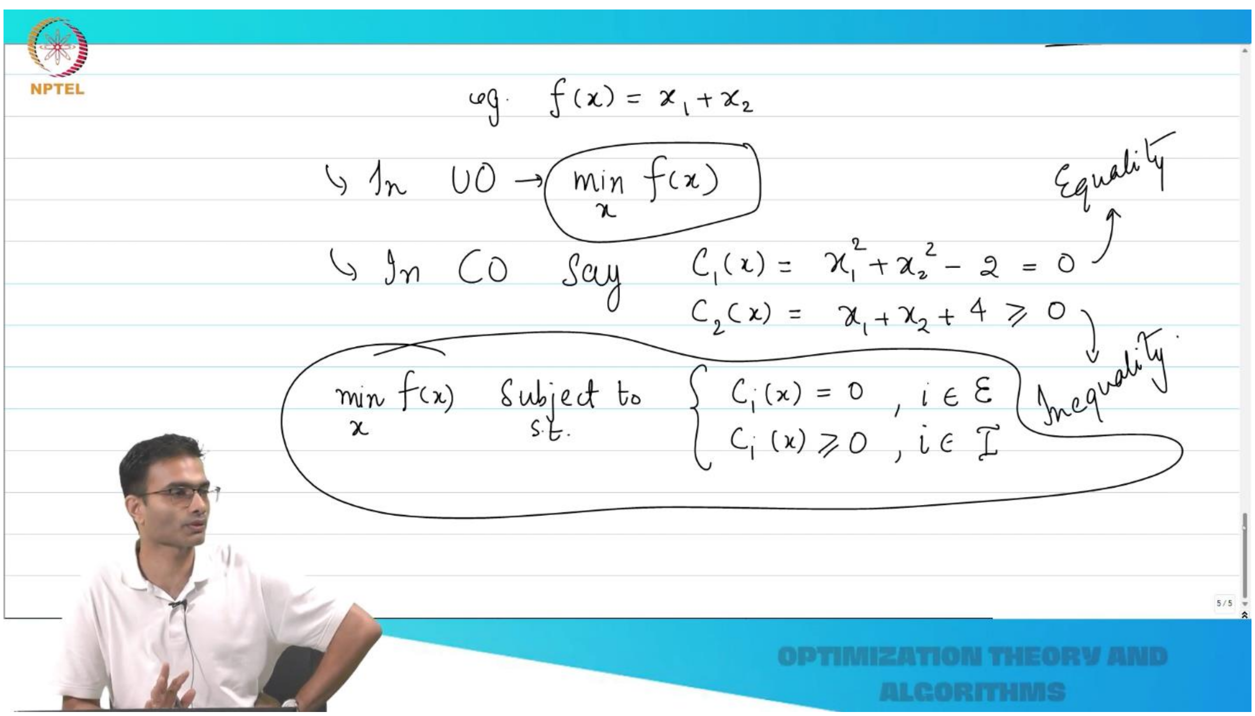
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 09
Lecture - 60

Constrained Optimization

Alright, so we are going to move to the next part of this course, the final leg of this course, which is a very important part: constrained optimization. So, for those of you following the book, this is roughly Chapter 12 of Nocedal Wright. Now, in talking about constrained optimization, I am going to take some extremely simple examples and they are going to be two-dimensional, and the advantage of doing that is that we can do some geometry. When we can visualize what is going on, then we will be able to apply it to, you know, more difficult problems.

So, let us start with something very simple. Let us say my objective function is $f(x) = x_1 + x_2$. Now, you will see that there are some very subtle differences between constrained and unconstrained optimization problems, which make no sense unconstrained, suddenly will begin to make sense in a constrained setting.

For example, if I give you a cost function like this, $f(x) = x_1 + x_2$, and I said minimize this, what would you say? That is crazy, right? Because it makes no sense to optimize something that is going down to minus infinity. However, in the constrained world, this may make sense. So, let us have a look at that, how.



So, let us call this C O, ok.

So, in the unconstrained (U) is unconstrained optimization, right? So, the simply the way we wrote it was. This is how the notation for the problem is, right? Unconstrained optimization I write it like this. In the constrained optimization world, I have to obviously introduce some constraints for you, right? So, let us introduce two constraints. So, I am going to say, I am going to give you two constraints, and for variety, what we will do is one constraint will be equality, and one constraint will be inequality.

Let us see what happens. So,

$$x_1^2 + x_2^2 - 2 = 0.$$

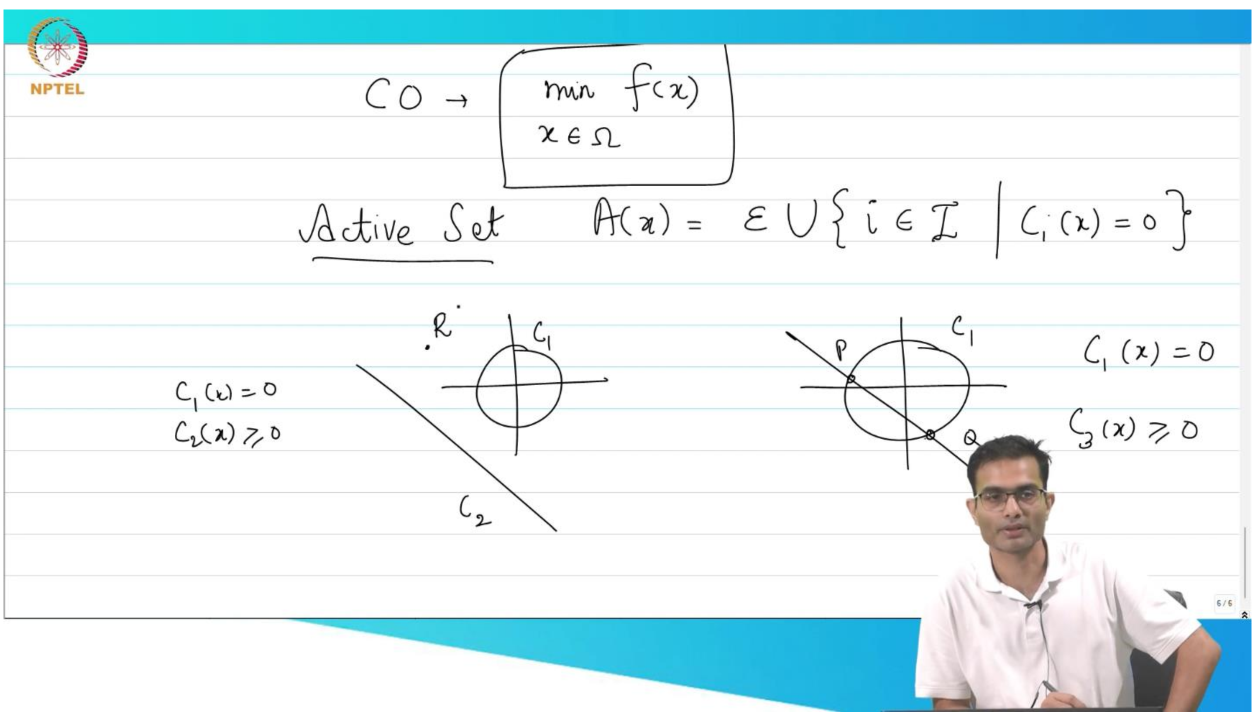
This is a constraint. Geometrically, what kind of a constraint is this? A circle, right? $x_1 x_2$ must live on the circle. Let me make matters more interesting and give you one more constraint:

$$x_1 + x_2 \geq -4.$$

This is what kind of a constraint? So, very roughly speaking, I have an equality constraint, and I have an inequality constraint. So, how do I write this now? So, I could write this as one way of writing it is

$$\min f(x), \text{ subject to constraints.}$$

So, many people will write it as S.T., which simply means "subject to," and your constraints come in basically two flavors: those two flavors—unlike Baskin Robbins—there are only two flavors here: equality and inequality. There is nothing else that you can really think of. So,



$$C_i(x) = 0,\ i \in E \quad \text{(equalities)} \quad \text{and} \quad C_i(x) \geq 0,\ i \in I \quad \text{(inequalities).}$$

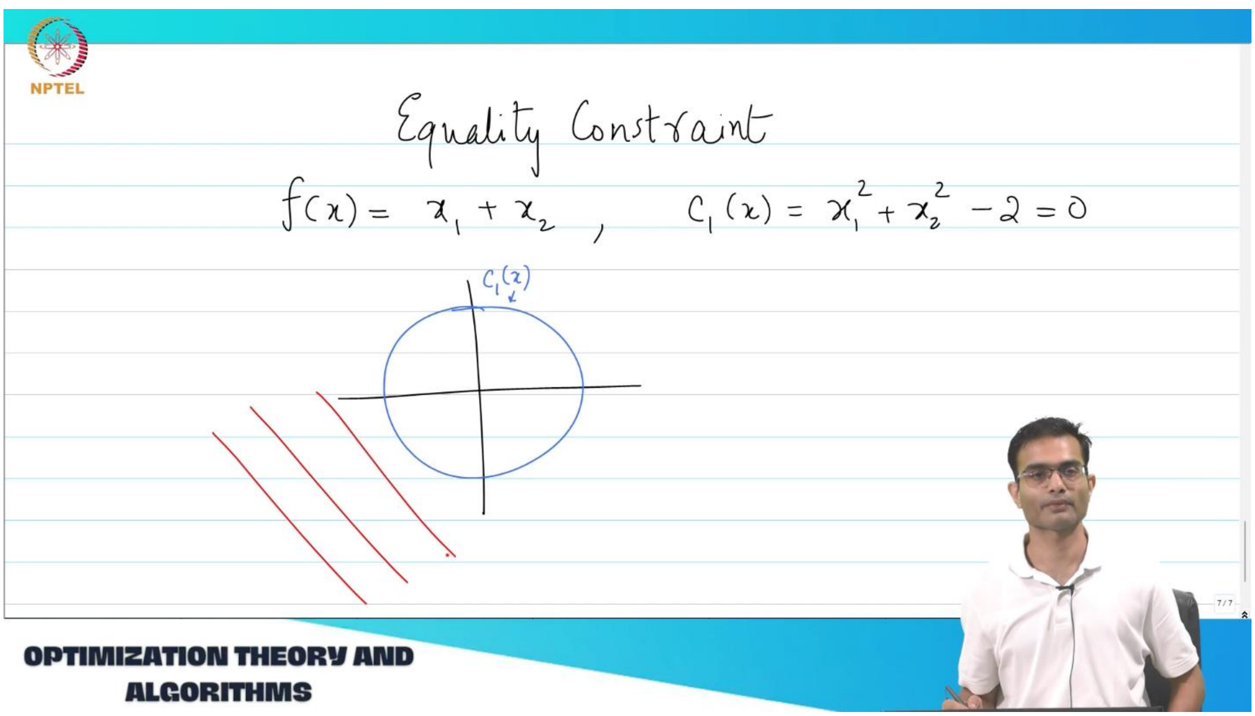
So, this is how your unconstrained optimization was a simple problem statement like this. Constrained optimization now, I have to write this whole thing, this is my problem statement over here.

Is the notation clear to everyone? $C_i(x) = 0$ for $i \in E$ for example, could be 1, 2, 3, that means C_1, C_2, C_3 are equalities, and $C_i(x) \geq 0$ for $i \in I$, where I could be 4, 5, 6, and so on. So, C_4, C_5, C_6 are inequalities. This is a simple way of just defining the problem.

Now, one thing I want to caution you against, which is a cause of a lot of confusion when you go from book A to book B to paper C. Every author seems to have a different convention of what an inequality is. So, half the literature will have $C_i(x) \geq 0$, the other half has $C_i(x) \leq 0$. So, everything will be off by a minus sign when you are trying to make sense of it. In the book that we are following and the notes that we are going to do, we are going to assume this form: $C_i(x) \geq 0$.

So, keep this in mind when you are reading the literature. If you go back to the very beginning of the problem definition, make sure which way their inequalities are going because otherwise, you will be off by a minus sign everywhere.

So, let us just make one note over here: convention. So, having written this, this looks like a very, like a clunky big way of writing a problem statement. We want to have a little bit more compact way of writing it. So, the equalities and inequalities together, they define a very common-sense word. They define a feasible set.



Means those, it defines those sets of x which are legal, right? That's also called feasible. So, the feasible set, the usual notation is Ω, right? It will simply be like this:

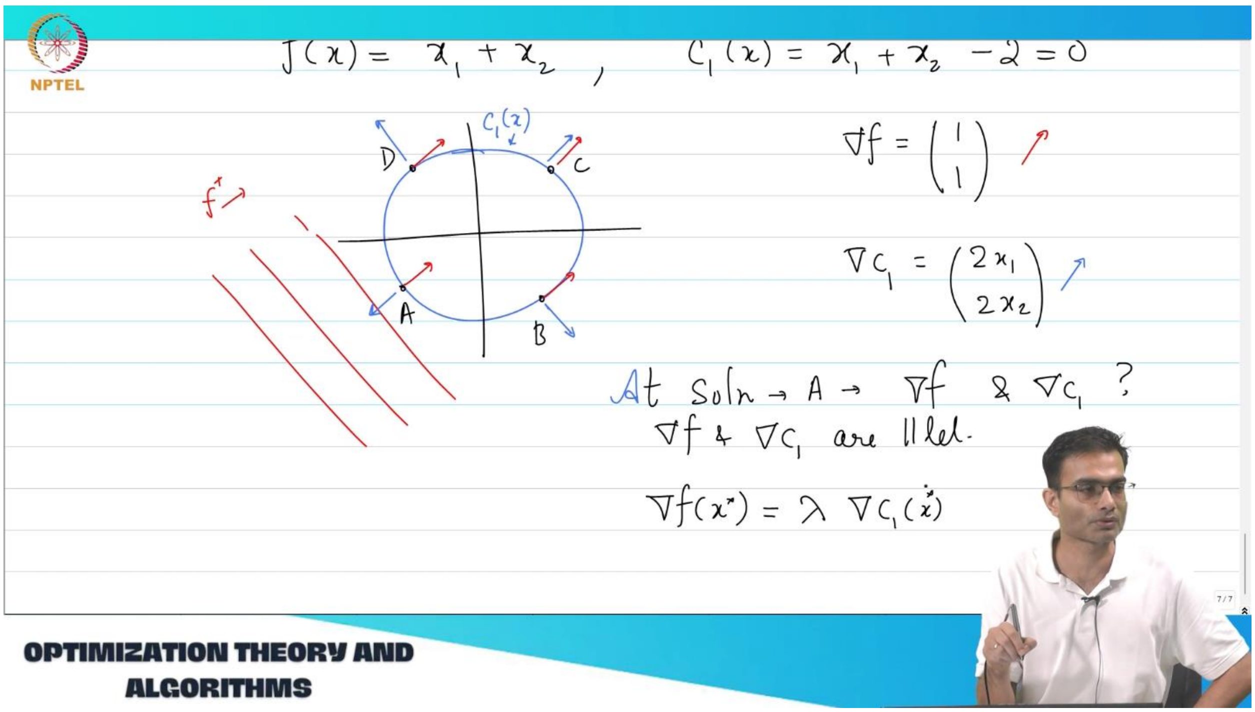
$$x \quad \text{such that} \quad C_i(x) = 0,\ i \in E \quad \text{and} \quad C_i(x) \geq 0,\ i \in I.$$

So this is your feasible set. Is it clear what I mean by feasible set? All those sets of x which satisfy the equalities and inequalities?

Having done this, the optimization problem simply becomes, so constrained optimization simply becomes:

$$\min_{x \in \Omega} f(x).$$

So, it is a compact way of just defining the problem. Earlier I just wrote x under the minimization. Now I am saying x belongs to a feasible set. This is the more common notation that you will find in research papers.



Everyone with me so far? So far we have just defined equalities and inequalities.

So, let us just now that we have done this, let us look at this constraint a little bit. These couple of constraints that we have got, let us sketch them. The first constraint is an equality constraint $C_1(x) = 0$, which we said is the locus of a circle centered at the origin, radius $\sqrt{2}$. So, I have this is my circle. So, what is this point for example? $(\sqrt{2}, 0)$, right? $(-\sqrt{2}, 0)$, this is $(0, -\sqrt{2})$, alright.

So, this is my $C_1(x)$. Where is my $C_2(x)$? What is the slope of this line $C_2(x)$? $-45°$. What is the intercept? -4, right? So, it is in the bottom left, it is like this. Right. So, this is my $(0, -4)$, and this point over here is my $(-4, 0)$, okay.

Very good. Now, what is this feasible set now? Is the entire circle, the entire circle is legal because it is, you know, it is quite obvious that this full region, the dashed green region, satisfies $C_2(x)$, and this circle itself satisfies both $C_2(x)$ and $C_1(x)$. So, if someone says, "What is the feasible region, or rather, the feasible set for this problem?" It is the circle, not the green region. The circle is the feasible set.

Now, on this feasible set, does it make sense for me to ask this question: Optimize $f(x) = x_1 + x_2$? It makes sense because the answer need not be minus infinity, right? So, even such a simple problem like this, the moment I start adding constraints into it, they will begin to have some solutions, which in the unconstrained world, this simply did not make sense.

Alright, so that is just a little bit of intuition over here. In terms of notation, I have told you what the equality and inequality constraints are. There is one more definition that gets used a lot in describing constrained optimization: it is called the active set.

So, the way we define the active set is very simple: it is those equalities, which are automatic, and all equalities are included in the active set. Those inequalities which are, what, when I say, "What let us?" We can arrive at this just by common sense. If I say an inequality is active, what is the first thing that comes to your mind? We are talking plain English here. An inequality is active, so let us go back to our example over here. So, in this one, which is the inequality? C_1 or C_2? C_2 is the inequality.

Supposing I am standing at this point, would you say the inequality is active? If I was standing at this point over here, would you say the inequality is active? Okay, so maybe it is not that intuitive. The meaning of active is that the inequality becomes an equality; remember, there is a "greater than or equal to," so those points where it becomes equal to 0 are where the inequality is said to be active.

So, this character, this point over here, is where the inequality is active. That's all, okay? So this is the definition of the active set. Could have used some other words also. So, inequalities are always active because they are always enforced.

It is an equality, so you have to satisfy it. The union with those inequalities, which are, quote-unquote, "active." So, i belongs to the inequality set, which means $C_i(x) \geq 0$, but for those points—so remember that this over here, the active set definition—does it depend on x? It depends very much on x; it is a question being asked at a particular point x. At this point, which is the set of inequalities which are active? Active, in some sense, in plain English, means it is making a difference, right?

If you look at—okay, let us just quickly draw this once more.

This was my C_1, okay, and this is my C_2, and another case is this is my C_1, this is my C_3, okay. Let us call this point P, this point Q, okay, so my $C_1(x) = 0$ is my equality constraint. $C_2(x) \geq 0$ is my second inequality over here. Once again, I have $C_1(x) = 0$ and I have $C_3(x) \geq 0$.

Okay, so if I were standing at a point R over here, is R even a point in the feasible set? It is not a point in the feasible set. So, I should not even be asking a question. Supposing I were standing over here, at this point over here, all I can say is that the inequality is active. The inequality is active because $C_2(x) = 0$ holds true at this point.

So, the inequality is active over here. If I was standing at this point on the circle, am I in the feasible set? I am in the feasible set. What is the active set at this point over here R'? Only $C_1(x)$, because only the equality always has to be there, right? So, $C_1(x)$ is there. So, the active set has the digit 1 inside it, okay. $C_2(x)$ is strictly greater than 0 at this point. At this point, $C_2(x)$ is greater than 0. It is not equal to 0, so therefore the inequality is not active.

That is the meaning of the active set.

Now, on the other hand, let us look at the second, the right-hand side example. My constraint is $C_3(x) \geq 0$, along with the previous one $C_1(x) = 0$. So, what is the feasible set now? The arc of the larger arc of the circle, right? So, this is the feasible set—not the full circle. If I ask, supposing I was, let us call this point Q over here, at the point Q, what is the active set? 1, right? $C_1(x)$.

Why? Because the equality is always there by definition, but inequality is not included in this because why? At the point Q, what is the value of $C_3(x)$? 0 or greater than 0? Greater than 0. Therefore, it is not included in the active set.

On the other hand, if I was standing at P or Q, $C_3(x)$ takes on the value of 0, right? So, if I ask you, "What is the active set at P or Q?" You would say 1, 3: 1 for C_1, 3 for C_3, shorthand, okay.

So, there are a lot of algorithms that make this distinction between active and inactive inequalities. The rough intuition is this. Supposing I am standing at point Q. Now, remember the whole game in optimization, in continuous-valued optimization: if someone who has not taken any optimization were to ask you, "What is the essential tool that you use in optimization?" What would you say? Someone who has not taken optimization asks you, "What is the fundamental tool that you use in optimization?" Whether considered or unconstrained, what would you say? Something very simple.

Anyone should be able to understand. I don't want such a technical answer. Is it not saying simply calculus? Calculus is your tool. Why? Because you work with a gradient, you work with a Hessian. and you move according to that that is the fundamental tool right if someone was doing SVD what the fundamental tool is linear algebra SVD here calculus is what allows us to move from a good point to a better point right the train that takes us from a good to a better point is calculus. I am taking first order Taylor's theorem, second order Taylor theorem, mean value theorem all of this is built on calculus right so at point Q I am sitting at point Q, do I have to worry about $C_3 x \geq 0$? Remember in calculus what is the, I use first order theorem, a Taylor's theorem, I use second order Taylor's theorem, what is the real distinction? I mean when can I use one and not the other? These are all theorems built for a small deviation about a point.

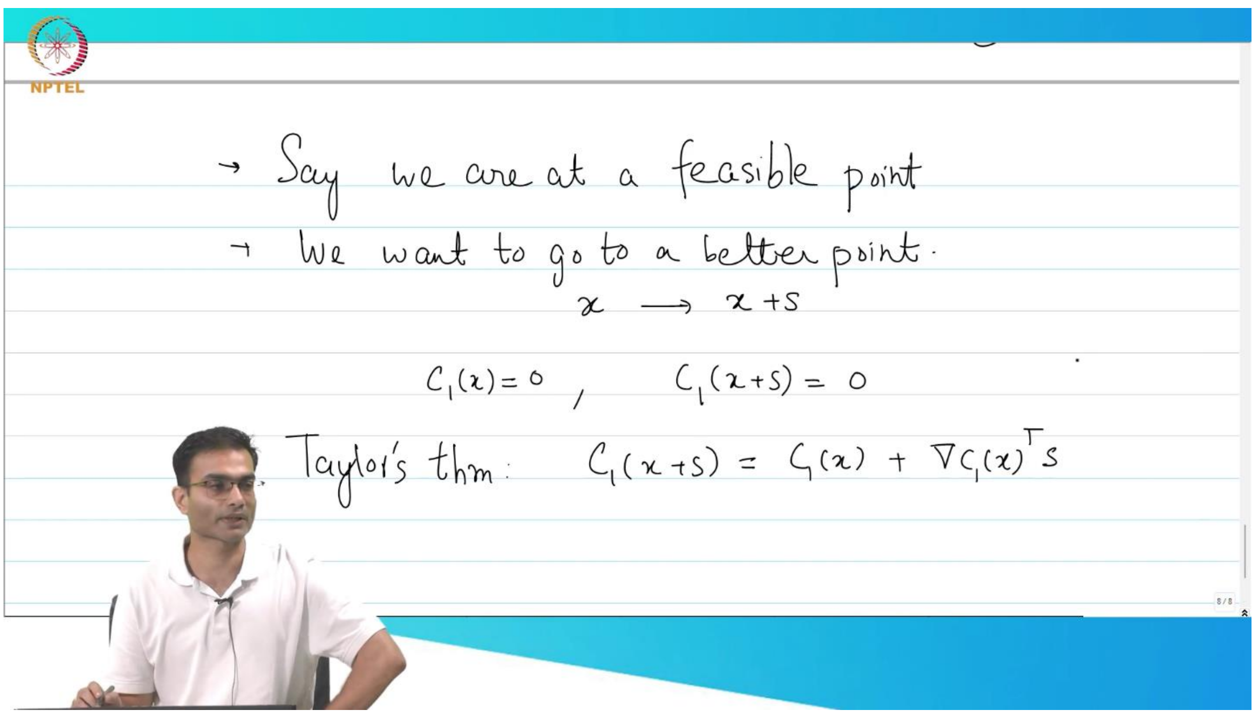
They are not applicable globally. If I want to write it globally, I have to take many, many terms of the Taylor theorem. But if I am taking only first order Taylor theorem, it's valid in a small neighborhood around the point. If I take second order Taylor theorem, it's valid in a little bit bigger area. That's the intuition behind first order, second order. Now, if I am standing at point Q and I am going to do first order, for example, gradient descent.

Second order, for example, Newton method. Those are the two things that I have used. If I am in a small neighborhood like this about Q, do I care about C_3? Right. So, in a small neighborhood about C_3, I am never going to care about Q. I am never going to get into trouble with C_3. That is the motivation behind saying C_3 is inactive.

I need not worry about. On the other hand, if I am sitting at P or Q, I do not have that luxury. I have to make sure that $C_3x \geq 0$ is satisfied no matter what infinitesimal Δx I move. So, that is why I say C_3x is active. So that is why in an algorithm when I am deciding how to move from a good point to a better point I need to do some you know keep check of what is good what is bad and that checking is happening by defining this active set. So while it looks like just some definition it is actually very useful when it comes down to the level of algorithm okay.

So this was about so far we have spoken about definitions, feasible set, active set these are the two new things that are sort of entered into the picture right. So now let us try to solve our first constraint optimization problem to make matters simple we will start with one constraint and even simpler which is simplest kind of constraint and equality constraint okay. What if C_3x is yeah then in that case that inequality would never be active by definition because it is greater than equal to I mean it is greater strictly greater than 0. So, it would never be active. However, it is unusual to find such a specification because it is not very well defined.

Right if it is C_3 is greater than 0 supposing that is my constraint. Now how close I get to C_3 without actually hitting C_3 is undefined I can get 10^{-10} away from it right. So, it causes a kind of a not a very well specified problem. So, what you will find in the literature is always greater than equal to kind of a problem. So, let us take our previous cost function simple $x_1 + x_2$ two dimensional C_1x is equal to $x_1^2 + x_2^2 - 2 = 0$.



So, I have an objective function and I have an equality constraint. Here is my C_1 ok. So, I mean this is one of those problems that you can solve graphically if it is very simple to solve. What are the contours of constant cost function value? What would they look like? Remember for ok, let us do a quick revision. For a quadratic form with a positive definite matrix A, what were the contours of constant cost function ellipses? For this function, what are the contours of constant cost function? Parallel lines.

Parallel lines. What are the slope of those parallel lines? $-45°$. $-45°$ right. So, if I draw lines like this, is that correct right? Which way is the cost function increasing? To the bottom left or top right right. So, f is increasing this way right if I go like this.

That is actually enough for us to solve the problem. What is the solution? Is there a solution to this problem? Yes. What is that point? The tangent to the circle at the in the third quadrant that is going to be the solution. So, I already know the solution to this problem without the constraint as I said earlier this problem makes no sense, but now it makes sense right ok. So, that is the reason behind choosing a simple problem like this that we know graphically what the solution is ok. So, let us still let us do some calculus over here to get intuition.

Now let us try to write down a few things that we will use throughout right calculus is what is the tool that we are using. What is ∇f scalar vector matrix vector size 2×1 why because $\frac{df}{dx_1}, \frac{df}{dx_2}$ right. So, $\frac{df}{dx_1}, \frac{df}{dx_2}$ what is its value? $f(x) = x_1 + x_2$, derivative is what? What is ∇C_1? $2x_1, 2x_2$. So let us see this first. Let us identify a few typical points on this graph that are interesting ok. So, I have taken a few points over here A, B, C, D ok.

I have taken these points over here. Let us denote ∇ by this color and ∇C_1 by this color ok. So, at all of these points, so ∇f is a constant, it is a constant vector. So, no matter which point I take on the feasible set, ∇f is always pointing in the same direction. So let us just draw that over here ok. Which way is ∇C_1 pointing? Inward, outward? Any relation to the circle? It is always radially outwards, you can check it right at any point if I take it, it is always pointing radially outwards.

So, at this point, it is pointing out radially outwards at point A which way is it pointing? Radially outwards backwards, right? It is pointing over here. Similarly, over here it is pointing here and over here it is pointing over here ok. So, at the solution, we know that the solution is out of A, B, C, D which is the solution point A right at the solution which is point A. Is there any relation between ∇f and ∇C_1? They are opposite to each other in general. What can you say? They are parallel to each other right whether it is pointing in the same way or with a minus sign they are parallel to each other right. So, ∇f and ∇C_1 are parallel ok. So, I could also write $\nabla f(x^*)$ which is the solution as $\lambda \cdot \nabla C_1(x^*)$.

So, I we have not derived this, this is just based on observation, you will agree with me ok. What is the trouble with this way of identifying the solution x^*? Is there any other point also where this holds true? It also holds true at C also, is true at point C ok. So, that is a bit of a bummer because we did not get it directly ok. So, let us say that we are at a feasible point. Let us assume let us make one very simple assumption that no matter what algorithm I use I always stay in the feasible. Let us assume that it is not allowed for you to jump out into you know the no man's land you are always on the feasible side.

So, let us say that we are at a feasible point and I want to go to a better point, that is the game that we are trying to play right, that is how it would be pick a random point, get better ok. So, we are here we want to go to point ok. So, I am at let us say x I want to go to $x + s$ which is better. Now, better would mean what? Value of $f(x+s)$ is less than $f(x)$ that is what is meant by better ok. Since I said that the game the rules of the game are that you have to stay on the feasible set right.

What does that mean for $C_1(x)$? What does that mean for $C_1(x+s)$? What are their values? They have to be 0 because they are I am only operating on feasible points this much is true ok. Now comes our favorite Taylor's theorem right. Taylor's theorem if I apply it to the function $C_1(x)$ can I do it? $C_1(x)$ is some function right $x_1^2 + x_2^2 - 2$, it is a function I can apply Taylor's theorem to it. So, $C_1(x+s)$, can you tell me what will it be if I apply Taylor's theorem? Remember the rules of calculus which I just told you, I want to make a small shift.

So, s is a small number comparatively. So, what would I write $C_1(x+s)$ as? First term $C_1(x)$, second term $\nabla C_1(x)^T s$. This is my first order Taylor's theorem. Given the rules of the game that I am always going to a feasible point, what does this first order Taylor's theorem imply? It is giving me some constraint on s. It is saying that $\nabla C_1(x)^T s = 0$.