as a local minimum point.

Now schematically let us have a look at this there is a function which is defined over the interval a to b in this particular case x 1 is a local minimum point and then x 2 is a local maximum point, x 3 is neither a local minimum point nor a local max maximum point. Because there is some point close to x 3 on the right side at which the function value is likely to be lowered and this will not be satisfied so this is not a local minimum point.

It is neither a local maximum point because on the left side of it the function value is likely to be higher, so this is a point of intersection. Now here you find it is again a local minimum point. So see the difference between this point, this point and this point here it is a clear local minimum point here it is not here it is here the function profile comes from upward and then becomes constant for a while and moves up again.

So, this is a minimum point here the function profile comes downward becomes constant for quite long and then goes down, but then whenever we say that it becomes constant for quite long it is not necessary that over an interval around x 3 it is constant it is just that it touches it is tangent more smoothly that is possibly first order and second order derivative both are 0. Here you find that a constancy kind of situation lingers for a little more smoother interval around it and then it goes up so it is a minimum point you will notice that even b is a minimum point. Now here the curve does not become flat, but since beyond b the function is not defined. So, this is also a local minimum point the function is defined only on the left side and all points on the left side are above the current point so b is also a minimum point.

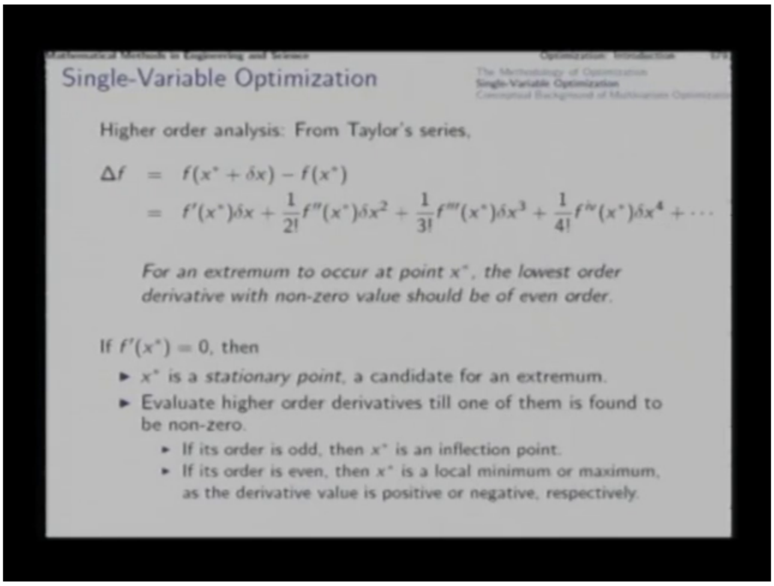
So, in this schematic x 1, x 4 and b are 3 minima. On the other hand a x 2 and x 5 are maximum x 6 is neither a minimum nor a maximum. Similarly x 3 is neither a minimum nor a maximum now that is according to this definition. And those points where the function is differentiable there you can find out certain optimality criteria based on the derivative and that is the first order necessary condition says if x star is a local minimum or maximum point and if the first derivative exists then it must be 0.

If the first derivative is nonzero at that points, then if it is positive then on the right side it will be it will go up the function value will go up on the left side the function will go down. Similarly if the derivative is negative then reverse will happen one side will be higher other side will be lower. So, that way it can be the point can be neither a minimum point nor a maximum points. Therefore for being local minimum or maximum point the first derivative can neither be positive nor negative and therefore, it must be 0.

Now, note in this case what is happening the first derivative is 0 the tangent is horizontal. What is the tangent before and what is the tangent after that is, what is the slope before that point and after that point. Before that point the slope is negative going downwards after that point the slope is positive at that point the slope is 0.

So, you find the first derivative is 0 and first derivative as x changes, as x increases the first derivative is first negative then 0 then positive that means, the first derivative is an increasing function of x which means that the second derivative is positive. So, that gives you the second order condition. Second order necessary condition is that the second order derivative is non-negative positive or 0 that is why this is also a minimum point this is also a minimum point. But this is not so it is necessary but not sufficient. Second order sufficient condition will be that the second derivative is positive. Now if the second derivative is 0, then it satisfies the necessary condition but not the sufficient condition. So, to resolve the situation you need to go further the way to go further is through Taylor series.

(Refer Slide Time: 14:50)



So, if you write the Taylor series of f of x around a given point x star the candidate point, then you get f of x star plus delta x is equal to f of x star plus first order term plus second order term plus third order term and so on. Now keeping f x star that is transposing f x star on the other side you talk of the change in function value.

So, the change in function value from x star to x star plus delta x at that neighbouring point the change in the function value from Taylor series is given like this. Now here you find that as long as the first derivative at that point is not 0 for small enough interval for small enough delta x this first term will dominate these things and the sign of this first order difference will depend on the sign of delta x that is whether you are taking the other point on the positive side or on the negative side, whether delta x is positive or negative. Depending upon that the first order term will change it is sign and that term is going to dominate the series for sufficiently close points for sufficiently small delta x.

That will mean that on 1 side this will be positive and the other side it will be negative. And therefore, the difference being positive on 1 side and negative on 1 side will include the possibility of the current point being a minimum or maximum point. Therefore for the current point to be a minimum or maximum point it is necessary that this derivative vanishes, the first ordinary necessary condition as we saw just now. If this point vanishes then for sufficiently small delta x values this entire series will be dominated by the second order term, and the sign of the second order term does not depend on the sign of delta x because delta x is appearing as a square.

So, irrespective of whether you go this way or that way delta x square is positive it will depend upon that derivative sign, if the derivative is positive then this will be positive for sufficiently closed points. Now if it is positive that means, the neighbouring point both sides have higher function value right, so that will qualify the current point as a local minimum point right.

Similarly, for negative values of value of this it will be a local maximum point, and that is why the second order derivatives being positive with the first order derivative 0 is sufficient condition for the current point to be a minimum point. If it is 0 if the second order derivative is also 0, then again the series will be dominated by the third order term the sign of which will again depend on delta x because you see delta x cube appears or power, so this goes on.

So therefore, now looking at the pattern you can say that for an extremum to occur at point x star the lowest order derivative with nonzero value should be of given order. If up to 3rd order it is 0, 4th order is positive, then again it is a local minimum point. So, that gives you a working rule for determining candidate points, and then classifying them as minimum maximum and so on. So first of all you evaluate the first derivative and set that equal to 0, and solving that you try to find out candidate points x star, so such a candidate point is to begin with a stationary point, it may be a minimum point or a maximum point or it can be a saddle point an inflection point.

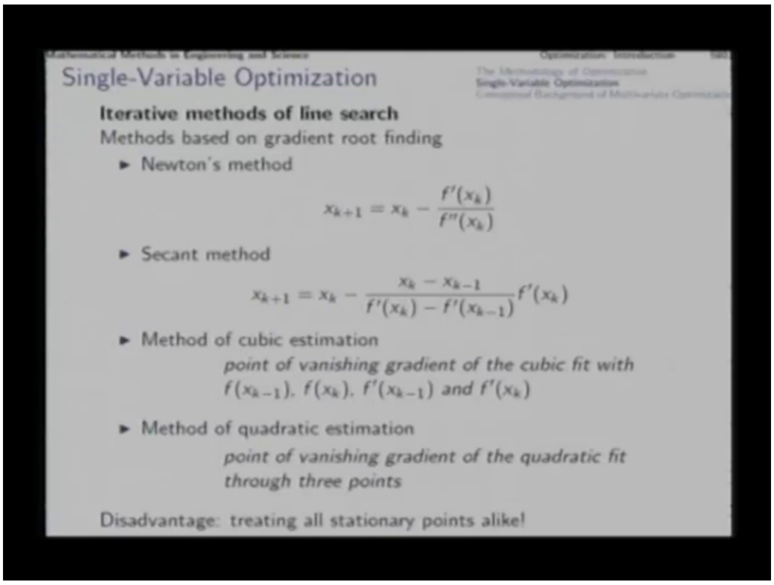
So, after we have captured certain candidates for further tests, then at that point at those point we evaluate higher order derivatives till 1 of them is found to be nonzero. If we go on finding derivatives and several of them are found to be 0 2nd order, 3rd order, 4th order, 5th order, then we stop at that point where first non 0 derivative is encountered, if it is order is odd then the current point x star is an inflection point coming like this going like this, or coming up and then going further up like this.

If the order of that first non 0 derivative is even that is either the 2nd or the 4th or the 6th and so on, then that will be a local minimum point or a local maximum point depending upon whether that derivative is positive or negative. So, this much you have studied long back in (Refer Time: 19:46) standard calculus itself, and this was the working rule for finding maxima and minima at that stage; however, it requires the solution of an equation.

Now, solution of an equation is always not an easy task in the previous lecture, we also discussed the situation where for solving an equation we formulate it as an optimization problem. So, equation solving is also not always extremely easy, there may be equations which are very difficult to solve. So quite often we do not try to rely on the equation solving process to capture the candidate points, but we follow an optimization based algorithm directly to find the minimum point.

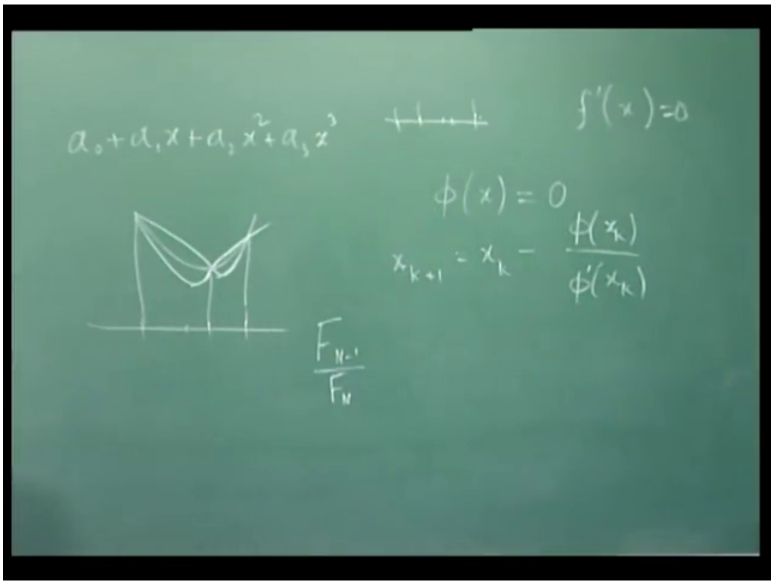
Now for that there are several methods now there are some of these methods depend on gradient 1 way or the other and some other do not depend. That is some of them depend derivatives, some of them do not depend even those which depend on derivatives some of them use derivatives explicitly and others do not.

(Refer Slide Time: 20:58)



For example Newton's method which is reminiscent of the Newton Raphson method of equation solving.

(Refer Slide Time: 21:12)



So, in the case of equation solving we had, when we required the solution of an equation like this, then our typical iteration was x k plus 1 is equal to x k minus phi of x k divided by phi prime of x k right.

Now, here are talking about minimizing the function f, now we know already that at the minimum value at the minimum point f prime is going to zero. So, why not try to solve f prime x equal to 0. So if we want to solve this equation x prime x equal to 0 then in case of phi, if we put f prime, then we get f prime here f double prime here that is second derivative. That is the typical Newton's method for optimization of a single variable problem single variable function.

1 difficulty of this is that this formulation will not differentiate between a minimum point and a maximum point, or an inflection point with 0 derivatives 0 first derivatives. Now here itself in the case of second derivative if we replace that with a finite difference kind of derivative formula, then we get this formula which is the second method. Now in this second method you will notice that we do not need the second derivative, but we need the first derivative and the function value at two points. So, Newton's method work with a single point up to second derivative which also means that the secondary derivative should exist, second method works with two points at a time; that means, in the initial guess we need to give it two points, and only up to first order derivative. Method of cubic estimation is another method which uses function values and derivative values at

two points. Starting with two points it evaluates the function value and the derivative value at these points and then; that means, that we have got 4 total number of four conditions four conditions in total, two conditions at this point and two conditions at that point function value derivative value, and with this kind of four conditions we can fit a cubic function in the local neighbourhood ok.

That is for the local neighbourhood we can approximate the actual function by a function of this kind a cubic feet. So, as we impose the conditions that is at x 0 and x 1 we are prescribing the function value, and the derivative value as we prescribe these conditions on this, we essentially get four conditions four equations for linear equations for that matter in the coefficients a 0 a 1 a 2 a 3 from that we can determine these coefficients.

And then we say that we look for that point where this cubic function is minimised or it is derivative is zero. So, that we will get in terms of a 1 a 2 a 3 etcetera and that becomes another point, now out of the two original points and this third point we retain two points and drop 1 of the old points, again at two points we evaluate the derivative and some derivatives and functions values are already there at a new point we evaluate the function and derivative and continue

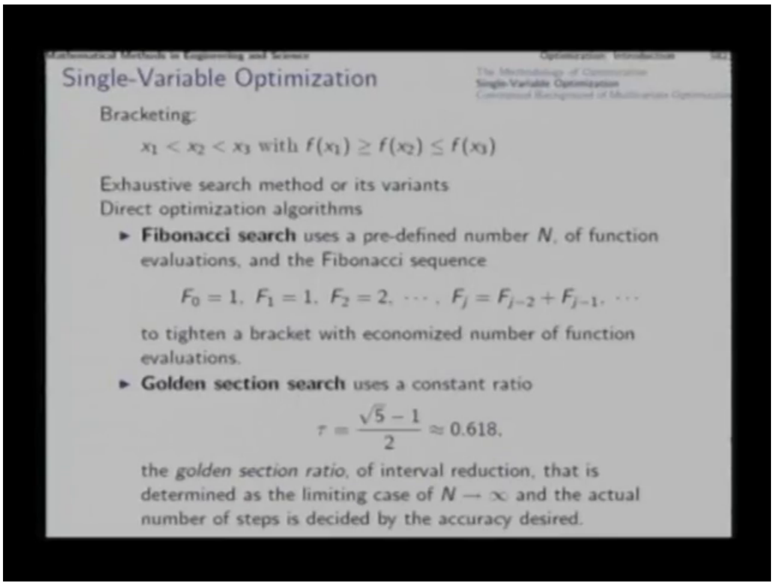
So, this is this method is called the method of cubic estimation. Similarly there is a also method of quadratic estimation, that operates not with derivative at all that operates only with function values, but at 3 points. So to being with you need to prescribe three points to this method and through those three points only with the help of function values the algorithm frames a quadratic, a 0 plus a 1 x plus a 2 x oh this is square a 2 x square only up to this and by fitting that quadratic with three coefficients because of with the help of three function values, and then asking for it is gradient to vanish it gets the new point.

So, this method of is quadratic explanation, so in which only function values are used no derivatives not note that whether some of these methods used derivatives here, up to seven derivatives here, and here up to first derivative, here no derivative, but then still all these four methods in indirect manner refers to the vanishing of derivative, because that is the test that is the requirement based on which the new point is generated.

So, the these are disadvantage of all these methods is that it treats all stationary points alive, and does not differentiate between a minimum and maximum that is a disadvantage in these methods. So, if the problem is such that it has a minimum and perhaps not a maximum, then any of these methods will work out nicely on the other hand for a problem which has lots of minima and maxima this kind of a method runs the risk of reaching a maximum points.

There are some other methods which first insist on a bracket and second do not make any difference not even an indirect one to the derivative. First what is this bracketing in the case of equation solving or root finding problem, we refer to the continuity of a function and said that if there are two points x 0 and x 1 and the sign of f x 0, and sign of f x 1 are different 1 is positive the other is negative; that means, due to continuity it is necessary that at one point in between x 0 and x 1 the function is bound to cross the 0 line and that is the root, so that was the way bracket the solution of an equation. In the case of minimisation problem the bracketing has a slightly different meaning.

(Refer Slide Time: 27:39)



If there are three points x 1 x 2 x 3 x 1 less than x 2 which is less than x 3, such that at x 2 the function value is lower than both x 1 and x 3 then we say that between x 1 and x 3 there is a solution ok.

So, if we have a pattern of the function which is like this, then we can say that in between these there must be a minimum point, because it is known that from this point the function value has gone down, and then it is known that from this point the function has gone up. So, in between what is the point here or somewhere where that actual trend is made, now these three could be like this in this case a minimum is here or it could be like in this case a minimum is here so, this is important.

So, bracketing in the case of minimization problem requires three points and the pattern or trend of downward and then upward should be established to identify a bracket. Now once such a bracket is there some of the optimization methods some of the single variable optimization methods try to continuously squeeze the bracket, bisection is 1 possible way for example, if we know that in between these there is a minimum point, then in a similar manner of bisection we can try to see that whether this half is going to constitute a bracket or this half is going to constitute a bracket, that is which half of the complete interval is going to retain the nature of a bracket. And then like that we can squeeze and find the solution; however, compared to bisection two other methods are found to be more efficient that is they conduct a same job with the same accuracy with less number of function evaluations.

One of them is Fibonacci search in which the interval reduction is not true half at every iteration but in a variable size in a variable fraction the interval is reduced the squeezing takes place at a variable rate, and the subintervals are decided based on Fibonacci numbers. So, f n minus 1 by fn in this way you try to reduce the size and through this measure what you ensure is that for this interval if you evaluate one point here, at evaluate the function at one point here, and at another point here at same distance from the two endpoints of the interval, and then either you retain this point and throw all of these or you retain this and throw away these.

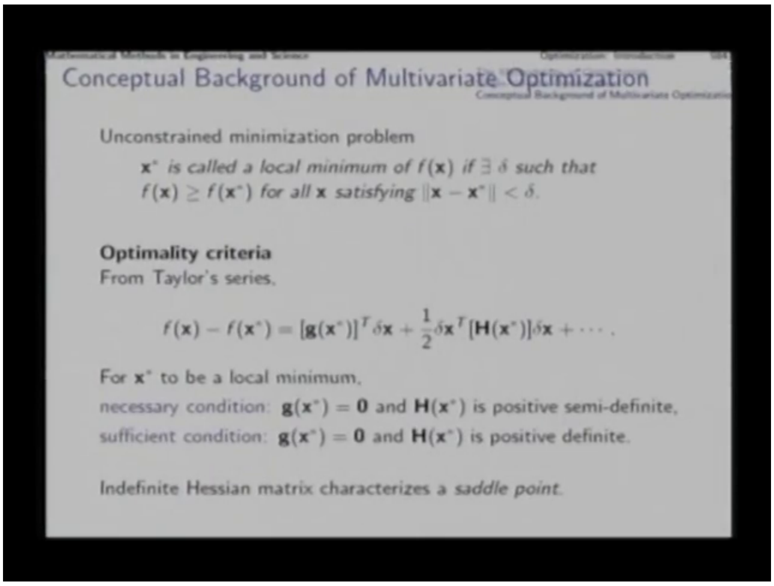
So, out of these whichever is larger than whichever maintains the bracket that is retained and the other is removed from here, and that way what happens is that in the next the way the fractions are generated with the help of Fibonacci number it becomes obvious that the, in the next round in this interval the two points that will be needed where the function will be evaluated out of that 1 will be this (Refer Time: 31:24) and the other will be symmetrically placed here, so at every new iteration the two new points the two new internal points integer points that will be needed one of them is one of the old points.

So at every in iteration only one new function evaluation is made and every function evaluation is used twice on an average, now from the Fibonacci search method itself one particular another search is developed which golden section search, in that the interval reduction fraction is not variable, but it is constant and it is equal to this golden section. When golden search a similar operation is done but at every iteration the interval reduces by this fraction which is the golden section ratio.

Now though this squeezing of the bracket; though this interval reduction, at every step there will be a stage where the interval is so small, that is smaller than your requirement of accuracy for example, if you wanted the solution up to an accuracy of 0.01 then by the time the size of the interval itself is less than 0.01 you say that any of these points is good enough as a solution. So, that is the way Fibonacci search and the golden section search method follow operate, and they keep on squeezing the bracket and finally make the bracket so small, that any point in that bracket is good enough for the required accuracy.

Now, with this much background of single variable optimization recapitulated now, we will go to discuss the actual problem of our focus which is multivariate optimization. First unconstrained optimization in this lecture and the next, and then will study a little constrained optimization.

(Refer Slide Time: 33:36)



Now, in an unconstrained minimization problem, a point x star is called a local minimum of the function, if there exists a delta whatever small you would like to choose. If there exists some delta such that within a ball cantered at the current x star and radius delta all points have the function value which is greater than or equal to the current point under question, then the current point x star is called a local minimum point. Now this is the basic definition of a local minimum point, and note that we are talking about local minimum point and most of the algorithms which we will be discussing scattered to the problem of finding a local minimum point only. Now you can talk of finding all the local minima several local minima, and then out of them choose the smallest one and hope that that is the global minimum that is one option.

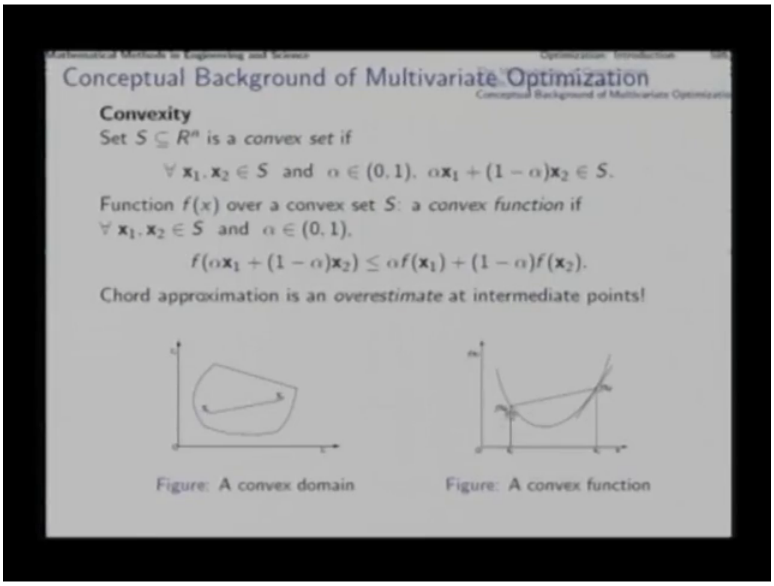
If you want if your problem demands to find the global minimum, now if the function is differentiable then you can work out some optimality criteria as we did in the case of first in the case of single variable problems based on derivative, say again making an approach to the Taylor series we find that if x is a point neighbouring x star, then the difference or function values f x minus f x star will be the first order change, where which is gradient transpose delta x where x minus x star is delta x plus half delta x transpose hessian the secondary derivative matrix into delta x plus the higher order terms.

So, up to this is the truncated second order truncated Taylor series, now for extra to be a local minimum again you argue in the same manner that as long as this gradient is nonzero, there will be some directions delta x along, which the function will increase and some directions along, which the function will decrease. Now a direction along which the function increases will ensure that in the opposite direction the function will decrease. So the current point cannot be minimum or cannot be maximum.

So, for minima or maxima for any extremum the first order term must vanish, which means the gradient as a vector the complete gradient vector must vanish all the partial derivatives should vanish. And then this second order term dominates the series for sufficiently small delta x, and in that case the sufficient condition is the positive definiteness of this hessian matrix at that point, which will ensure that for all delta x it is positive that is sufficient condition.

Necessary will be that it is positive semi definite and indefinite sufficient matrix with some eigenvalues positive and some eigenvalues negative will characterise what is called a saddle point. Note that we can talk of the first order condition or second order condition only when, the function is first order differentiable and second order differentiable and so on. So, only for first order functions which are differentiable up to that first order we can take of you can talk of first order condition, and only for functions which are differentiable twice you can talk of the second order condition. Now with these optimality criteria we will proceed towards a few further issues which will be found quite useful when later we consider multivariate optimization methods, and the most important issue in that direction is convexity.

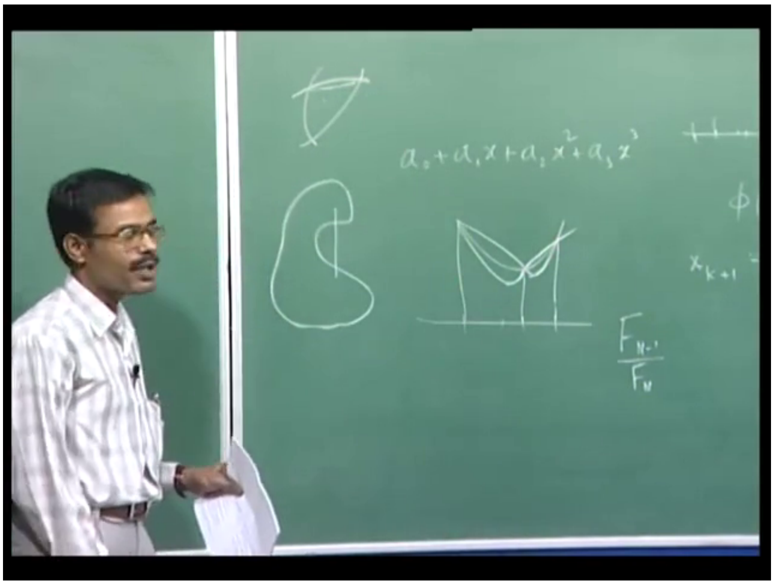
(Refer Slide Time: 37:48)



So, there are two aspects of convexity, a convex set, or a convex domain and a convex function.

Now, in the r n that is n dimensional real space a set S or a region is a region is called a convex set, if for all pair of points belonging to that set the complete state line segment joining them is also inside the in the set that is this region is not convex.

(Refer Slide Time: 38:10)



Because in this you can find two points which are which belong to the set, but the state line segment joining them does not completely lie within the set. Now on the other hand this region is convex because for every two points inside the set the straight line segment joining them will be completely inside the set inside the region, now for unconstrained optimization problem the region question will not arise, but it will arise in constrained optimization problems, but then it will be important because further we will defining a convex function, which can be sensibly defined only in a convex set or convex domain.

Now saying that the straight line segment joining the two points is saying this that is for a alpha belonging 0 2 1 interval alpha x 1 plus 1 minus alpha x 2 that is for alpha equal to 0 you get x 2 for alpha equal to 1 you get x 1 and for any intermediate value you get a point in the line segment joining x 1 and x 2.

So, such a set for which in which for every two points this will hold that the entire straight line segment joining the two points also, we can will be will belong to the set will belong to the region such a region is called a convex region or a convex domain. Now over a convex set over a convex domain, you can define a function F x which will be a convex function if for every two points belonging to that region and alpha again between 0 and 1 the function value at an internal point in that line segment is less than equal to the corresponding linear interpolation between the function values at the end point. Now if this is not very clear then think of it this way, that is you have the function value at x 1 that is f x 1 and you have the function value at x 2 that is f x 2, and you want the function value at a point which is intermediate between x 1 and x 2 say at 0.2 fraction of the distance from x 1 to x 2 that is from x 1 to x 2 in that line segment 0.2 distance from x 1 and point eight distance from x 2.

Now 0.8 into x 1 plus 0.2 into x 2 is that point you evaluate the function at that point, and that is the function value here, now rather than evaluating the function value at that point if you had tried to interpolate it from the function values at that two endpoints then you would get this approximation right, 0.8 into f x 1 plus 0.2 into f x 2.

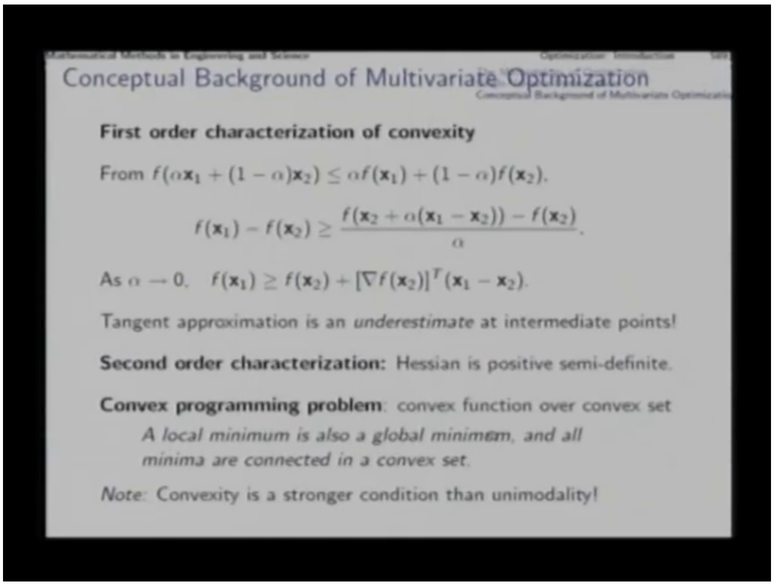
Now, the function is called a convex function if in every such situation, this interpolated function value is always an overestimate, that is chord approximation interpolated approximation if it is always an overestimate compared to the actual function value at all intermediate points, then it is called convex function equality is permissible that is it can be equal.

So, schematically seeing this is a convex domain because any two points that you can take in this will ensure that the straight line join of those two points is completely inside, now this is an example this graph of the function that is shown is a convex function because you see that if you take two points x 1 and x 2 and the function values here, now a linear interpolation between them a chord approximation will be given like this.

So, at this point the function value through a chord approximation will be found to be this whereas, the actual function value is here, actual function value is lower and the chord approximation is higher chord approximation is an overestimate. So this kind of a function is called a convex function that is for being a convex function such a thing must happen at every intermediate every pair of points for every intermediate value ok.

Now, you will see that other than the chord approximation if you wanted to make a tangent based approximation, that is you know the function value here and you want to you know the derivative also at this point and based on the function value, and the derivative through a first order Taylor series you want to approximate the function values somewhere else that is the tangent approximation. And the tangent approximation will be always an underestimate it will be lower see the tangent is going lower, on this side as well as on that side compared to the actual curve actual graph the tangent approximation is lower on left side as well as on the right side. So such is the property of a convex function and the chord approximation is an overestimate that actually in a way implies, that the tangent approximation will be an underestimate.

(Refer Slide Time: 43:50)

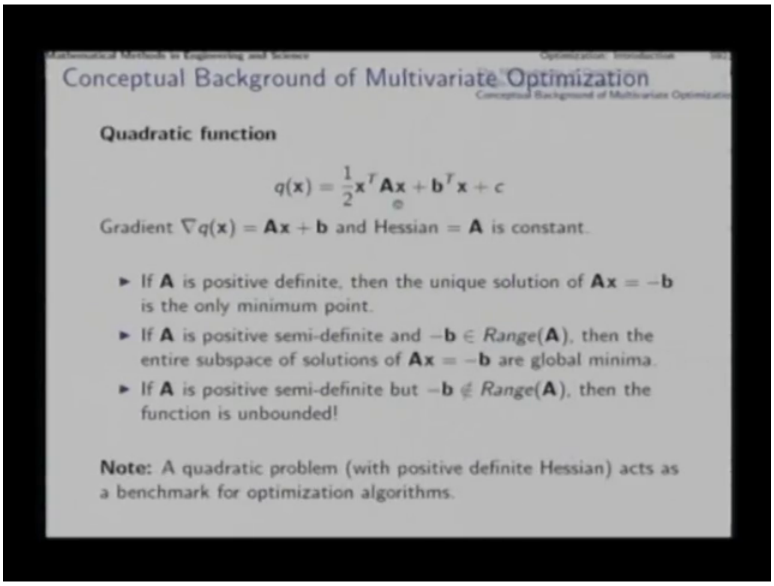


And that you can show through a few small steps which we are we will not going to the detail currently, the only thing that we need to stress at this point is that this gives you a first order characterization of convexity. This is the 0 third order characterization of convexity which is the definition in terms of only function values and this is equivalent to the first order characterization of convexity, which you can talk of if the function is first order differentiable, that is f x 1 is greater than equal to f x 2 plus gradient transpose x 1 minus x two.

You can work out a second order characterization also, through another few small steps and that is actually quite straight forward. The second order characterization of convexity is that the hessian matrix the second order derivative matrix is positive semi definite, it is the function is strictly convex if it is positive definite on the other hand if it is possible semi definite, then it is just convex there is a certain class of problems in which the region the domain the feasible domain is convex and the function that we want to minimise is also convex, such a problem is called a convex programming problem.

That is we try to minimise a convex function over a convex set, and in that kind of a situation a local minimum is also a global minimum and all minima are connected together in a convex set. So convexity is a very strong condition on a function further we know nicely behaved function we will find in a quadratic function, which could be convex or could be non convex also, but a convex quadratic function turns out to be a benchmark problem.

(Refer Slide Time: 45:58)



Against which all optimization algorithms are qualified. A quadratic function is like this and if you try to find out it is gradient and it is hessian, then you will find very easily through first order and second order derivatives, and the variant is this A x plus b and the hessian A is constant quadratic function, so second order derivative should be constant and hessian is this matrix A the second order derivative.

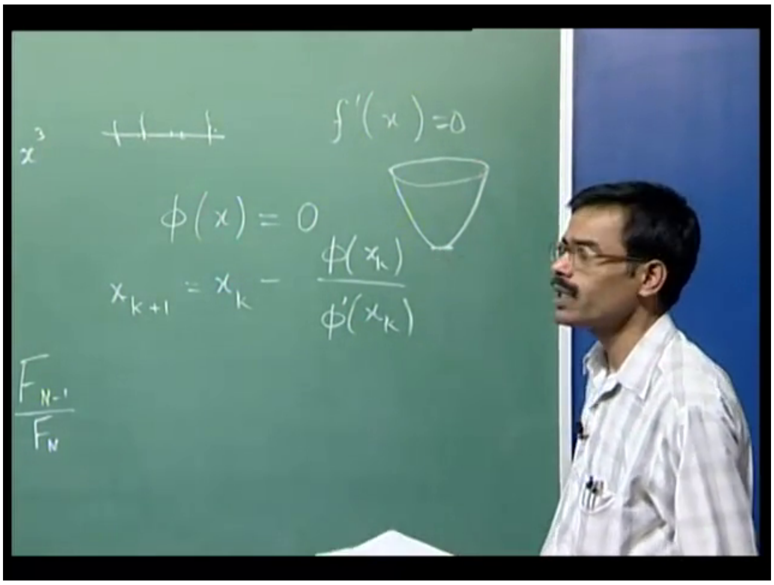
Now what kind of a matrix is this hessian A, if it is positive definite then you will say that it is a convex quadratic function, and quite often when we use a quadratic function as a benchmark problem, then we consider convex quadratic function in which the hessian a is positive definite. Now if A is positive definite then it is non singular as well which means that this equal to 0 will have a unique solution A x equal to minus b will have a unique solution, and that unique solution will satisfy the first order condition gradient is 0, and the hessian is positive definite anyway that together will satisfy the sufficient condition for that particular point to be a local minimum, and that is the that since 0 gradient has that as the unique solution.

So, that unique solution is the only minimum point of the function, if A is positive semi definite, then it is singular as well and in the case of singularity, and this system of equations this equal to 0 this may be constant or may not be ok.

so if it is consistent that is minus b is in the range of A, then in the case of singularity of A once it is semi definiteness you will have infinite solution and all those points, all those solutions, all those infinite solutions of this are local minima, and global minima as well and they are together connected that is they are distributed over an entire line or entire plane like that.

So, that is again a convex set if A is a positive semi definite, but it is this system of equation this equal to 0 is inconsistent that is f minus b is not in the range of A that will mean that the convexity is not a problem, but 0 gradient condition is not satisfied anywhere. So, in that case the function is unbounded, and A minimization problem has no solution, these second and third cases the first case is very simple the first case is like this ok.

(Refer Slide Time: 48:47)



This is the shape of the function. So, this is the minimum point 0 gradient 0 derivative condition is satisfied only here, and the function is convex everywhere so second derivative matrix is positive definite everywhere it is constant. So, this is the unit minimum.

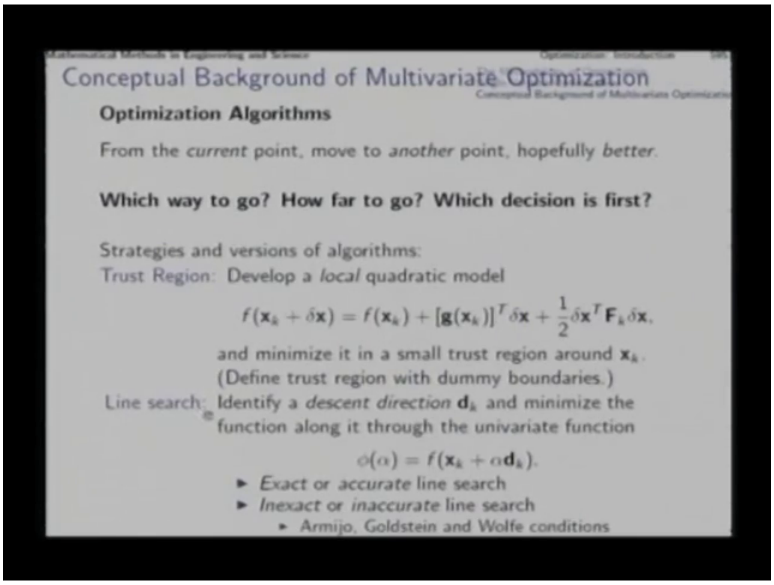
On the other hand the second case here, A is positive definite positive semi definite that is singular and minus b is in the range of A; that means, that there are points where the gradient vanishes, that is a function profile which is like this cylinder. In this cylinder you find that it is convex semi definite, because you see along these directions it is straight and along this direction it is convex like this.

So, this is a semi definite case and when you try to solve a x plus b equal to 0 zero gradient 0 slope slow, then all these points satisfy the 0 gradient condition and therefore, this entire line is the solution of the minimization problem, all these are the points at which the level of this function is lowest. On the other hand the third case here, this 1 is a same cylinder, but not placed horizontally, but like this. In this case again it is semi definite because there are directions in which there is no convexity concavity and there are directions in which there is con convexity.

So, here 0 gradient condition this A x plus b equal to 0 that condition is not met that it is there is no point which satisfies the 0 gradient condition, as you can see on this surface there is no point at which gradient is zero, so that is why the function is unbounded that is along this direction you can go on going downward and there is no end to it, compared to this case, where you cannot go downward.

So 0 gradient condition is not met in this kind of a situation at any point, so that is why in this case there is no solution now as benchmark problem we typically consider those quadratic functions for which the hessian is positive definite non singular (Refer Time: 51:08).

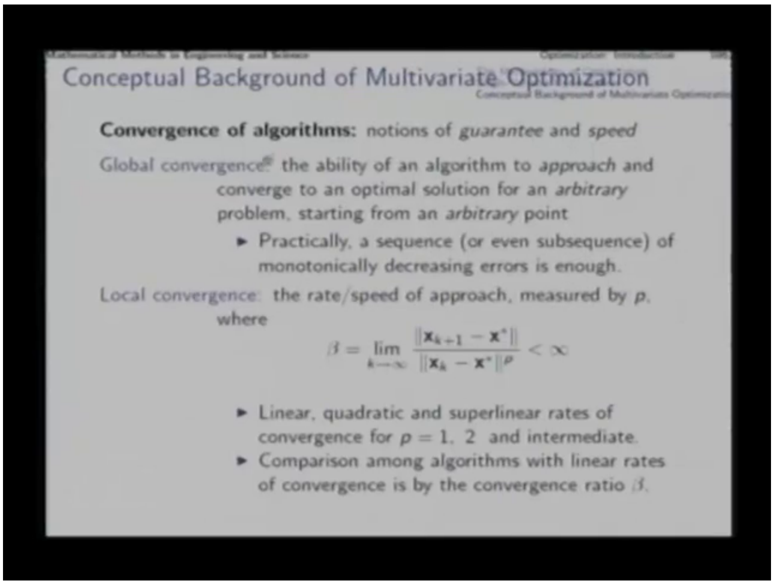
(Refer Slide Time: 51:11)



Now, for an optimization algorithm we need to have a good picture, a clear picture of how a typical optimization algorithm operates. Typical way to operate for an optimization algorithm is to start from a current point move to another point which is hopefully better than the first. Now there are three questions that arise in this process first is which way to go, second is how far to go, and which decision is taken first. If we first decide the direction which way to go and then decide how far to go in that direction this gives us 1 strategy of automation algorithm that is called the line strategy.

On the other hand if we first decide that within this much distance we are ready to go and then, we decide that within this much distance in all directions which direction to take, and how far to go that is how far we are ready to go, if we take that decision first and this decision we take later then that is a strategy which is called trust region strategy. Now there are some algorithms which can be implemented in both the strategies some of them can be implemented in only 1 of the strategies. For any optimization algorithm there are two questions that arise one is the question of global convergence, that is whether at every step the algorithm makes an improvement in the function value whether it approaches a optimum point and that is the issue of global convergence.

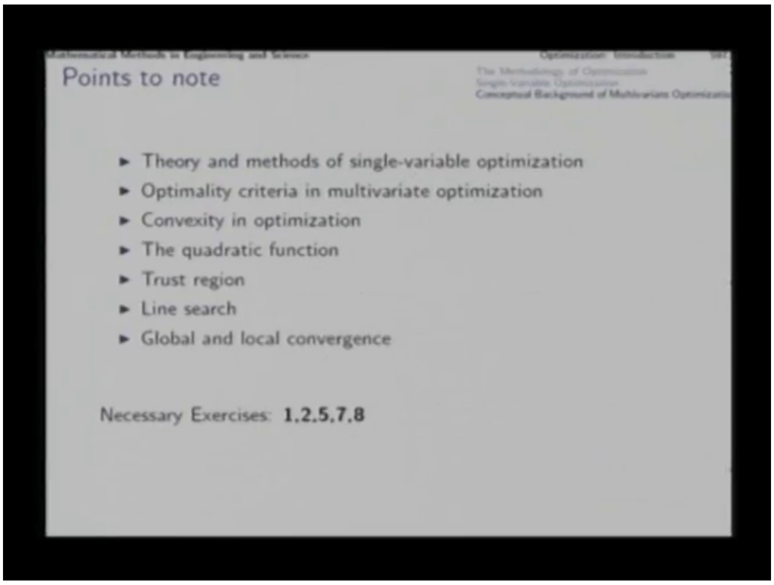
(Refer Slide Time: 52:44)



It is in terms of guarantee whether there is a guaranteed decrease of the functional value, whether there is a guaranteed approach towards the minimum point.

The other issue that arises, in terms of convergence of algorithm is the local convergence that is what is the speed of approach is if we start sufficiently close, so global convergence refers to the guarantee of approach from anywhere in the solution space, local convergence refers to the speed of approach if started sufficiently close. So, some of the methods have linear convergence rate which are typically the slower methods some have quadratic convergence rate, which have which are typically first methods and there are algorithms which are in between.

(Refer Slide Time: 53:33)



Now with this much background, in the next lecture. We will try to study optimization methods and currently the points to note are here, and quite a few exercises are there in this lesson in this chapter of the book, and some of them you must attempt on your own to be very conversant with the idea behind this subject matter.

Next lecture we continue into optimization method

Thank you.