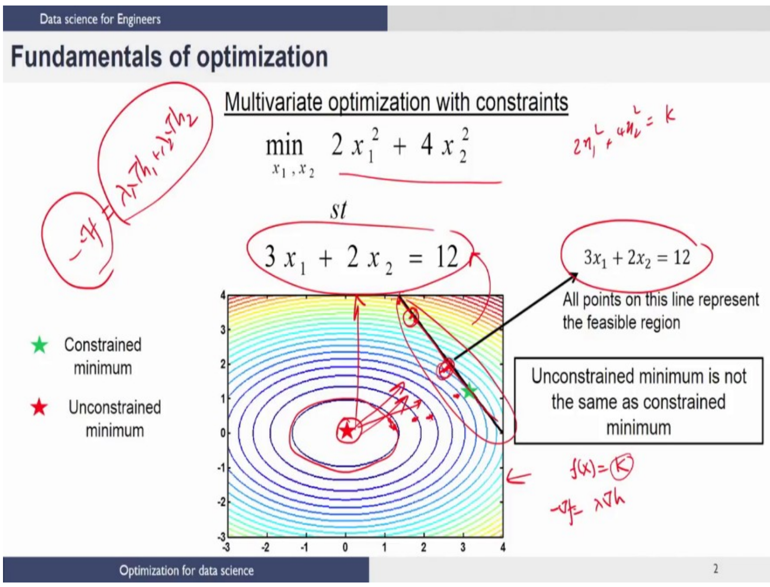
Data Science for Engineers
Prof. Raghunathan Rengasamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 27
Multivariate Optimization with Equality Constraints

Let us continue our lectures on optimization for data science. In this lecture we will look at multivariate optimization non-linear optimization. However, in the previous lectures we saw the unconstrained version of this problem. In this lecture we will look at problems of this type where we have what are called equality constraints. I will explain that is presently.

(Refer Slide Time: 00:46)



And I am going to explain this using a very simple example, but before we dive into this example let us set some context and then ask the question as to why we should be interested in this in a course on data science. The reason why we should be interested in this in a course on data science; the reason why we are interested in constraints in optimization is as I mentioned before, we look at optimization from a data science viewpoint because we are trying to minimize error in many cases, when we try to solve data science problems. And when we minimize error, we said we could use some kind of gradient based algorithm what we called as the learning algorithm to solve the problem.

In some cases while we are trying to optimize or minimize our error or the objective function, we might know some information about the problem that we want to incorporate in the solution. So, if for example, you are trying to uncover relationships between several variables and you do not know how many relationships are there, but you know for sure that certain relationships exist and you know what those relationships are, then when you try to solve the data science problem you would try to constrain your problem to be such that the known relationships are satisfied.

So, that could pose an optimization problem where you have constraints in particular equality constraints and there are several other cases where you might have to look at the constrained version of the problem while one solves data science problems. So, it is important to understand how these problems are solved. Once we look at equality constraint problems we will look at in-equality constraints which is even more relevant. For example, the algorithms for inequality can with inequality constraints are very useful in data science algorithm that is called support vector machines and so on. So we going to look at both equality and inequality constraints.

Now, let us look at this problem right here. So we are interested in mini-mizing the function here which is $2\ x_1^2 + 4\ x_2^2$ and like we discussed before this is an objective function in 2 variables x_1 and x_2. So, real visualization of this would be in 3 dimensions where the objective function value is plotted in the z direction. However, we know that we can work with contour plots and if you look at this picture here, you see these contour plots. These are plots where each of this contour is a constant objective function contour and again you would see that these are ellipses because if you look at constant contour plots then you have $2\ x_1^2 + 4\ x_2^2$ equal k this you will see is an ellipse that is what is plotted here.

Now, if you look at the optimum value for this function and just by inspection you can see that the optimum value is 0 because this are functions where I have terms which are squares of the decision variables. So, there are 2 terms here $2\ x_1^2 + 4\ x_2^2$ and the lowest value that each one of these could take has to be 0 and that basically means the unconstrained minimum is at $x_1 = 0$ $x_2 = 0$ the objective value at that point is also 0.

So, which is what is shown here in this point right here. So, if I had no constraints I would say the optimum value is 0 and it is at 0 0 the star point here. Now let us see what happens if I introduce a constraint in this case I am going to introduce a very simple linear constraint. So, let us assume that we have a constraint of this form which is $3\ x_1 + 2\ x_2 = 12$.

Now what this basically means is the following. You are looking for a solution to x_1 and x_2 which also satisfies this constraint that is what it means. So, though I know the very best value for this function is 0 from a minimization viewpoint, I cannot use that value because the 0 0 point might not satisfy this equation. So, you will notice that if I put $x_1 = 0$ $x_2 = 0$ it does not satisfy this equation. So, the unconstrained solution is not the same as the constrained solution.

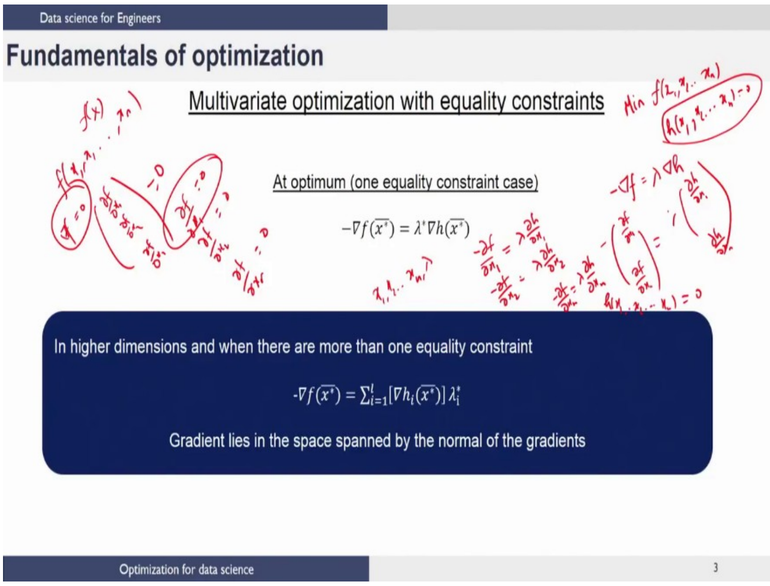
Now, let us see how we solve this problem. To understand this we first start by representing this constraint in this 2 dimensional space. So, since this is a linear constraint you will see that it is a line which is here which is what is given by this constraint. Now since we said that whatever solution I get it should satisfy this constraint would translate to saying, I can only pick solutions from this line because only points on this line will satisfy this constraint. Now, you notice that you could pick many many points on this line. So, for example, you could pick this point and the objective function value at that point would be corresponding to a contour which intersects this point. So, if you pick a point here then you would have a corresponding objective function value. Now you could pick a point here then you would see that it would have another objective function value which would be corresponding to the constant function contour that intersects the line at that point.

So, you notice that as you pick different points on this line the objective function takes different values and what we are interested in is the following. Of all the points on this line, I want to pick that one point where the objective function value is the minimum. To understand this let us pick two points and see what happens. So, if I pick a point here and let us say I pick a point here and ask the question which one of these points is better from a minimization of the original function view point. The way to think about this is the following.

So, when I pick a point here I know that the objective function value would be based on the contour which intersects at that point and if you compare these 2 points you will see that this point is worse o than this point, from a minimization viewpoint. This is because if you look at these contours these are contours of increasing objective function values and the contour that intersects this point is within the contour that intersects the line at this point. So, basically what that means, is because this is the direction of increasing objective function value the contour intersecting this point is inside the contour intersecting at this point. So, that basically means this function takes a lower value at this point on the line.

So, as you go along the line you see that the value keeps changing and my job is to find that particular point where the objective function value is the minimum. So, this is a basic idea of constrained optimization solution that we are looking for. The key point to notice here is that the unconstrained minimum is not the same as the constraint minimum. If it turns out that the unconstrained minimum itself is on the constraint line then both would be the same, but in this case we clearly see that the unconstrained minimum is different from the constrained minimum.

(Refer Slide Time: 10:20)



So, when I have just only one constraint, how do I solve this problem? I will first give you the result and then explain the result again by going back to the previous slide and then showing you another viewpoint and then how that leads to this solution.

So, if you typically take a multivariate function f(x) in the unconstrained version, let us assume that x is x_1 x_2 and x_n. We know that the optimum solution is grad f = 0 which basically means that $\partial f/ \partial x_1$, $\partial f / \partial x_2$ all the way up to $\partial f / \partial xn$ equal 0. So, this when I expand this further I will get equations $\partial f/ \partial x_1$ equal 0 $\partial f / \partial x_2$ equal 0 $\partial f / \partial x_1$ 0 and so on.

Now, notice that if there are n variables there will be n equations of this form. So, I have n equations in n variables. So, I can solve for it and find a solution and once I find a solution to find out whether it is a maximum or minimum or a saddle point I calculate the second derivative and then construct the hessian matrix and then depending on whether the hessian matrix is positive definite positive negative definite or semi definite and so on we can make judgments about whether the point is a minimum point maximum point and so on.

Now, let us see what happens if I am trying to solve a problem where I have to minimize f(x) which is x_1 x_2 all the way up to x_n. But like I said before let us assume that I also introduced one constraint and I am going to introduce let us say a constraint which is of the form h of x_1 x_2 all the way up to $x_n = 0$. So, this is an equality constraint. So, we can always write a constraint in this form, even if you have some number on the right hand side you can always move to the left hand side and write this constraint as something equal to 0.So, basically my job is to find the minimum point for f subject to this constraint. I will just give you the result and then we will see how we get this result.

So, when I want to solve for this problem the result is in this case grad f = 0 itself gave us the result. In this case it will turn out that the result is the following. So, we can write the negative of ∇ has to be equal to some λ ∇ of h. So, you can write this as negative or drop the negative and the sign of the value λ will take care of that, but I am writing in this particular form.

So, just to expand this basically says I have something like this I have domain is $\partial f / \partial x_1$ all the way up to $\partial f / \partial xn = \lambda\ \partial h / \partial x_1$ all the way up to $\partial h / \partial xn$. So, if we expand this further I am going to get n equations. In this case I got these equations to be 0 in this case I am going to get equations of the form - $\partial f \partial x_1 = \lambda\ \partial h / \partial x_1$ will be one equation, the second equation will be - $\partial x_2 = \lambda\ \partial h / \partial x_2$ and so on, the last equation will be - $\partial f / \partial\ xn = \lambda\ \partial h / \partial xn$.

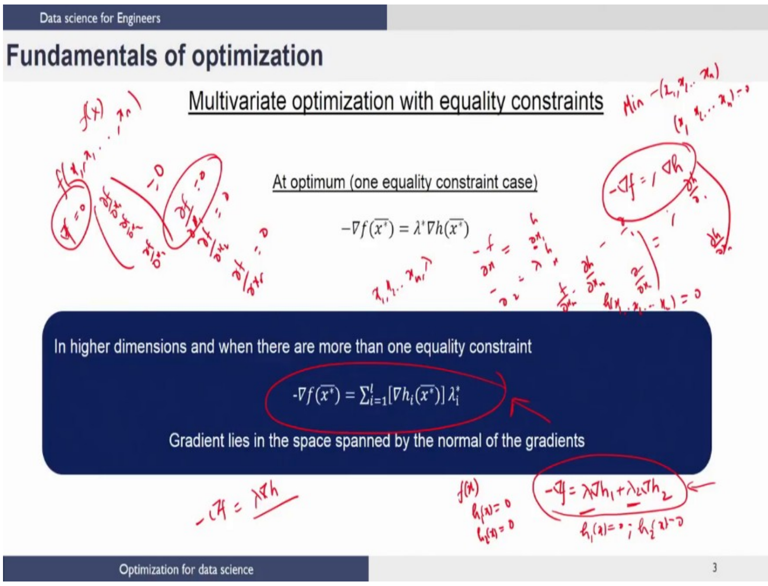
So, now notice much like before I have n equations the difference being in the unconstrained case I had zeros on the right hand side in the constrained case I have these terms on the right hand side. Nonetheless these equations are in now n + 1 variable because I have my x_1 x_2 all the way up to xn and I have also introduced a new variable λ right here. So, my equations are in n + 1 variables. I have only n equations at this point, but notice that I have one more equation that I need to use and that equation is the following if I find some solution x_1 to x n which satisfies all of these equations.

Then that also has to satisfy the constraint. So, here we are only talking about the gradient form of the various functions, but the equation which represents the constraints also needs to be satisfied by any solution that we get for the constrained optimization problem. So, with these n equations I will also get another equation which is that h of x_1 x_2 xn has to be = 0.

Now, you notice that in this case with one linear constraint I have n + 1 equation in n + 1 variables. So, I can solve this, so to reiterate the difference between the constrained and unconstrained case was, in the unconstrained case we had n equations in n variables, in the constraint case with just one constraint we have n + 1 equations in n + 1 variables.

Now, you might ask what happens, if there are there is more than one constraint there are let us say 2 constraints.

(Refer Slide Time: 16:17)



So, we will see what happens if there are 2 constraints, so in this case we are going to minimize f(x) and now let us say we have 2 constraints we are going to say h_1 x = 0 h_2 x = 0. In this case what is going to happen is the following the solution to the constrained optimization problem is going to be - ∇f = λ 1 $\nabla$of h_1 + λ 2 $\nabla$of h_2.

So, if you look at this and the one constraint case you will see the similarities. In the one constraint case we introduced one extra parameter, in the 2 constraints case the x introduced 2 extra parameters. In the one constraint case, we simply said grad f - ∇f equal λ ∇h, in this case this λ ∇h becomes a sum of λ ∇h_1 λ 1 ∇h_1 + λ 2 ∇ has 2 and so on.

Now, if there are 3 equality constraints, then you would have 3 terms here and if there are l equality constraints you will have something like this here where this is sum of l terms. So, that part is clear the other part that we need to worry about is if I have enough equations to solve for all the variables when I have 2 constraints, that basically means I have introduced 2 new variables λ_1 and λ_2.However, this equation irrespective of the number of equality constraints you have will always be n equations.

Because grad f would be ∂f/ ∂x_1 ∂f /∂x_2 ∂f n/ x n so these are all n by 1 vectors. So, this will just give me n equations. Nonetheless if I had 2 equality constraints I need to find the 2 extra equations which are directly given by the constraints. So, since the optimum point has to satisfy both the constraints, I get one extra equation x_1 x = 0 and the other extra equation is 2 x = 0. So, if you have 2 equality constraints in n variables I will have n + 2 variables and n + 2 equations and we will always find this to be true because if I had 3 equality constraints, I will have n + 3 variables which would be x_1 to x_n $\lambda 1$, $\lambda 2$, $\lambda 3$ and this gradient equation will always give me n equations and the 3 extra consent would have given me the 3 extra equations.

So, I will have n + 3 equations and n+ 3 variables which can be directly generalized to this form where I have l equality constraints. So, let us see in the single constraint case how we get an expression like this that that would be easy to see in the single constraint case. This expression where we have the sum of these terms is slightly more complicated to understand. I am not going to do that. I am going to give you an intuitive feel for why this is true in the single equation case and once you understand that with a little bit more effort you should be able to think about why it should be true for more equality constraints and so on.

So, we go back to the previous slide and when I was discussing this slide and explaining how equality constraints affect the optimum solution. I said there are many points on this line which could all be feasible solutions. Feasible solutions meaning those are all solutions which can satisfy this equation. Nonetheless, there are some points out of those or one point which would give me the lowest objective function value. So, when we looked at candidate solutions for this optimization problem we were looking at the points on this line that that we are interested in because that is a constraint.

Now, let us take a slightly different viewpoints and then look at the same problem from an objective function viewpoint. Now from an objective function viewpoint if you did not constrain me at all and you said you could do anything you want, then I would pick this point as a solution. Now when you look at this point and then say well this is a best point I have let me find out whether it satisfies my constraint and then you substitute this point into this and then you figure out it does not satisfy the constraint.
So, you say look I have to do something because I am forced to satisfy the constraint. So, you will say let me lose a little bit in terms of an objective function perspective and then see whether I can meet the constraint.

So, when we say I want to lose a little bit basically you know as we mentioned before these are contours where the objective function value increases and those are actually not good from a minimization viewpoint. So, while I am here since the constraint is not satisfied, I am willing to lose a little to see whether I can satisfy the constraint and maybe I go to a point here and then this is a constant objective function contour point. So, if I am willing to give up something that basically means I am going and sitting on different points on this contours and as I am pushed away and away from this minimum point I am losing more and more in terms of the objective function value. That is I am increasing the objective function value. Now logically if you keep extending this argument you will see that, let us say this is the first point I moved here which is basically worse than this because you see this is a contour which is going to be outside of this contour, but I moved here I made my objective function worse, but I still am not satisfying the constraint. So, I give some more I come to this point and I see a contour and this point is worse than this because this contour is outside this contour and if I extract this argument let us say I keep making things worse and the only reason I am making these worse is because I am forced to satisfy this equality constraint.

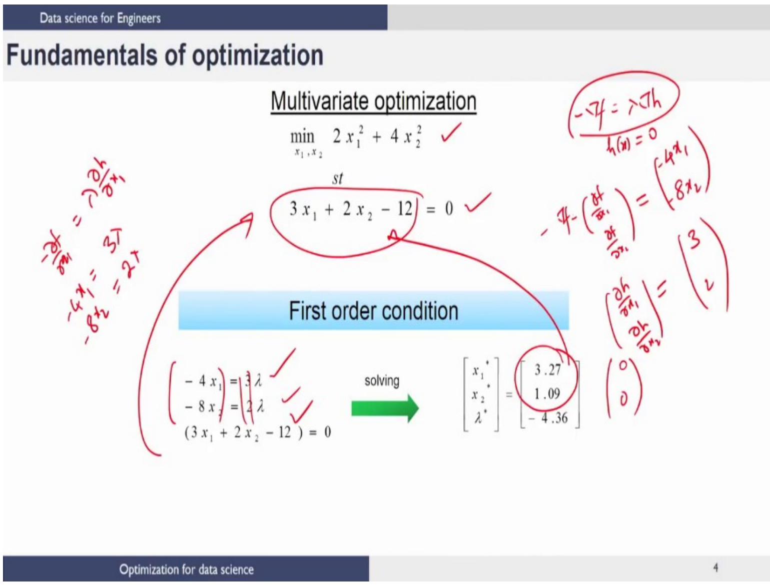
So, I come up to let us say here and this is still a contour which is much worse than my original solution, but it is still not enough for me to satisfy my constraint. So, if you keep repeating this process, you are going to find a contour here where I touch this line for the first time. So, when I touch this line for the first time is the point at which for the first time, when I give up my objective functions value, I am also able to satisfy the constraint. Now once I find a contour like that which touches this line then there is no incentive for me to go further beyond because going further beyond would mean I would be making my objective function worser.

So, when I just touched that line that is the best compromise because that is where I become feasible for the first time and going any further would only make my objective function worse. So, geometrically what this would mean is I keep making my objective function worse till a contour just touches this line. So, at that point this line will become a tangent to that contour and remember what that contour is that contour is $f(x) = k$ for some k

So, we have to choose a k in such a way that this contour, when I have the constraint, the constraint becomes a tangent to this contour and optimum point. This geometric fact when you represent it as equations, you would get the form that I showed which is - grad of f is λ times grad of h. So, this is for the one constraint case, now when you have many equality constraints while it is not as easy to see that as it is in the single equality constraint case the statement that - grad f let us say if I have 2 is λ 1 grad h_1 + λ 2 grad h_2 this statement basically says that this gradient negative of the gradient has to be written as a linear combination of grad of h_1 and grad of h_2 and that has a geometric interpretation.

So, that is the reason why we get these conditions for optimization with equality constraints the key point that I want you to remember is that as you have more and more equality constraints you will have to introduce more and more parameters λ1 λ2 and so on. However, there will always be enough equations and variables.

(Refer Slide Time: 26:47)



So, let us look at a numerical example to bring all the ideas together. We are going to use the same example that we had looked at in the previous slides. So, we are trying to minimize 2 x_1^2 4 + 4 x_2^2 as objective function subject to this constraint. So, we wrote the following equations for the optimum point for identifying the optimum point we said - ∇of f has to be λ ∇of h and then we have h of x = 0. So, this is what we wrote let us do the computation so that we understand this.

So to calculate ∇of f which is basically ∂f/ ∂x_1 ∂f/ ∂x_2. So, if you look at this objective function. So, f ∂x_1 will be simply 4 x_1 and ∂f /∂x_2 will be 8 x_2. So, we have / of f so negative grad of f would be negative. Let us look at grad of h. So, h is this equation so if I do ∂h / ∂x_1 ∂h/ ∂x_2. So, this is going to be = ∂h /∂x_1 will be simply 3 and ∂h ∂x_2 will be simply 2 so we have this.

So, now, let us see what this equation becomes. So, this equation you would see is if you put this in a bracket and you will see this easily. So, you have the first equation is - ∂f/ ∂x_1 = λ ∂h /∂x_1 and we have - 4 x_1 = 3 λ, similarly the second equation would turn out to be - 8 x_2 = 2 λ and this equation is basically the same equation as the constraint equation that we have here. And as we mentioned before we thought the constraint that would have been 2 variables and you would have got 2 equations which would have been ∇f = 0.

But with an equality constraint we have added a new parameter λ. So, we need 3 equations in x_1 x_2 and λ we do have these 3 equations here and when you solve these 3 equations you will get this solution and this is your optimum solution in the constrained case which is different from the optimum solution in the unconstrained case which would have been 00. So, in other words we have given up on the value of the objective function.

However, this is a point which would satisfy this equation of the line. So, that is how we deal with equality constraints in an optimization problems. I already described how these are useful or why we should study them in the first place from a data science perspective. With this I will conclude this lecture and in the next lecture we will look at how we handle inequality constraints and I will also explain why we are interested in understanding how optimization problems are solved with inequality constraints from a data science perspective.

Thank you and I look forward to seeing you again in the next lecture.