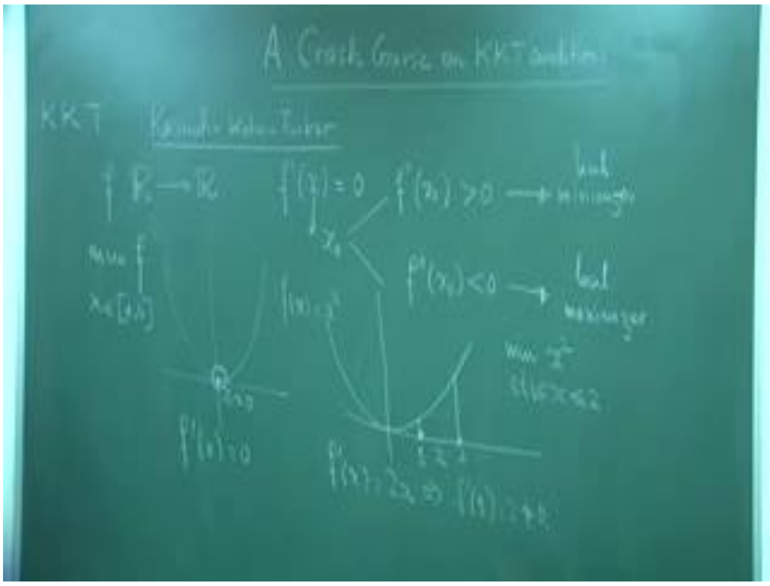
Probability and Stochastics for finance-II
Prof. Joydeep Dutta
Department of Humanities and Social Sciences
Indian Institute of Technology, Kanpur

Lecture – 05
A Crash Course on KKT Conditions

So, today we are going to learn portfolio optimization. And as we start portfolio optimization, we have to know certain basic facts of optimization so instead of directly jumping to the portfolio optimization the first course, the first last lecture of the first week should end, we equipping with tools in order for you to understand portfolio optimization in a better way. And hence today's lecture is a crash course on the KKT conditions.

(Refer Slide Time: 00:53)



KKT actually is name of 3 persons Karush Kunt and Taucket so these people are responsible to begin out the Lagrange multiplier rule when you have inequality constants. What is the Lagrange multiplier rule? The Lagrange multiplier rule that you have learnt in school, in your may be first year college of minimizing of function or maximizing of function with respect to one constants and that constant is equal to 0, one constant. What means that x whatever x that you choose, in order to find maxima and minima should be satisfying that inequality.

So, that idea of doing it to the method of Lagrange is bought to it is more general form when you also include inequalities. But modern applications have inequality there is a hallmark other than inequality. It has you will very soon see as you seen already in the modals that we have constructed optimization modal that we have constructed, and that is the basic structure. Inequalities are the hallmark of modern optimization so we are going to study the Karush Kunt Taucket conditions right. You all aware in high school, let be study you take functions like this and we want figure out whether these are the maxima and minima provided the function if is also given to be the differentiable.

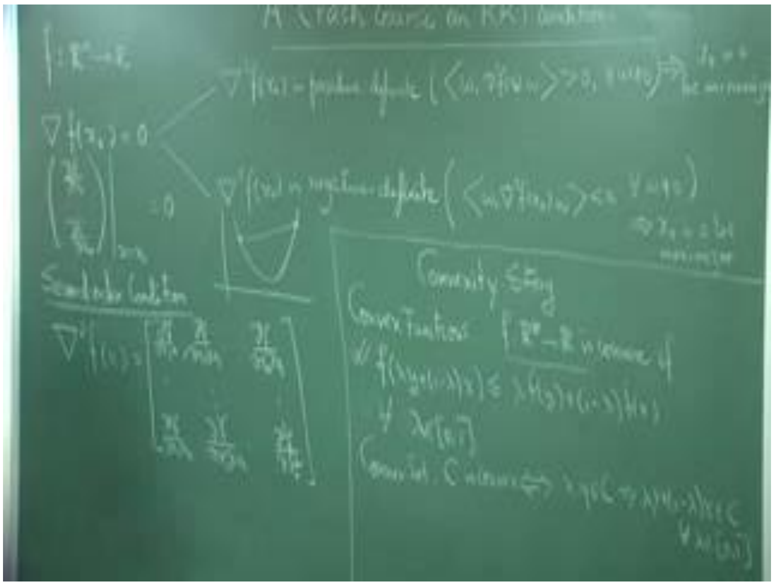
So, what we do is, which are the figure out what x would satisfy the in equation that f dash x is equal to 0, derivative at x is equal to 0. Suppose such point is x not, but this condition you must remember is a necessary condition and not a sufficient one. That is when you get a x naught by solving in this equation it does not guarantee you that there is truly a local minimizer or global minimizer or local maxi miser or local maxi miser all this problem. And for that you need to talk about second order condition. We should aware in your school studies high studies. And now you try can take the secondary, were if I cannot take the secondary activities, then there is big issue. If there is if can take the secondary with if then and compute it that x not and see what happens. If it is strictly bigger than 0 you know this is will be minimizer local minimize, local than x not is the local minimizer. Actually this something calls as a strict local minimizer, but we are not going to get into those optimizations in deeper optimization issues at this moment. Rather for the course you did not get too much bogged up by those issues and if these happen then you get what is called a local maxi miser. This is some; this is the repetition of what you already know from high school, these just recollections.

Now, is the simple tool, simple idea or simple mechanism by which you can possibly find the maximizer and minimizer of the function from R to R, is this applicable when I additionally put a constant that is, if I minimize function say from R to R, subject to x element of some interval. Can I use the same idea to have it? Answer will surprisingly turn out to be no. For example, if you take the function f x equal to x square so now, you want to minimize towards the whole R, this is called inconstant minimization. There is no addition or restrictions on, from where the variable should come it is free so here if you look at the function y equal to x square, then at x equal to 0 you obtain the minimizer. And of course, f dash 0 in this case is 0, but now suppose I take the same function y equal to x square, but I want to restrict it from, so now want to choose x within 2 and 1.

So, it is the same problem minimize x square subject to x belonging to interval whose interval to 1. There is no doubt that designs look at the picture and immediately determine 1 has the minimizer, 2 has the maximizer. In that these are global maxi miser. So now, on this interval no one more x equal to 0 is present. So the minimizer here is 1. And let me compute f dash 1. So f dash x is equal to 2x. This implies f dash one is equal to 2, not equal to 0; so if minima, if you are looking at a constant minimizer; it does not necessarily fallow f dash equal to 0. Here thus is a paradigm shift from going from unconstrained from to the constraint problem. And here you see I have constraint informal inequalities. And I could wright these are x less than equal to 2. And mines x less than equal to minus 1. So there is the huge paradigm shift.

So, the question is, if I have a constant problem how do I write, what sought of necessary conditions, of being followed so that, I would be able to use of necessary conditions at least compute as point which I can suspect to with the minimizer. And then test it through other conditions. So let us jump to higher dimensions, and that would make you have a much better understanding what is coming. So in order to move towards the Karush Kunt Taucket conditions, it is important that we again recall what would have happened.

(Refer Slide Time: 08:05)



If I look at a function f from in higher dimension on set up there is a function f moving from Rn to R. Then how do you decide how to find the local maximizer or minimizer. In that case you are first step of course, is to find point x not which satisfies the gradient equal to 0. Basically it is gradient simply means, del f del, x 1 so vector of n components del f del x n, if I calculate them at the vectorial point, x equal to x not that should be a mean the 0 vector.

So, find as a point x not that is all. But once I find the point x not, I do not have much options I cannot say that x not is the minimizer or maximizer. But in many problems you can actually try to rubbing to the problems structure and try to argue, but in many cases we are to realize on what is called second order conditions. Exactly progressing has we have done in the case of the function from real line to the real line. The same idea which mean progressed. But of course, you are in a much higher dimension setup so thinks cannot be just writing in bigger than 0 lesser than 0. So what you have to look into to the following. You have to know how to compute what is call the hessian matrix of f.

So, the hessian matrix is nothing, but the Jacobean matrix of the gradient. Or rather is the matrix whose individual rows are the gradiance of del f del x 1, dot, dot, dot del f, del x n, so if I take the gradient of del f del x 1, so I take it with first with x 1, del f del x 1, square del to f del x 2 del x 1, del 2 f, del x 1 del x, and so I have taken the gradient of sorry, del x n del x 1, I have taken the gradient of del f del x 1, del f del x 1 actually of function right from R n to R. So you have del 2 f, del x 1, del x n so these are the matrix is the symmetric matrix provided if the function is differentiable twice differentiable and all these are continuous, and by young theorem this matrix is becomes twice continuous in differential function, then the hessian matrix becomes symmetric.

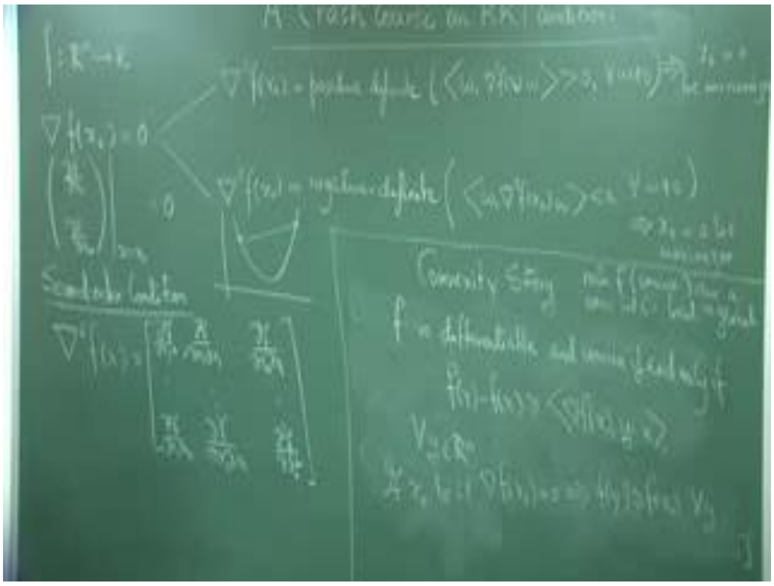
So, those were some studied in some basic calculus. Now what we should not be very much intimated by this sought of expression. So once of find x not what I am supposed to do? Second that to 2 fold parts. So see if this is positive definite. And then again see an also check negative definite. So if it is positive definite, positive definite means that so take any vector w, and take this inner product first operate f, f or x not on w, and then take in a product with w, and this is strictly bigger than w for all w not equal to 0, we call this to be positive definite. So if this is 0 grad effects not 0 and this is positive definite, and imply that x not is the local minimizer. While the opposite is of course, true we just have to change signs. And that would imply that x not is a local maximizer. Once I know these two, so we have a complete picture and how to handle unconstraint problem of course, it is not all easy to find out this, but we still have a complete picture to handle unconstraint problem.

Now, to begin the KKT story, which would go back to something more, is there any the convexity story? So here it is talking about finding secondary event that is exits and then you can do all these. Can there we some other conditions on the function so that whenever I have these and x not satisfying these that would give me a minimizer. In that case, one of the main assumptions on function f is that of convexity. So we are going to look at convex functions. So convex functions of functions of this form, so f from or it is convex, if f of lambda y plus 1 minus lambda x, is less than equal to lambda f y, plus 1 minus lambda f x, for all, this is the symbol of for all inverted a for all lambda line in the close interval 0 1. Of course, this simply means that this is an element then the line segment joining x and y. This is something you have already known in have coordinate geometry on in basic high school. And also corresponding there is a concept of convex set.

So, C is convex if an only if, these the symbol of if an only if, x y element of C, implies that in the whole line segment joining x and y, would be inside C. Observed that here if I would lambda equal to 1, I would get y if I would lambda equal to 0 I would get x so as I move from 0 to 1 I change lambda from 0 to 1, and actually moving along a straight line from x to y. So a major idea! Why convexity is important? Convexity is important, because if we want to minimize convex function over a convex set, that is constant optimization, in fact, then there no local minimizer. Every minimizer is a global minimizer. So if you would you know our whole space R and itself you also convex set. So we will largely we bother about functions of R n to R because in will literature you can also talk about functions some given convex at C to R and see whether they have this behavior convexity from a more geometrical point of view, means the following. Means that if you have a convex function and if you have you take 2 points on the graph of the function. And if we joined them by a code, that code would always line of the graph of that function line, after between these 2 points.

So, the portion of the graph of the function, line between this of this point will always line below this particular line segment joining these 2 points. That is exactly the meaning of the statement. So what we are now going to show, is that if we have convexity, right, if we have convexity of the function f, then you do not look have to look for second order condition. The reason is this. Suppose I have a function which is differentiable and convex.

(Refer Slide Time: 18:21)



Then we have following nice expression that f will satisfy a following nice inequality. F is differentiable and convex if and only if, said this is write iff, but would the whole thing if and only if, f y minus f x is greater than equal to, gradient away if at x in our product with y minus x. And this is for all y elements R n, which is bother about convex function from R n to R.

So, what happened? Suppose so this will so effect fix upon x, for whatever y I take this will true. So if x not be such that, grade effect not is equal to 0, employing this in equal little it will mean, f y is greater than equal to f x not, for all y. So this is the very important thing. That for a convex function, so if you minimize f convex over a convex set C, then local is equal to global and every local minimizer is a global minimizer. These are very short hand way writing. Do not think I have written mathematical in equation, local equal to global. Some small they just a simple way of writing, so we will now look come to the constant case and explain to you what is the Karush Kunt Taucket condition. So now, as we promise will start our discussion on the constant optimization problem.

(Refer Slide Time: 20:33)

Here, we will not focus; on equal focus on this in is general structure, what standard structure. So it is the one we had already written down. In the last class, however, we are not going to include x element of capital x. We will not be going to include extra, take this most simple format. Then we are going to discuss what sought of conditions, these would follow. If we have a local maxima minimize, we are going speaking down minimizer only, as I have already told in finance also a uni minimizing always one, in the optimization. Only so let us listen here, that here we would assume f, g i, h j, for all smooth. Smooth meaning continuous differentiable. They are differentiable and then the partial derivatives also continuous.

Now, once I do that, suppose x not is a local minimizer. Then what? So x not is a local minimizer, what sought of conditions do x not satisfying in terms of the given data of the problem. So it was Fridgejohn in 1948, Fridgejohn is one name is not 2 names. Usually some people also called the john conditions. He wrote of paper giving the necessary condition and also use it to device certain algorithms. And send it due general of mathematics which rejected it. Later on it published in some conference proceedings. But this is one of the most fundamental conditions that we need to know when we studying some basic optimization. Fridgejohn condition do not require any addition condition it tells ok, if you have smooth functions, this is the problem it is not is the local minimizer then this will happen, but Fridgejohn says that if x not is a local mean, then there exist scalers lambda not, which would be associated with the grade aware, greater than equal to 0 lambda i, which is associated with the grade of g, and mu j, it is just an element odd would have any sign.

So, this is the sign call, they are exists. On those were not familiar with mathematical symbols, this simply means they are exist. So there exist scalars, so there exist scalars. This nature sees when you have inequality is your sign known the so called multiplier, these something like Lagrange multipliers. So then they are scalars such that, lambda not grade of f x not, plus summation i is equal to 1 to m lambda i grade of g i x not plus, summation j is equal to 1 to k, mu j rad a j x not is equal to 0. And lambda i g i x not is equal to 0 for all i. We call to 1 to m. So this is the standard optimize condition and lambda not, lambda 1 lambda m, mu 1, mu k, this full vector cannot be a 0. Vector means all of this cannot be simultaneous were all of these are simultaneous. It is 0 it; obviously, satisfy this condition. So we have to find out vector which were it cannot all simultaneous is 0, and of satisfied these. This is what Fridgejohn proved in 1948. And this condition is called the complementary slackness condition of very important tool in optimization very important notion.

This simply means that, you cannot have both of these holding restrict inequalities at the same time. You cannot have a lambda is strictly bigger than 0, as well as the same time g i is strictly less than 0. This is not possible. Where is lambda is strictly bigger than 0 g i x not must be equal to 0, if we that is the weight is because then you will maintain this inequality, and this is very fundamental. So the drawback; however, of the Fridgejohn points was that, lambda not could be 0. If lambda not could be 0, then the roll of the objective function goes away from the problem. And that something you would like. And there are many, suppose if I now up, if some convexity happens on the f and g, then unless I know with lambda on strictly bigger than 0, I have no way to prove then under such conditions x not will again give me, although one minimizer. And they work there could be feasible point which are not optima. They can still satisfy the Fridgejohn condition. The Fridgejohn condition as the very important role is that it provides a negative negativity certificate. That is in the sense negative certificate, that is if there is x not which violates the Fridgejohn condition, then it is definitely not of the local minimizer of global minimizer.

So, how to divide the condition at this annual is a Fridgejohn condition. Rather like lambda not becoming 0, go goes away and that was frame by Karush Kunt and Taucket, so originally frame by Kunt and Taucket in 1951, published and in 51, and then they found at Karush had already done some work of similar type, in his MS thesis, and then Kunt wrote to him and then the or the historical facts straighten. So Kunt and Taucket 1951, is the paper which every what you referred to. What they say is the following. So of the same conditions, so with the Fridge John condition, add with it some conditions on the constraints, that is called constraint qualifications. Once you do that, you can show that, by choosing a proper constraint qualification lambda not is not strictly bigger than 0. Then you can divide everything by lambda not, and then you knew multiplied where these conditions would be maintained. And the Kunt Taucket condition would look like grade effects not plus lambda i dash grade g i, x not plus mu j dash that h j x not. So this can be normalized to 1. So basically would once you know that lambda not is simply greater than 0 divide whole thing both side well lambda not.

So, that is what is Kunt Taucket condition. Fridgejohn condition plus and constraint qualification will give you this. Suppose I say that I am in situation do check that all the gradient vectors of g i and all the gradient vectors of a j i nearly independent then, even easily show that lambda not strictly greater than 0. So rather go ask you go to the web and figure out what certain constraint qualification are there. Now the importance of Kunt Taucket condition comes when you are talking about convex functions. So if f is convex if g i is convex, and h j is a faint function are all discuss when you are talking about linear programing, h j is a faint, and KKT condition holds at a feasible x not, then x not is a global minimizer.

So, would like that at home, you would write down, figure out what are constraint qualifications. And then you can show mathematically, you tried to show mathematically, this is happening then lambda not strictly greater than 0. I can divide also it is by lambda not basically choose lambda not equal to 1 and then write down the Karush Kunt Taucket conditions.

So with this I would end todays talk, and when we start next week we really start with art of finance, we will talk about portfolio optimization. And that would be doing 2 parts, and then will go to the capital asset pricing modal.

Thank you very much.