Artificial Intelligence for Economics

Prof. Adway Mitra

Artificial Intelligence

Indian Institute of Technology Kharagpur

Week – 02

Lecture - 07

Lecture 07 : Constrained Optimization

Hello everyone, I am Adway Mitra, an Assistant Professor at Indian Institute of Technology Kharagpur and I welcome you to this course on Artificial Intelligence for Economics. Today is the lecture 7 of this course and our topic today is going to be Constraint Optimization. So, in the previous lecture which is lecture 6, we had dealt with the we had introduced the concept of optimization and we had also discussed why it is extremely useful in economics. in many economic applications we deal with the like the problem of optimization. And we had also seen that like in case of optimization problems there is an objective function which we want to maximize or minimize, but most of the mathematical techniques that we use like look upon it as a minimization problem. So, that even if it is a maximization problem we change the objective function so that it becomes a minimization problem.

Now, we had also seen several algorithms some of them are like analytical that is based on differentiation and equating to 0 and some of them are numerical that is for example, based on gradient descent where we start with an initial solution and move step by step towards the actual solution. So, those are all useful techniques. The problem is that in those techniques are useful only for unconstrained optimization problems. That is to say like there is an objective function which we want to minimize, but we do not put any constraints on which solutions are acceptable.

So, we had earlier discussed the concept of feasible regions or the feasible solutions. So, a feasible solution when we are considering the feasible solution which is not the entire space of variables I mean the entire space which the in which the variable leaves. In that case we the methods like gradient descent will not work because even though it will minimize the function it will the solution at which the function is minimized may not be a feasible solution that is it may not satisfy the additional constraints which the like which may have been imposed on the solution. So, in this case we are now going to talk today about the another problem which is called the constraint optimization which is you can say one step away from or one step further from the problem of unconstrained optimization which we had discussed in the previous lecture. So, in case of constant optimization as usual there is an objective function which we want to minimize or maximize with respect to one or more variables.

For example, we may want to allocate resources to different sectors to maximize the net revenue or the net utility from all of them. The concept of utility function we had discussed in details in lecture 6. So, once again we consider the same problem of resource allocation which arises frequently in economics. you want to allocate some resources into in different sectors let us say. However, not all possible solutions are acceptable.

So, it like it might be that you it the for the best results you if you like if as you go on increasing the amount of investment to each of the sectors then it is expected that you will probably get the best returns. So, you just go on increasing how the your investment and that is all you need to know. But in general this is not a feasible solution why because the total amount which you have to for investing the budget which you have is limited. You like that is whatever allocation of resources you want to do to the different sectors the total amount of resource however may be limited. So, that is in some cases in some way we can say that is a constant which you may have.

Apart from that you may have some other kinds of constraints also. For example, there may be a particular sector to which you do not want to allocate more than a particular sum of amount or you may want that the allocation of resources into the different sectors it should be fair allocation. That is it should not happen that you are pouring 90 percent of your resources into one sector and 10 percent resources in all other sectors together. that is not allowed. Let us say that like you set some limit that no sector will get let us say more than twice the resources which is allocated to any other sector.

So, these are various constraints which we can define on the allocation. So, then this becomes a constraint allocation problem. It is no longer enough to minimize or maximize the net utility which you are talking about across all the sectors, but in addition to it the whatever allocation you are proposing that allocation must also satisfy certain constraints like the ones which we just said. So, the solutions so it is possible that the solution to the first problem that is the an allocation which maximizes the utility function it may not be feasible because it may violate more like one or more such constraints right. So, like only those solutions which satisfy all the constraints we will consider and those are the feasible set.

Now it may happen that if you consider only or if you are restricting yourself to only those solutions which satisfy all of this constraint then this So, like utility function will not be maximized as much as it would have been and we not had this constraint, but there is no way around it. So, like restricted to these few constraints you have to do as best as you can. So, that is called as the constraint optimization problem. So, let us start with the simplest of constraint optimization problem which is the constraint linear optimization. So, what happens here as the name suggests here the objective function which has to be maximized or minimized that is a linear function as you can see like $c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$.

So, there are n variables as you can see each of them has its own coefficient. Like there might have been a constant term also, but that is not relevant anyway because we are looking for the variables only. So, this is the thing which we want to maximize. However, we have certain constraints as already mentioned. So, now these constraints they can be of these types.

These are the constraints are also of the linear like linear constraints. Like you as you can understand each like here we are saying m inequalities each of them each of these m inequalities involve the n variables. So, furthermore we have one more criteria that is all the variables they are non-negative. So, these are like we can say some $n+1$ sets of constraints sorry $m+1$ set of constraints which we have. Note that each of these constraints like they are first of all there it is a linear relation and secondly it is a less than or equal to relation.

So, like the standard practice is the if we have this kind of thing like b_1 or b_m we bring them to the left hand side and just make it and just bring it to the form like say $h(x) \leq 0$ or something like that. So, that is a typical inequality constraint. So, now we convert the problem the constraints into a system of linear equations with the help of slack variables. So, we right now we are having inequalities. Now that means what that means if we add some quantity to the I mean like here inequality means that the left hand side is less than or equal to the right hand side.

Now, so that means, if we can now add some quantity on the left hand side I mean a non negative quantity if I add on the left hand side then it should become equal or it can become equal to the right hand side. So, the quantity which has to be added so that this inequality becomes an equality that is known as a slack variable. So, we have right now we have a system of inequalities we convert them into a system of equations by like introducing some new variables which are known as slack variables. So, obviously, each of these so, total m inequalities are there each of them have to be converted into equations. So, we have to add m new variables.

once we have that then we have got a system of linear equations and we often know how to solve a system of linear equations. We already had n variables now we have added these new m slack variables. So, total m plus n number of like variables we have got and we also have got m equations. So, now in general we know that like we need as many system I mean as many equations as there are variables, but in that this case this that is not happening. So, we will not be able to apply the standard methods and of algebra and just get the get a unique solution.

but instead we will have to consider like a family of solutions. So, it may happen. So, this is typically it is an under determined system that is there may be more than one unique solutions here. What do I mean by unique solution? Unique solution is one particular assignment of values to all of these m plus n variables. So, m plus n variables m equations.

So, it is entirely possible that there are multiple solutions to this problem. So, now what do but our aim is not to solve these constraints our aim is to maximize the objective function. So, each solution of these constraints. So, let us just say there are 100 solutions or 100 possible values of the these m plus n different variables which satisfy all of these constraints ok. So, then we have the question is out of these 100 possible solutions which of them maximizes the value of p right.

So, we will have to plug in those 100 solutions to this p and c what is the I mean which of those maximizes the value of p. So, that is the very broad idea. Now, like just to see an example here we had like this was our original problem, this is the objective function which we want to maximize and these are the constraints. So, now these constraints which are inequalities we convert them into equation by adding these new slack variables $s_1, s_2, \ldots, s_p$ and so on. So, now we using this we like define something known as a basic solution.

So, what is a basic solution? Even a linear programming with n decision variables and m constraints a basic solution of the corresponding initial system is a solution of the initial systems in which n of the variables are equal to 0. So, total as I already mentioned m plus n variables are there. If any at least n of them are equal to 0, it could be that all of these original variables which are called decision variables, one possibility is that all of them are equal to 0 and the remaining are non-zero that is one possible basic solution. it could also be that some of these are 0s and some of these are also 0s that is the more general solution. So, like each of the these configurations where at least n of them are equal to 0 that is known as a basic solution.

Now, we define any further What is the basic feasible solution? Now, if a basic solution of the initial system corresponds to a certain point in the feasible region of the original linear program, then it is called as a basic feasible solution. What does that mean? So, point in the feasible region of the original LP this basically means that such a point or such an assignment to these variables such that all of the constraints are satisfied. That is to say we are talking about like assigning values to these variables x_1, x_2, s_1, s_2 etcetera such that at least two of them are equal to 0 and additionally like they satisfy together they satisfy all of these equations. So, like if we plot it graphically then it will become clear that like these basic solutions are really the corner points of a like a high dimensional polygon or whatever we can call want to call it. So, like in like let us consider this two dimensional space.

So, there are two linear decision variables in this case x_1 and x_2. Now, we are defining the feasible region in this way. that is like. So, each of these straight lines we have two if you see it has two straight lines the first one and the second one. So, the feasible region is like the quadrilateral which is defined by these two in addition to these sides which enforce the this constraint that just a second.

which enforce these constraints that is the positivity constraints. So, $x_1 > 0$ $x_2 > 0$. So, those are these xs. Now, when we are talking about the first two criteria that is $4x_1 + 2x_2 \leq 32$ and thus other one. So, the first corresponding to the first inequality there is one line like this So, every point which lies on this side of the line they satisfy the inequality.

Similarly, for the second constant also there is the another line like this and everything that lies on this side of the line they satisfy the second inequality. Now, both inequalities have to be satisfied. So, we need to consider the intersection of these two regions where both of those I mean functions they will have a value which is less than or equal to 0 that is this quadrilateral and we also have these bounds which are imposed by these constraints. So, basically like we are searching for like in I mean the feasible region is this quadrilateral. So, only points which are within this quadrilateral they are acceptable and they are feasible and when we are trying to minimize the objective function that is this function we must only restrict ourselves to points within this quadrilateral.

That is if it were possible we would have to calculate the value of the objective function at every value within this quadrilateral and at whichever point that objective function is maximized we have to consider that point as the actual solution. Now we come to what is known as the fundamental theorem of linear programming. The fundamental theorem of linear programming basically tells us that. If at all a solution exists to this problem, then that value must occur at one or more of the basic feasible solutions of the initial systems. Now, what is basic feasible solutions where at least n of the variables are 0.

So, in this case in this example n is equal to 2 that means, out of the 4 variables that is 2 original variables x_1, x_2 as well as the 2 slack variables s_1 and s_2 out of these 4 at least 2 have to be 0. So, which obviously mean that they will have to be corner points like this.

Like for example, in this one we can understand that 2 of them are 0 because like first of all $x_2=0$ and secondly the like if this point lies on one of those constraints. So, the corresponding slack variable is also equal to 0. So, in the same can be said about this point that is x in this case $x_2=0$ and also it lies on one of the constraint equations.

So, that constraint I mean the corresponding slack value is also equal to 0. In this case of course, both x_1 and x_2 are 0. In this case neither x_1 nor x_2 are 0, but since this point lies on the intersection of those two equations. So, the corresponding so, the like the equalities hold anyway that is the corresponding slack variables s_1 and s_2 they are equal to 0. So, total 4 points we are seeing like this where at which are the feel like the basic feasible solutions.

Now, the what does the fundamental theorem of linear programming tell us it says that the true solution must lie in at any of these 4 points. So, we do not really have to go about searching every point in the quadrilateral we just have to evaluate these 4 limited points which are is of course, easy to do at each of these 4 points we calculate the value of the objective function p. and whichever gives them the like whichever point gives the highest value of P, we report that to be our solution. Say for example, in case of $(0,8)$, the solution is obviously 32.

In case of this $(8,0)$, the solution is 40. In case of $(6,4)$, we see that this becomes $30+16$ which is 46. and so on and so forth. So, this way like at each of these points we can evaluate the objective function and whichever of these points it gives us the higher maximum value we declare that as the solution. So, like we do not really have to consider every point in the feasible region it is enough if we consider the boundary points. So, that is what the fundamental theorem of linear programming tells us.

Now, we come to another concept which is known as the Lagrangian multiplier like. So, this is you this can be used when we are having equality constraint, but it can also be used when we are having inequality constraint. So, let us start with the equality constraint case first. So, we want to now we are considering a minimization problem which is not really a big deal because any maximization problem can be converted into a minimization problem by taking either the negative or the reciprocal. So, we want to minimize this loss function f this objective function $f(x)$ such that we have all these k equality constraints are satisfied.

So, $h_1, h_2, \ldots, h_k$ these are all linear it could be non-linear also these are like all functions of the variables x note that x can again be like it need not be scalar it can be a vector valued variable also. and we want and this is the objective function. Now in this case we are like earlier when we are talking about linear programming we are talking about f and the h all of them being linear. But in this case we relax that requirement now they can be non-linear also. So, now what we do is in the method of Lagrange multipliers we define another alternative problem.

So, we basically what we do is all these constraints are absorbed into the objective function. Now, when we are absorbing we multiply them those functions or constraint functions with certain coefficients like this $\lambda_1, \lambda_2, \ldots, \lambda_k$ etcetera and these coefficients are known as the Lagrange multipliers. So, these are their values are not known to us, but we so we have to estimate them also along with the values of x. Once again remember that x is a vector or x can be a vector which is basically a collection of variables like $x_1, x_2, \ldots, x_n$. So, now we have an unconstrained minimization problem with additional variables.

So, like this is the total new objective function we can call it as the augmented objective function. So, we now try to minimize this, but there are no other constraints because the constraints have been absorbed into this. So, now we treat it as our unconstrained minimization problem and we solve it using either direct method analytical methods or using gradient descent. So, now what happens is so, what does that give us it gives us basically the stationary points of f with respect to all the variables that is the partial derivatives of f with respect to all the variables that will be equal to 0. So, that is like the problem which we are right now trying to solve.

So, now the like there is something known as a Lagrange multiplier theorem. So, what does that theorem tell us? The theorem basically tells us that if you are able to solve the minimization problem on this augmented objective function which also includes the these additional terms as well as the additional variables. So, note that just like in the previous case we had additional slack variables in this case also we have additional variables which you call as the Lagrange multipliers. So, we have to minimize this augmented function with respect to all of these parameters I mean the or all of these variables the initial variables as well as the new variables which are known as the Lagrangian multipliers. So, the what is the approach? The approach is to find the stationary points of this augmented objective function f with respect to like all the additional variable with respect to all the variables including the x's as well as all the lambdas.

And since it is a stationary point their derivative of a capital F with respect to all of these variables I mean their partial derivatives will all be equal to 0. So, that is what we are trying to find such stationary points. and the this Lagrangian multiplier theorem is telling me that these stationary points of these f they will also be the solutions to the original minimization problem small f. That is I just construct a new function combining the original objective function with all the constraints and I calculate the stationary points of this new function $f(x)$. So, remember that stationary function means what it means that the derivative disappears derivative equal to 0 with respect to all variables.

So, that can happen at maxima it can happen at minima it can also happen at the saddle points. So, we are not claiming that we have found a minima that is we have found a minima of capital F. We have only found a stationary point of capital F, but the Lagrangian multiplier theorem is telling me that that is enough. Using that solution you can find the value which minimizes the original problem $f(x)$ also. Now, you also involve these kinds of inequality constraints in the objective.

So, earlier we are having only these equality constraints like this. Now, let us say we also have the inequality constraints. there is a small error here the $g(x)$ will be less than equal to 0 those are the inequality constraints and $h(x)$ will be equal to 0 that is those are the equality constraints which we earlier also had. So, also there will be any number of inequality and equality constraints like in this case we had k number of equality constraints similarly we can have l number of inequality constraints also. So, now when we are constructing the augmented objective earlier also the what we did the same thing we do once again for the equality constraints we just like earlier we incorporate them into the like we that is we merge them with the original objective function which is $f(x)$.

We merge them by multiplying them with a Lagrangian multiplier and adding them. Now, in case of the other variables that is the $g(x)$ is less than or equal to 0 pardon the typo here. So, like those are also converted into equality constraints as earlier by introducing the slack variables. So, we so, like the slack variables is something which converts those equal inequality constraints into equality constraints and So those new equality constraints which we are obtained from the inequality constraints we like incorporate those also into the augmented objective function and corresponding to them also we have some Lagrange multipliers. So now we have two sets of Lagrange multipliers one with respect to the inequality constraints and another with respect to the equality constraints.

So, now what do we do? We once again have to find the stationary points of this capital F. Now we have so now we are the solution to the original problem that is minimization of small f is characterized by certain criteria. So, which are known as the KKT conditions. So, what are these conditions? So, first of all the gradient of f we should be equal to 0. So, that is the stationary point of this augmented Lagrangian which we had earlier also.

So, that is unchanged that is we have to calculate the derivative it must be such a point at which the derivative of capital $F(x)$ vanishes with respect to all the variables which are in question. Then secondly $h(x)$ should be equal to 0 and $g(x)$ should be less than or equal to 0 which basically means that all the constraints the equality constraints as well as the inequality constraints they are all satisfied. The third criteria that must be satisfied is known as the complementary slackness criteria. This is for the slack variables s_j which we have now considered. So, that is basically that the $\mu \cdot s$ should be equal to 0 that is the like this relation should hold.

What is μ? The μ are the Lagrangian multipliers corresponding to the equality constraint. what are the s_j's? s_j's are the slack variables for the corresponding inequality constraint. So if we the complementary slackness criteria is basically saying that like we if we multiply these pairs of things that is basically this thing and this thing and we add them up over like all the inequality constraints that were there we will get 0. So, this is an important criteria which is may not be that obvious, but this is important. So, this is a complementary slackness criteria and finally, the sign condition on the inequality multipliers.

So, the μ all these μs they must be non negative. So, any so like as already mentioned I have this $f(x)$ this is the augmented objective function. So, we find the stationary points of all these we do it by calculating the derivative with all of these variables I mean the original x variables the new the Lagrange multipliers lambda's the Lagrange multipliers μ 's as well as the slack variables. So, we find the those stationary points now among the stationary points the true solution that or those for which all of these things are satisfied. And what is the true solution? The true solution is such a solution which minimizes $f(x)$ such that all these constants are also satisfied.

So, this is how we solve with Lagrange multipliers. Apart from these there are also a set of methods which are known as the interior point methods and the exterior point methods. So, the basic idea of interior point method is you start with a point which is within the feasible region that is which satisfies all the constraints even though it need not be optimal with respect to the objective function. But, now what do we do we perturb the solution towards the optimal that is maybe by using something like gradient descent ok. So, that is we try to minimize or maximize the objective function with respect to the Now, whenever we move to a new from the current solution or the from the initial solution we move to a different solution with the hope of improving or minimizing our objective function. But the question is as we move to the new solution is it still feasible or have we somehow overshot the this feasible region.

So, we have to make sure that as we are whenever we are updating our position that is from the initial position which may be random to we are trying to move towards a solution which lies or which minimizes the objective function at every candidate solution we have to check whether the constraints are satisfied or not. So, that is of course, a tedious task suppose we find that the constraint is not satisfied then what will we do we cannot backtrack. So, the alternative is we introduce something known as a barrier function that is once again we are we look for an augmented objective function. So, we have our original objective function to that we add something called as the barrier function which takes a high value which is outside when for points outside the feasible region and low values for points which are inside the feasible region. So, that means, the presence of that barrier function will make sure that we at whichever point we are going we do not go outside the feasible region.

Why? Because now we are trying to like minimize both the this barrier function as well as the original objective function. So, that is we are considering a linear combination of them. So, if one of like if we somehow step out of the feasible region, then the objective then the barrier function will become very high. So, the augmented objective function that will also become very high.

Hence the algorithm will prevent me from going to such a place. So it will the algorithm will force me basically to stay within the feasible region so that none of the constraint is not violated. So this is basically what the procedure looks like. So you set an initial tolerance value epsilon and a decrease factor β as well as an interior starting point x_l which satisfies all the constraints and an initial μ. So, like we will understand this later. So, now we formulate this augmented problem as I already said $f(x)+\mu B(x)$ where μ is this factor and $B(x)$ is the barrier function which I mentioned.

So, now use x_i as the starting point and let the optimal solution be x_{i+1} that is this is the thing this is augment instead of minimizing $f(x)$ you minimize this thing this new thing to get a candidate solution x_{i+1}. Now if this happens if this condition is satisfied then you stop else you update the μ. So, remember that β is a decreasing factor so as I multiply so μ decreases. Now what is the point of decreasing μ or what is the point of μ at all μ is basically telling you how important the barrier function it is to maintain the barrier function. It is of course very important to like the barrier function is very important because in case I overshoot the feasible region then it is the barrier function which alerts me which does not let me overshoot.

So, like I so high value of μ basically means I am giving a lot of importance to the barrier function. that is I am like if I try to escape the feasible region this part will be because μ is already high. So, this part will become really high and like it will not like the solution will not be accepted as it will not be able to minimize this augmented function. So, like so I should start with a high value of μ, but if I start with too high value of μ that means, I will be always be present in the this thing the feasible region that is true, but I may not be minimizing the $f(x)$.

So, once it is ensured that I am inside the feasible region. I can actually reduce this μ a little bit I can give a little less importance to the barrier function and a bit more accordingly a bit more importance to the original objective function f x. So, that I can like achieve the best of both worlds that is I am inside the object the feasible region as well as I am not as well as I am trying to find the I am trying to find as low value of f as possible. So, an example of this is given I am not going into the details of the example, but you will see that the this μ is decreasing step by step. And you will also see that this term this is also decreasing step by step that means what that means that. I am like I am gradually approaching the minima and the final convergence is obtained at this stage 0 where x has a certain value.

So it can be seen that like if you go back to the original problem then you will understand that these values they indicate that the constraints are all satisfied and that is where the minima has been obtained. The other approach is just the opposite of it that is the exterior point methods. Here the idea is you basically start at an initial point which is outside the feasible region, then optimize with respect to the objective function plus the constraints, then you gradually increase the constraint weights till they are satisfied and you this automatically pushes the solution towards the feasible region. So, this is like this is called an exterior point method because here you are the original value with at which you are starting the original solution at which you are starting it indeed minimizes f x even though it may be outside the feasible region. So, now you try at every stage you now try bring that it closer to thefeasible region like with the help by gradually increasing theequivalent of the barrier function.

Earlier we had the barrier function here we do not call it as a barrier function, but we like call it as a like as the constant, but we are in this case we call it as a penalty function. So, when we are outside the like exterior point that is since we are considering exterior So, like it is we are out that is we are violating some of the constraints we are outside the feasible state that means this is quite high. Now, we try to progressively reduce the violation by bringing the solution closer and closer to the feasible region. So, that is what is known as the exterior point method. So, this is the algorithm which unfortunately I will not be able to explain in detail due to the lack of time, but I hope you will be able to understand it especially when you see this numerical example.

So, here also you will see the only difference will be that in this case we will see that we are I mean like that is as we are approaching the optimization like the like we are reducing the sorry we are increasing the μthat is we are increasing the importance of the penalty function and as a result this penalty function is gradually increasing. Unlike in like in the previous case where the like where the aim was to decrease the μin this case it is important to increase the μ. So, like basically we discussed to like several approaches of constraint optimization. So, we started off with the simple problem of linear programming where all the equations and constraints are linear in nature and we saw that it can be solved with the help of this fundamental theorem of linear programming.

There are also efficient algorithms like the simplex algorithm which we did not study here. Apart from this we can also have more general settings where they are non-linear and for such cases we can take the help of Lagrange multipliers. Alternatively we can utilize the interior or exterior point methods. So, with that we come to this at the end of this topic on optimization. So, in the next lecture we will deal with some other topics. So, till then everyone please take care and stay well. See you soon. Bye.