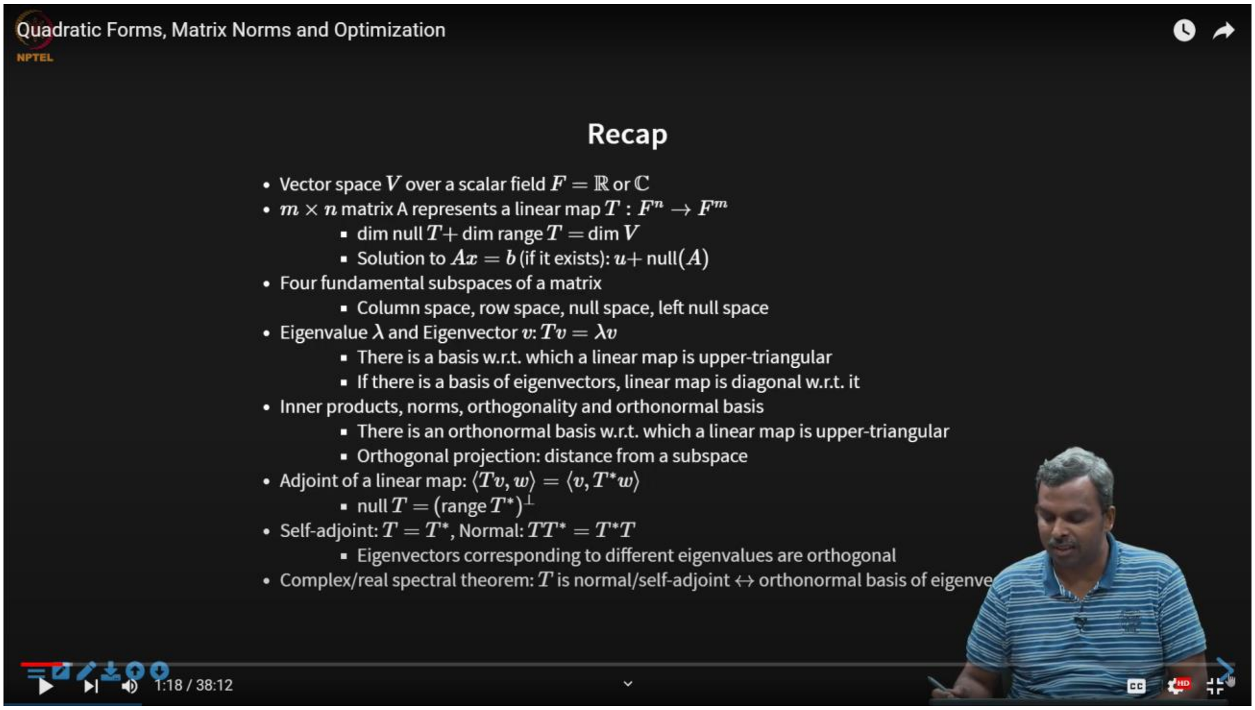
Applied Linear Algebra
Prof. Andrew Thangaraj
Department of Electrical Engineering
Indian Institute of Technology, Madras

Week 11
Quadratic Forms, Matrix Norms and Optimization

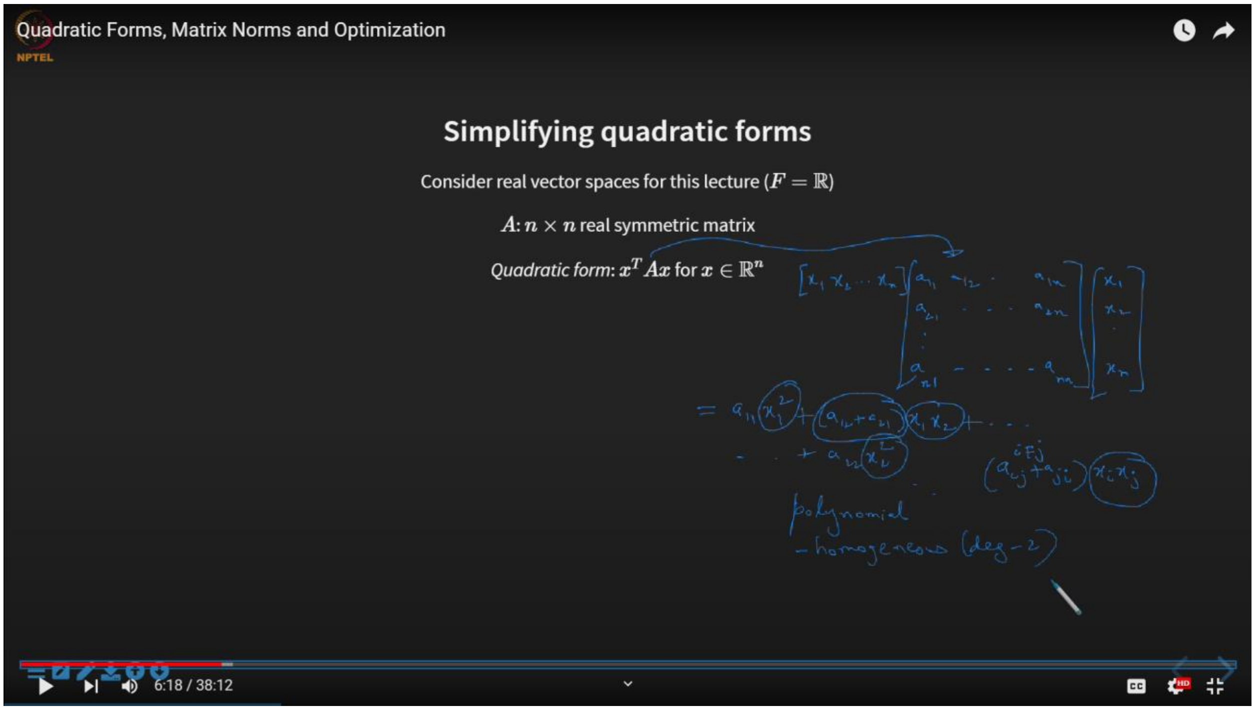
Hello and welcome to this lecture. This lecture is primarily going to talk about two or three different ideas, mostly sort of in maximization-minimization, optimization sort of framework in relation to matrices, quadratic forms, you know, matrix norms and other just pure multivariate optimization problems. And it's very surprising how the simple idea of, you know, positive operators, eigenvalues, real spectral theorem, all these things play an important role in giving us a sense of, you know, the maximization-minimization in these kind of forms. So you will see it's got lots of interesting applications, and this lecture will walk you through that mainly. So let us get started. So I will skip the recap for this lecture.

(Refer Slide Time: 01:18)



I think, so in the previous lecture we were looking at positive operators which basically are self-adjoint operators which have, you know, non-negative eigenvalues, right? So that gives you a sense of what positive operators are. We saw many of their interesting properties. We'll put some of them to use, at least show some applications of them in this lecture and also applications for the real spectral theorem. I think it's really more applications of the spectral theorem than anything else. I mean, if you look at positive operators, really the spectral theorem is everything in that, right? So the fact that it's self-adjoint and its eigenvalues are non-negative lets you define the square root, and I mean the whole characterization is revolving around that, okay? So you can think of this as applications of the real spectral theorem.

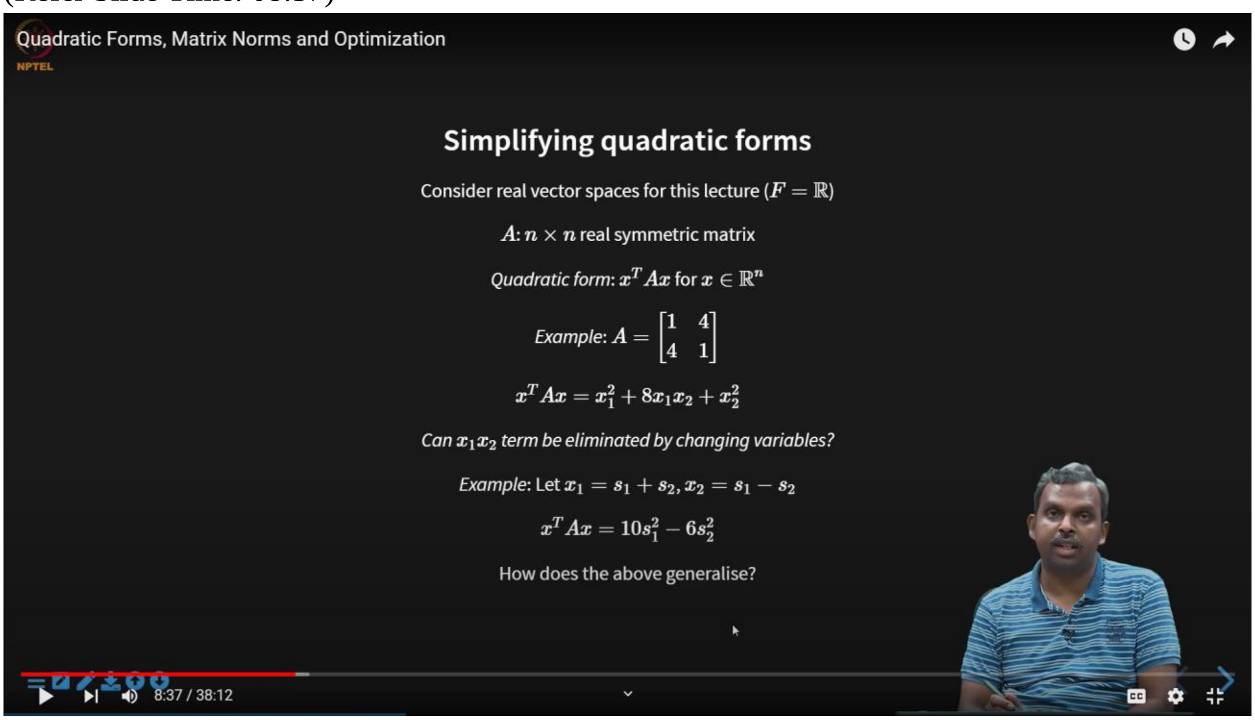
(Refer Slide Time: 06:18)



Okay. So the first thing we're going to look at is, we're going to look at quadratic forms in some detail. These things came up even in the previous lecture. I was in the context of introducing positive operators, right? So we saw how the quadratic forms really, I mean, play a crucial role in the definition of positive operators. The quadratic form is non-negative means the matrix defining that is positive, right? So that's the way we defined it. So now we're going to specialize to real vector spaces. I mean, it's not because things are not easy in the complex vector space, it's the same thing except that just it's easier for me to discuss and present the ideas in real space. So I will start with an $n \times n$ real symmetric matrix and I'll look at the quadratic form $x^T Ax$, okay? So this has a certain form. I don't know if you realized it, I think I didn't really spend some time thinking about it. So in general, if you want to write, say, $[x_1 \ \ x_2 \ \ldots \ x_n] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, right? So this is the quadratic form, isn't it? So this quadratic form is basically this. So in case you are wondering why people say quadratic and all that, you will see it is basically a degree 2 function. So if you look at, if you multiply this out, you are going to get $(x_1, x_2, \ldots, x_n)$ and then you will get a column vector here. The first entry will be $(a_{11}x_1 + a_{12}x_2 + \ldots)$, like that. So you can do this multiplication first, and then you can multiply by this, right? With x_1 you can multiply, okay? So if you look at it, you will get $a_{11}x_1^2$, right? When this multiplies. Plus, remember, this is symmetric, so a_{12} and a_{21} are the same. But, I mean, for just comfort, I'll keep it like this. You will get $(a_{12} + a_{21})x_1x_2$, right? So you will actually get $a_{12}x_1x_2$ and a_{21} will also give you x_2x_1. So I am just writing it like that so you will get terms of this form, okay? So I am not going to write everything. So you will get a term like, you know, a_{22} will be what? x_2^2, right? So you will get, and then generally you will be getting this $a_{ij} + a_{ji}$, okay? If $i \neq j$, you will get this plus this x_ix_j, right? So basically you get terms of this form, right? These are degree two terms or quadratic terms, right? x_i^2 or x_ix_j, you get terms like that and then you get a linear combination. So that is why this is called a quadratic form. I didn't quite emphasize this, I thought you could do it yourself. But this is the reason why these kind of forms that are written succinctly, x^TAx, these are called quadratic forms for that reason, okay? So you have to keep this expression in mind. So generally quadratic form will be a long series, long polynomial, multivariate polynomial of, you know… It's a homogeneous polynomial. What's homogeneous? Every term has got degree two, right? So degree two, this is degree two, this is degree two, every term has degree two. And, you know, it could be x_1^2, x_1x_2, x_1x_3, like that. And the coefficients are multiplying that and adding it up. So it's a polynomial, so it's a polynomial which is homogeneous degree two, right? And coefficients come from the matrix, right? Suitably they come. So this is how this quadratic form looks, okay? So if you take small n, you can write explicitly how it will look, right? If you take $n = 2$, it will just be, you know, it will have x_1^2, x_2^2 and x_1x_2, that's it, okay? And larger n you'll have more and more and more terms, okay? So this is how, this is the picture you should have in your mind when you think of a quadratic term, quadratic form. It's a form like this, but it also has this homogeneous degree two polynomial expansion in this form, okay? So both are okay. If I give you the polynomial, you can go to the matrix, right? How do you go to the matrix given a polynomial, right? You can always go to a symmetric matrix, no? You can take this coefficient, divide by 2, put it in a_{ij} and a_{ji}, you will get, you can go back to the matrix, okay? So this is some basic thing about quadratic polynomials, quadratic forms. I didn't quite describe this in the previous lecture. I thought I should do it here, okay? Hopefully this is clear, okay? So this is how a quadratic form looks. Okay. So let us move on.

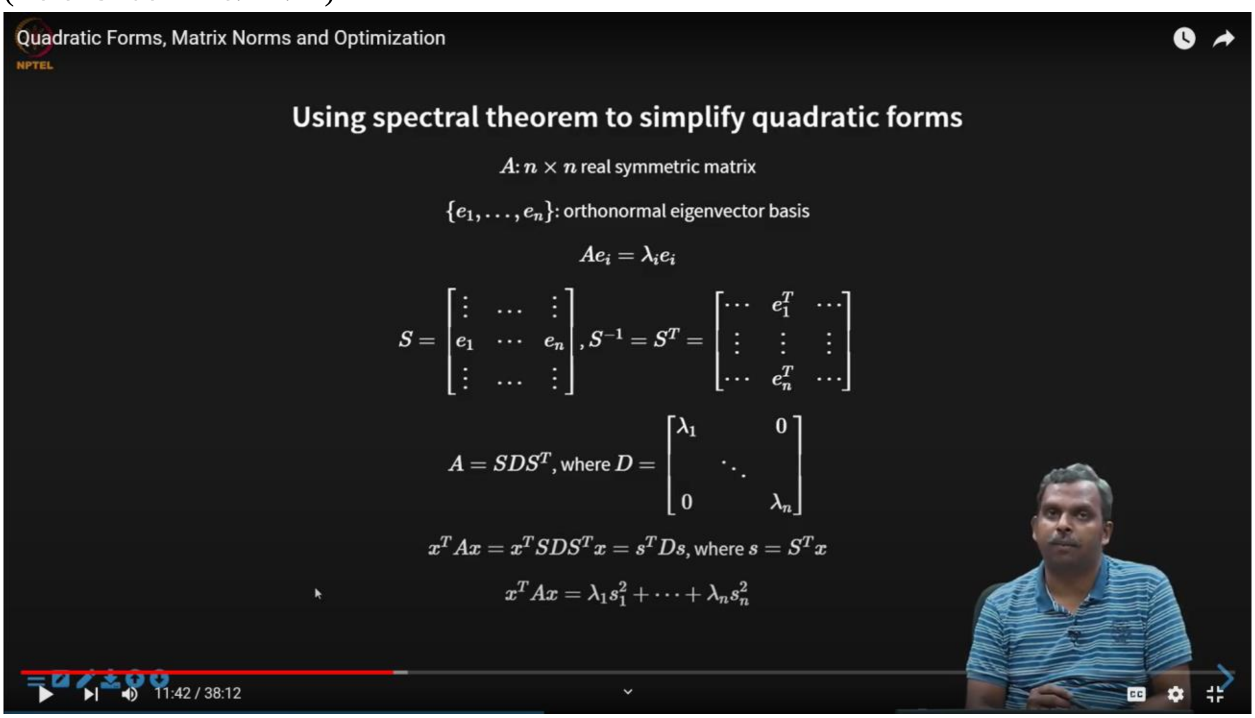
So here is an example, okay? I have a very simple example for $n = 2$. $\begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix}$. You write down the same expression and add it up, you will get $x_1^2 + 8x_1x_2 + x_2^2$, okay? So now when you think of how to plot this, like, for instance, if you want to plot this quadratic form as a function of x_1 and x_2, it will actually be a 3 dimensional plot, right? And $x_1^2 + x_2^2$ is very easy, right? You can sort of imagine how it will be. But this $8x_1x_2$ is a little bit confusing. So in general, in this polynomial, in a quadratic homogeneous polynomial, these cross terms, x_1x_2 terms slightly complicate the picture. If it was only, if these cross terms are not there, $x_1^2 + x_2^2$ is considerably simpler than dealing with the cross terms, okay? So this is one constant thing about quadratic forms. You like quadratic forms without cross terms better because you can deal with them easier. It is easy to see that $x_1^2 + x_2^2$ will be positive everywhere. You know it will be 0 at 0 and then after that it will increase. So you can think of all very easy ways of characterizing $x_1^2 + x_2^2$, right? So if you say $x_1^2 + x_2^2$ is a constant, you know what you will get, right? You will get a, you will get a circle, you know? Everything is easy to characterize when you don't have the cross term. When you have the cross term, things become complicated, okay? So one of the goals that people do is: can you do a change of variables, some linear change of variables to eliminate the x_1x_2, okay? So that's sort of like diagonalization, right? You can see how that is similar to diagonalization. If this were diagonal, if the matrix were diagonal, you will only have the square terms, you won't have the cross terms. So you have a symmetric matrix, you want to make it diagonal, you can do a change of variable. So this is where everything comes together, what we've been studying with symmetric matrices and diagonalization and quadratic forms and simplifying them by eliminating the cross variables, okay? So here is a change of variables. This is sort of inspired by the diagonalization. But, you know, if I didn't tell you that, it will look like magic, you know? If you say x_1 is $s_1 + s_2$, x_2 is $s_1 - s_2$, you see the cross term disappears, okay? It will disappear and you will get a $10s_1^2 - 6s_2^2$. So the question of course is how does the above generalize? And the generalization is through the real spectral theorem and diagonalization, okay? So you can sort of expect that and that is what will happen, okay?

(Refer Slide Time: 08:37)



So this is how you use the spectral theorem to simplify quadratic forms. Supposing somebody gives you an arbitrary $n \times n$ real symmetric matrix. It has off-diagonal elements, so the quadratic form has, you know, cross terms, $x_i x_j$ terms, quite a few of them, okay? So you start with your orthonormal eigenvector basis. Maybe you write coordinates for these basis vectors with respect to the standard basis, right? So we will always think of it like that. And we know Ae_i is $\lambda_i e_i$. So this λ_i's are your eigenvalues, right? For this real symmetric matrix, okay? So now I can make a matrix S which is $\{e_1, \dots, e_n\}$ written in columns, okay? Then I know S^{-1}, because this is an orthonormal basis vector, basis vectors, so S^{-1} will be the same as S^T, right? We discussed this before. So if you look at S^T, you multiply $S^T S^{-1}$, you can quite immediately see that, you know, only the diagonal entries will be 1, everything else will be zero because of the orthonormal property, okay? So this A itself can be quite easily written as SDS^T, okay? In using this orthonormal basis decomposition. And D becomes λ_1 to λ_n on the diagonal, okay? I hope this is clear. I hope… I've got the ordering of S and S^T wrong. I always get messed up, mixed up on this. Maybe it's, is it $S^T DS$ or SDS^T? One of the two, okay? So it doesn't matter what it is, it sort of works, okay? So this is the diagonalization, the diagonalization is important. Of course it matters, the sequence matters, but I mean the idea is basically the diagonalization, okay?

(Refer Slide Time: 11:42)

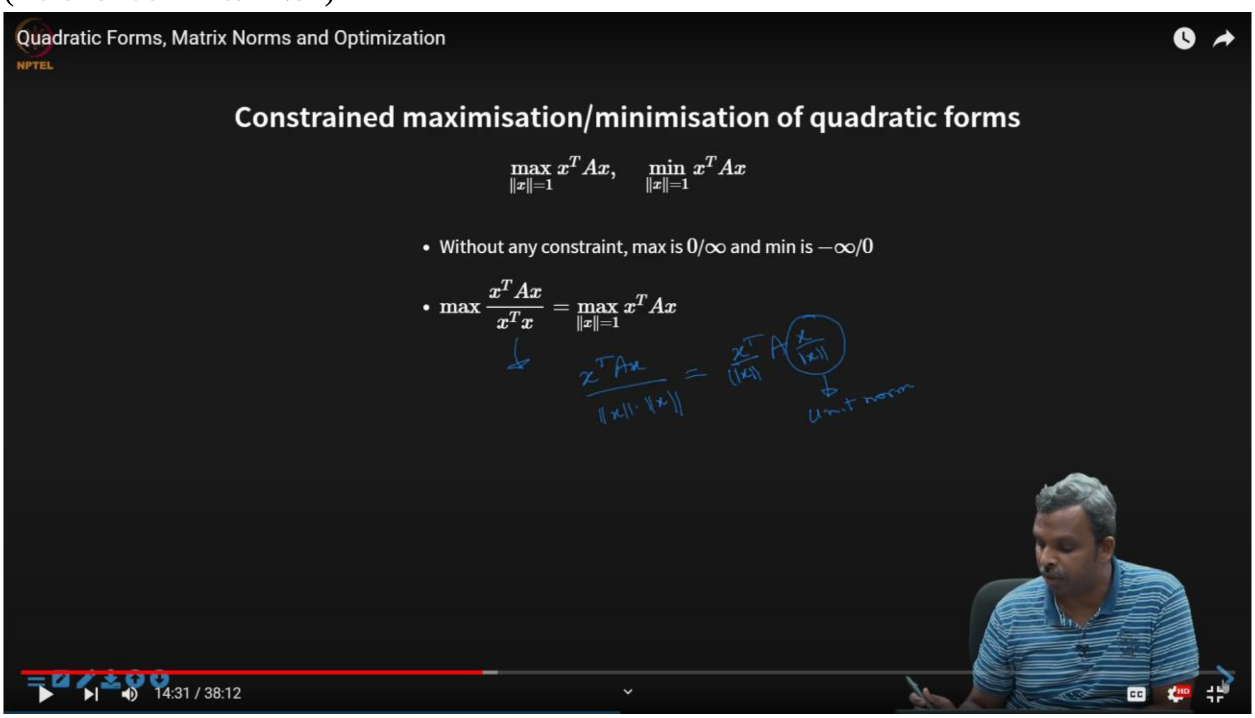


All right. So once you have a diagonalization like this, with a suitable S, SDS^T, then your $x^T Ax$ becomes x^T… Instead of A I will put SDS^T, okay? Once I put SDS^T here, I would get this and then I can define an s which is $S^T x$. And then what will happen to $x^T S$? That will be s^T, okay? So my $x^T Ax$ becomes $s^T Ds$ and D is diagonal, okay? And once D becomes diagonal, I know that this x^TAx is simply $\lambda_1 s_1^2 + \cdots + \lambda_n s_n^2$, there are no cross terms in this form, okay? Once I get to the diagonalization, okay? So the change of variables that takes you from, you know,

$x^T Ax$ gets… The change of variables that gets rid of all the cross terms in your quadratic form is strongly connected to the diagonalization and the spectral theorem like we studied for symmetric operators, okay? So you might say: what if A is non-symmetric? This all, this does not go, but you remember how I showed you that any quadratic form can be written as $x^T Ax$ where A is symmetric, right? So it's possible to do that. So symmetry is good enough. And this is a very nice way to get to the answer, okay? So you see that quadratic form simplification has some very nice connections to real spectral theorem, okay?

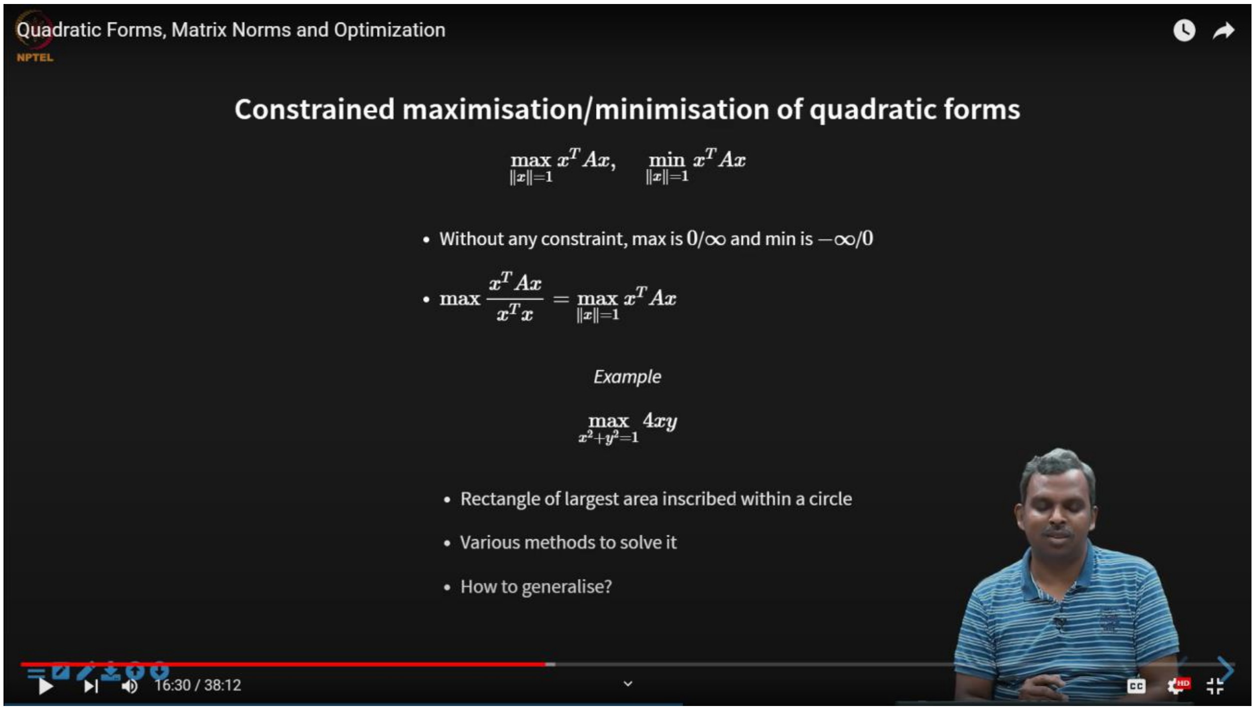
So let us move on. The next, very interesting question with respect to quadratic forms is what is called constrained minimization and maximization of quadratic forms, okay? So… Okay. We have simplified $x^T Ax$. We know how to simplify it. Can we put the simplification to some use, okay? In particular, can we find what is the maximum possible value for the quadratic form or the minimum possible value for the quadratic form with some constraint on x, okay? Now that is the sort of problem I have defined here. I am going to, I want to $\max_x x^T Ax$ with the constraint that $||x|| = 1$, okay? The same thing with minimization. So these are very interesting problems. They have a lot of applications. They show up quite often in practice, okay? So it's good to do it.

(Refer Slide Time: 14:31)



But why is this constrained? What if you do not put this constraint, you might ask. If you do not put the constraint, usually you will get zero or infinity, okay? So it depends on A, whether it's negative, non-negative and all that. But, so if you don't constrain, you can quite easily solve this problem. So it's not a big deal if you don't constrain. You can just drive it to infinity. You can take some examples and see that it's easy to drive it to $\infty, -\infty$ or, you know, it may be 0, okay? So it is easy to do this if there is no constraint. So when there is constraint, it becomes more interesting, okay? So another way to view this problem, if you do not like this constraint, if you want to think of some other version of this problem, is to look at this problem. You $\max_{x} \frac{x^T A x}{x^T x}$. Now $x^T x$ is just $||x||^2$, right? So you can push that into the x and the x here. So it's, both of these are equivalent, okay? So there is no, it is not difficult to see that both these are equivalent. In case you are worried about it, you can you can do this. You can see, no? $\frac{x^T A x}{||x||^2}$, $||x||$ into $||x||$, right? And that is the same as, you know, I can write $\frac{x^T}{||x||}$, okay, sort of, you know, hope you understand the notation, $\frac{x}{||x||}$, okay? So this guy is the same as, you know, something with unit norm. So this now, this whole thing has unit norm, right? So whether you are maximizing over all x or you are maximizing just $x^T A x$ with the norm being 1, it is the same thing, okay? So this problem is also equivalent. So this is something nice. So this sort of tells you the effect of A on the $x^T x$, okay? You have a norm of x, what A does to the quadratic form when compared to the norm of x, if x is very large, what does it do? So this is, this ratio is meaningful in so many different ways, okay?

(Refer Slide Time: 16:30)

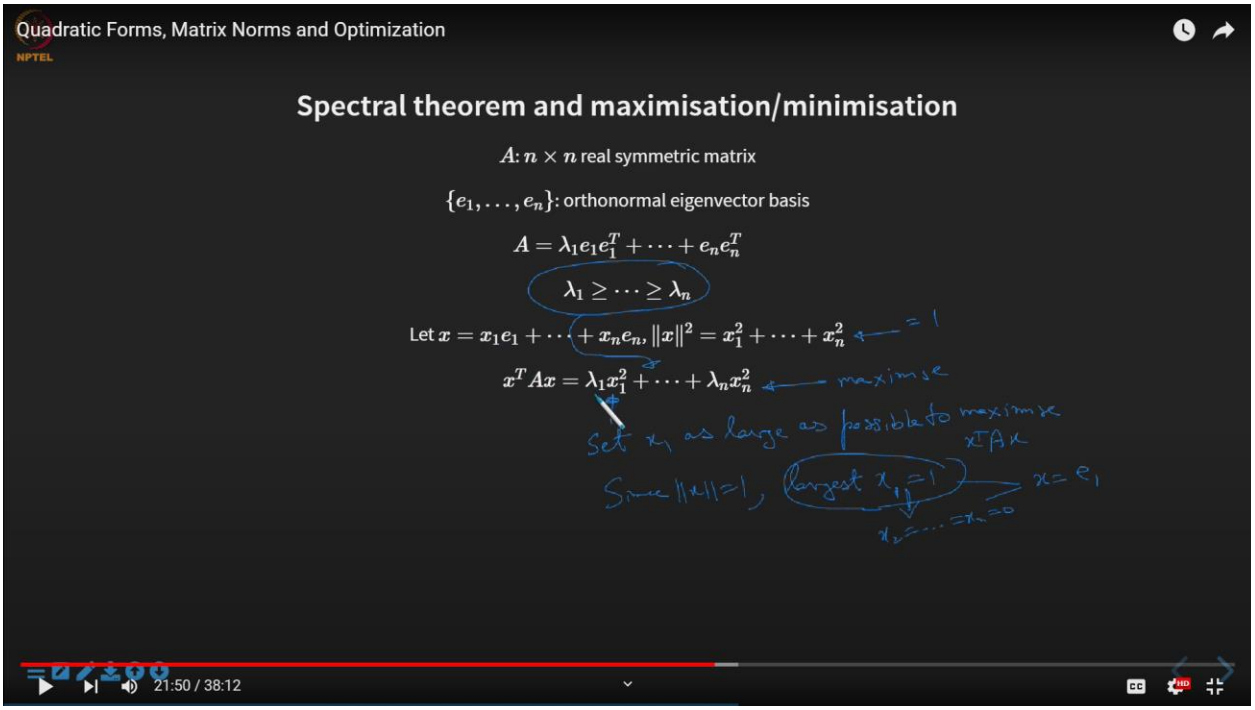


Here's a very simple classic example. You might have solved this problem in so many other ways before or you might have, you might be familiar with this. $\max_{x^2+y^2=1} 4xy$. $4xy$ is clearly a quadratic form, okay? So this is the constraint that, you know, $||x|| = 1$. What is $4xy$? This has a very simple interpretation of being rectangle of largest area inscribed within a circle. Think about why that is true, you know? You put a circle $x^2 + y^2 = 1$ describes the circle. Any (x, y) that you pick on the circle, you know, $(-x, -y), (x, y), (x, -y)$, right? That's the rectangle. That area is $2x \times 2y$, which is $4xy$. So $\max_{x^2+y^2=1} 4xy$. So this gives you that there are various methods to solve it and in this case it's not very hard, right? You can change it to one variable, right? You can make y as $\sqrt{(1-x^2)}$ and then you will get $4x\sqrt{(1-x^2)}$ and then after that it's only a single variable maximization problem. You can differentiate, equate to zero, you can do such things. So it is possible to do all that here using calculus. There are various other methods, also geometric etc. that you can use, okay? But the question here is how to generalize, right? How do I take, how do I take this problem and then sort of generalize it to n dimensions, you know? Any quadratic form with this constraint, how do you solve, okay? n dimensions, arbitrary A, okay? A quick thing I want to point out here. You can also check that without any constraint. Supposing I don't put $x^2 + y^2 = 1$. If you only want to max $4xy$, you know what's going to happen, right? You can go to $-\infty$, $+\infty$ quite easily by choosing large and large and large values for x and y, okay? Or positive, negative like that, right? So you pick it carefully, so you can drive it anywhere you like. So max and min without the constraint for quadratic forms is not so interesting. With the constraint it becomes very interesting, okay? So, anyway. So that is that sort of motivates the problem for you and the real spectral theorem is or the spectral theorem has a wonderful application in this context. We will see some quick applications and we will see how easily it solves the problem. And also, you know, getting rid of the cross terms is also very important here, okay? So you will see all that plays around.

Okay. So here is how the spectral theorem is involved. We have the $n \times n$ real symmetric matrix and associated with that matrix is the orthonormal eigenvector basis. Again think of this e_1 to e_n in terms of coordinates with respect to the standard basis. That's a, that's the easy way to write things down. It helps us to write things down very easily. So now once I know that, I know the spectral theorem tells me A can be written as $\lambda_1 e_1 e_1^T + \cdots$, I forgot the λ_n there, okay, so the λ_n should come there, $e_n e_n^T$. And this λ_1's I will pick like this, okay? So I will pick λ_1 to be my largest eigenvalue. Then λ_2, then λ_3 so on to λ_n, okay? So λ_1 is the largest eigenvalue of A, okay? λ_n is the smallest eigenvalue of A. It could be positive, negative, whatever. I do not care. But I know it is real, right? All eigenvalues for A are real. Because it's symmetric and then, you know, λ_1 is the largest eigenvalue. λ_n is the smallest eigenvalue, okay? So this is how it goes, okay? So this is A, this λ_n I'd forgotten here, so typo, you should put it there.

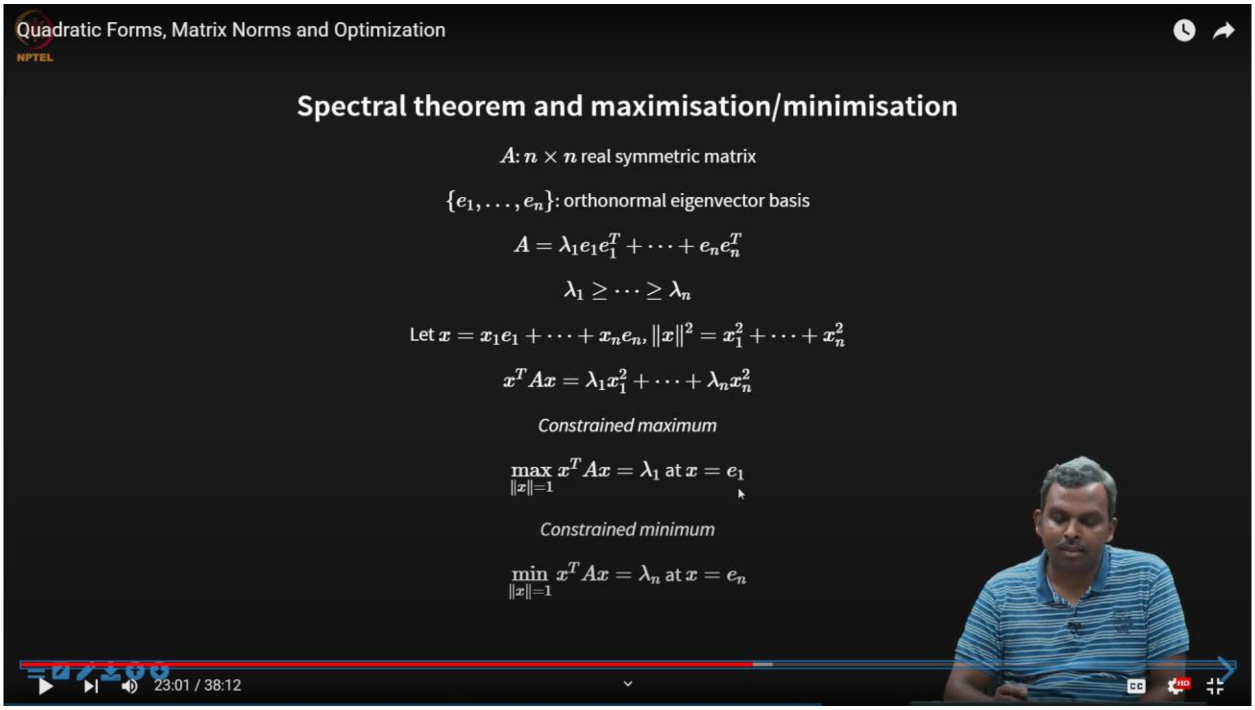
Okay. So now let's see how this can be used, okay? Now in a quadratic form, right? $x^T Ax$. I'm going to think of x, right? So this is same as what we did with the diagonalization. I'm just describing it a bit differently. But this description helps you in the maximization more directly I think, okay? So you express x in terms of the orthonormal basis $\{e_1, \dots, e_n\}$, okay? So x_1 is its coordinate with the, of e_1, x_n is its coordinate of x which multiplies e_n, right? So x becomes $x_1e_1 + \cdots + x_ne_n$, right? So you basically express x as a linear combination of e_1 to e_n. I know this is always possible because this is the basis. And you get x_1 to x_n. So clearly $||x||^2$ is $x_1^2 + \cdots + x_n^2$. Is that okay? So why is this important? So, because I am going to constrain my norm to be 1, okay? So this $x_1^2 + \cdots + x_n^2$ have to sum up to 1, okay? So this is something that has to happen. So now what is x^TAx, okay? You put A in this form and x in this form and you substitute x^TAx and simplify, you know what is going to happen, right? All the cross terms will disappear, you will get $\lambda_1x_1^2 + \cdots + \lambda_nx_n^2$, okay? So now this is our problem, okay?

(Refer Slide Time: 21:50)



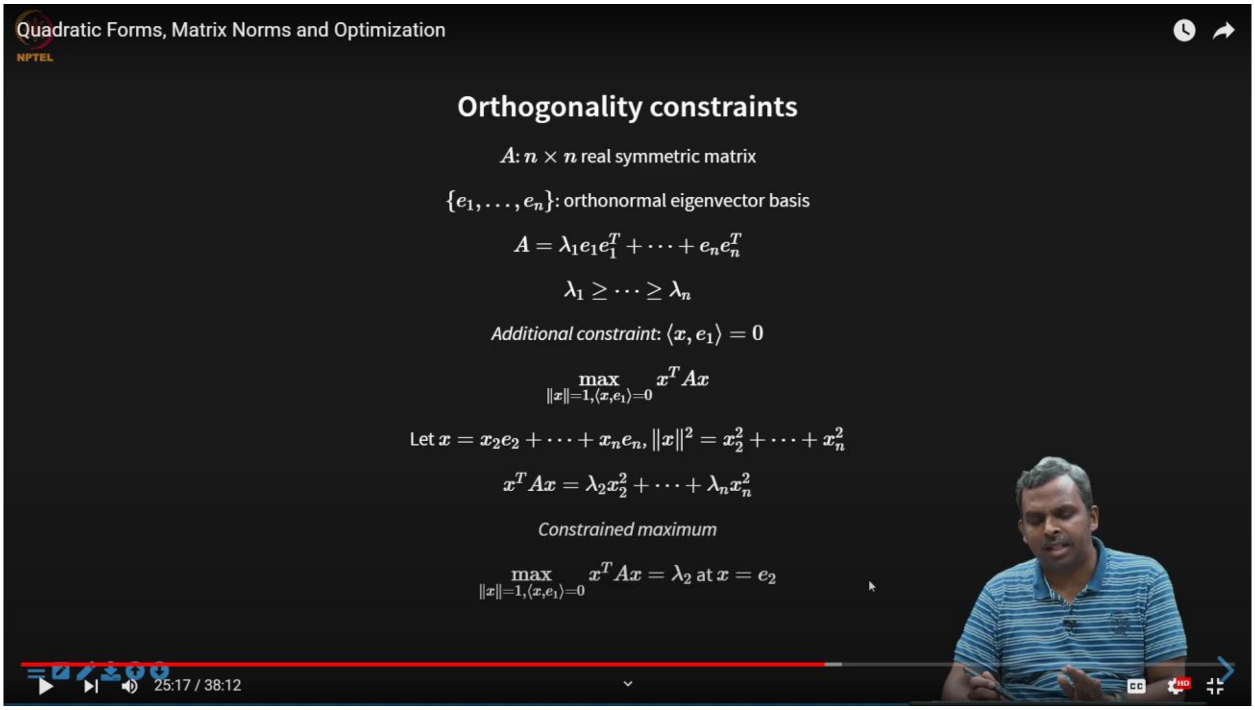
So I want to maximize this or minimize this, okay? Let us look at maximizing this with the constraint that this is equal to 1, okay? And what do I know? I know this guy, okay? This is very important, okay? λ_1 is the largest, okay? So if I want to maximize this, what should I do? I should make sure that whatever multiplies λ_1 is as large as possible, right? Right? That would give me the maximum value for this summation because there is nothing else, you know, that's confusing this. Every term is just $x_1^2, \ldots, x_n^2$, right? It's all positive, every term contributes, in some sense, positive, and then the λ gets multiplied by a positive number only, right? So λ_1 is getting multiplied by x_1^2 which is positive. So if I want to maximize this sum, I want to, I want to make sure that x_1 is as large as possible, right? Now suppose I do not put any constraint on x. I can make x_1 ∞ or something very large. And then just simply the sign of λ_1 will sort of dominate what happens, whether in maximization or minimization. Let's say λ_1 is positive there, right? So then if you want to maximize, you simply let x_1 become larger and larger and larger, you will get plus infinity, right? So that's what will happen, okay? But we have this constraint that $x_1^2 + \cdots + x_n^2 = 1$. So what is the largest possible value that x_1 can take? That is 1, right? Largest… So we want to set x_1 as large as possible to maximize, right? Okay? So that is the main story. And given the constraint, okay, since $||x||$ is constrained to be 1, largest x_1 equals 1, okay? So you cannot go above 1, okay? So the largest x_1 you can put is x_1 equal to 1. And once x_1 becomes 1, what happens to x_2 to x_n? All of them become zero, okay? So this corresponds to $x_1 = 1$ and that implies $x_2 = x_n = 0$, and that implies $x = e_1$, isn't it? So if you put x_1 is 1, everything else is 0, x becomes e_1. So to maximize $x^T Ax$ under the constraint $||x|| = 1$, you simply have to pick $x = e_1$ and the maximum value will be equal to λ_1. It cannot go above it, okay? So that's a very simple result that you get once you get this form. If you did not have this form, if you had x_1x_2, x_1x_3 and all, you can't make these easy calculations, okay?

(Refer Slide Time: 23:01)



So notice how the spectral theorem is playing a huge role here, okay? So the constrained $\max_{x,||x||=1} x^T Ax$ is simply λ_1 and it's achieved at $x = e_1$. That's it, okay? How simple is it to solve this using the spectral theorem, right? So there is no calculus, there is no differentiation, there is no… Nothing more is involved, just the answer comes out very cleanly using the spectral theorem, okay? The same thing can be very easily done for the constraint minimum, right? Supposing you want to minimize, what should you do, right? I want to make x_n as large as possible, right? Anything else I do will only increase my quadratic form's value, right? x_n should be as large as possible. So you pick $x = e_n$. So x_n is 1, everything else is 0. And the minimum becomes λ_n, okay? So maximizing and minimizing quadratic forms with a constraint is quite easy and that simply is directly connected to the maximum eigenvalue of A and minimum eigenvalue of A and achieved at the corresponding eigenvectors, okay? The maximum, the eigenvector corresponding to the maximum value or the eigenvector corresponding to the minimum value, okay? Simple enough, right?

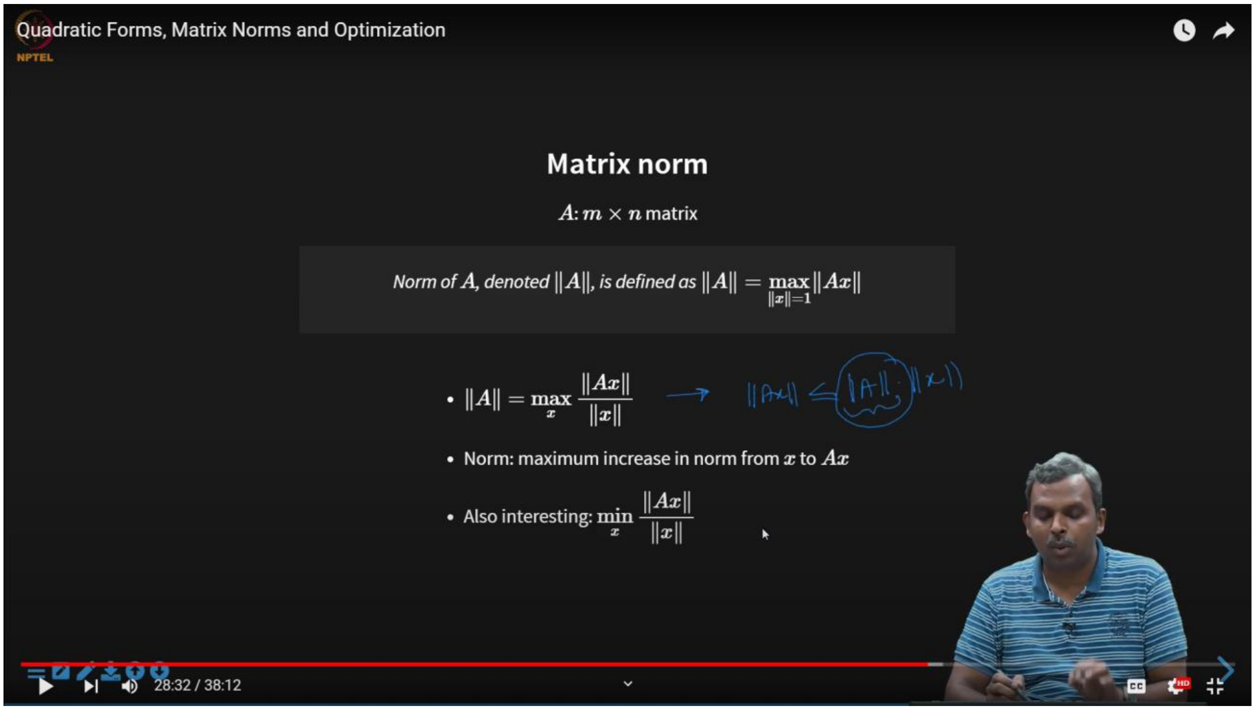
(Refer Slide Time: 25:17)



So we can further constrain the problem. So let us say I am not interested in the maximum eigenvalue, right? So I don't care so much about the maximum eigenvalue. I want to make sure that I am not in any way related to the maximum eigenvalue, okay? So you may have a problem like that, okay? So this basic problem I may want to put an additional constraint saying $< x, e_1 >$ has to be 0, okay? So I do not want anything to do with the maximum eigenvector. I do not want to be anywhere near there but I want to maximize the quadratic form given the constraint that $||x|| = 1$ but $< x, e_1 > = 0$, okay? Even these kinds of problems are easy enough to do, right? So if you just think about this problem, just fundamentally, you have maximization of a quadratic form, degree two, right? (x_1, x_2). What are the constraints? There is a degree two constraint and a degree one constraint, okay? So it's sort of complicated from a pure calculus point of view if you think of it as an optimization problem. But because of the special properties of quadratic forms and because of the spectral theorem, even these kinds of problems have a very, very simple solution. What do I do? If I know $< x, e_1 >$ is 0, then I know x has to have this form, right? x has to belong like this. The e_1 is 0, everything else is non, could be non-zero and $||x||$ will simply be $x_2^2 + \ldots + x_n^2$. And $x^T A x$ is this, right? And what is the maximum among this? λ_2, okay? So the constrained optimization is quite easy to do. You see that its λ_2 and it's achieved at $x = e_2$, that's it, okay? So these are all easy and nice results that one can get using the spectral theorem when you are maximizing with some constraint, when you are maximizing the quadratic forms, okay?

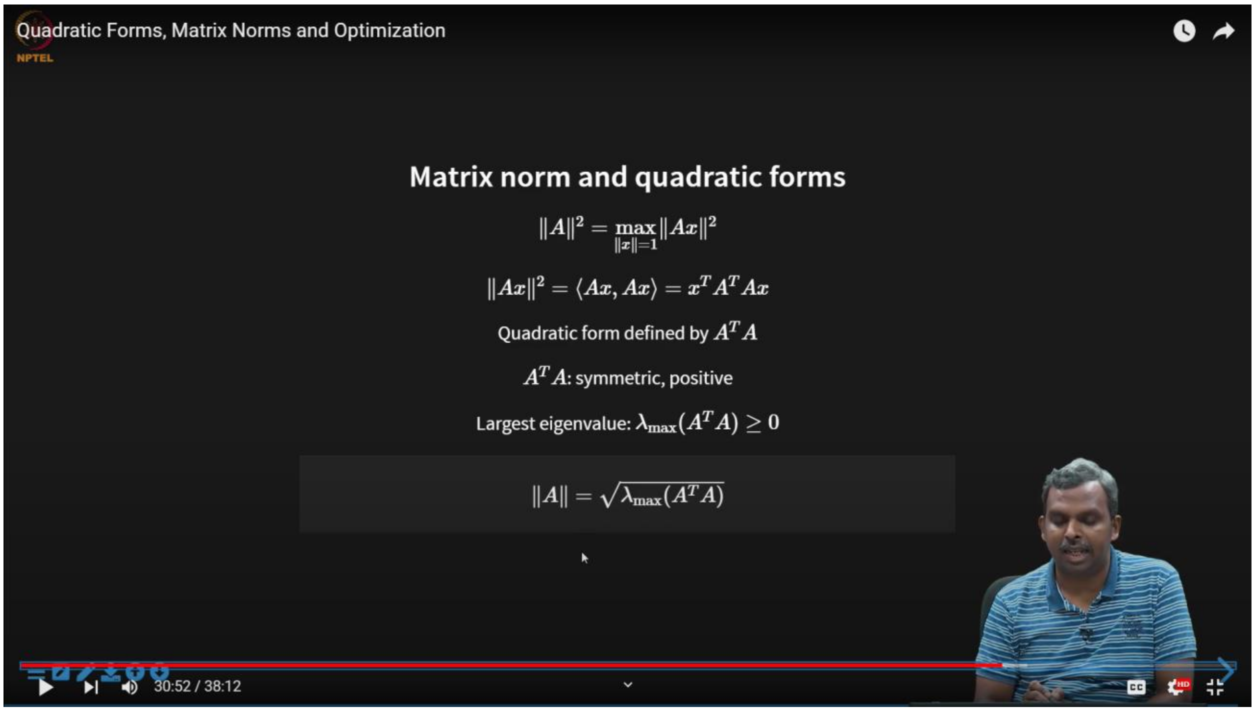
So we have seen simplifying quadratic forms and maximizing quadratic forms. Spectral theorem simplifies matters tremendously. So you can do this in general for any size, right? So that is the power of this. So there is an analogous way to think of complex also, and, you know, self-adjoint matrices and all that. The same thing goes through, there is no problem there except that, you know, you have to write it in a bit more complicated language. So I have chosen this simple one for description, okay? All right. The next couple of things I want to point out, I'll go through a bit faster, and these are not… Slightly advanced ideas, not so important, but I want to point out how easy it is to use the ideas we have discussed to quickly provide definitions at least and give you some motivation for why these are interesting from a theoretical point of view, okay? Supposing I have an $m \times n$ matrix. People associate a norm with the matrix, okay? So, so far we've been thinking of norm with vectors. You can also associate a norm with operators, okay? So A of course represents an operator. I am thinking of it in the matrix representation, think of it as an operator only if you want. So the norm of A, denoted $||A||$, you define it like this. It's a $\max\limits_{Ax, ||x||=1} ||Ax||$, okay? So there is a good motivation for why this is defined this way, okay? So you look at all x which have norm 1 and look at how much A increases that norm, right? Ax is a vector, right? But remember, Ax is a vector in $\mathbb{F}^m$, right? x is a vector in $\mathbb{F}^n$, okay? So, but still, you know, it's a vector. So if you, if you start with norm 1 vectors, how much does A increase norm by, okay? So it's sort of like a reasonable thing to know, right? If you have an operator, you want to know how much, you know, amplification it's doing to norm, right? So that represents the norm of the operator, okay? x has a norm, Ax has a norm. What is the norm of A? It is the maximum amplification it can do to norm, okay? So there is a sense in which this works out, you know? Just like we looked at before. $||A||$, instead of doing the constraint of $||x|| = 1$, you can say $\max\limits_{x} \frac{||Ax||}{||x||}$, okay? So it's the same thing because it's linear, you can push this x inside. So you have norm of, you can think about why this is well defined. So this is going to work out and this is correct, okay? So this and this are the same. So in this sense, $||A||$ represents the maximum amplification that x will get from A, okay? So that is the norm, okay? That is good to know. So maximum increase, but this is maximum, you know, multiplicative increase, sort of maximum amplification, okay? So there is another way to write this. People write this also differently. So this implies, you know, $\left||Ax|\right| \leq \left||A|\right|\left||x|\right|$, okay? If you know what $||A||$ is, $||A||$ plays this kind of a role, okay? So quite often you want to know after operating with A how bad can my norm be, how big can be norm be, that gets bounded by $\left||A|\right|\left||x|\right|$, okay? So that is why this is very interesting in practice. One might also want to look at the minimum, okay? Even though we do not look at that too much from the norm, the minimum does not play at all, but the minimum in practice can also play a role, you know? How much will A, you know, what is the opposite of amplification, whatever, that is how much A reduces the norm of x, okay? So that also might be interesting from some point of view, okay? Anyway. This is norm. I don't want to go too much into why norms are useful, but you can see that this is well defined and one can look at how to operate with it, okay?

(Refer Slide Time: 28:32)



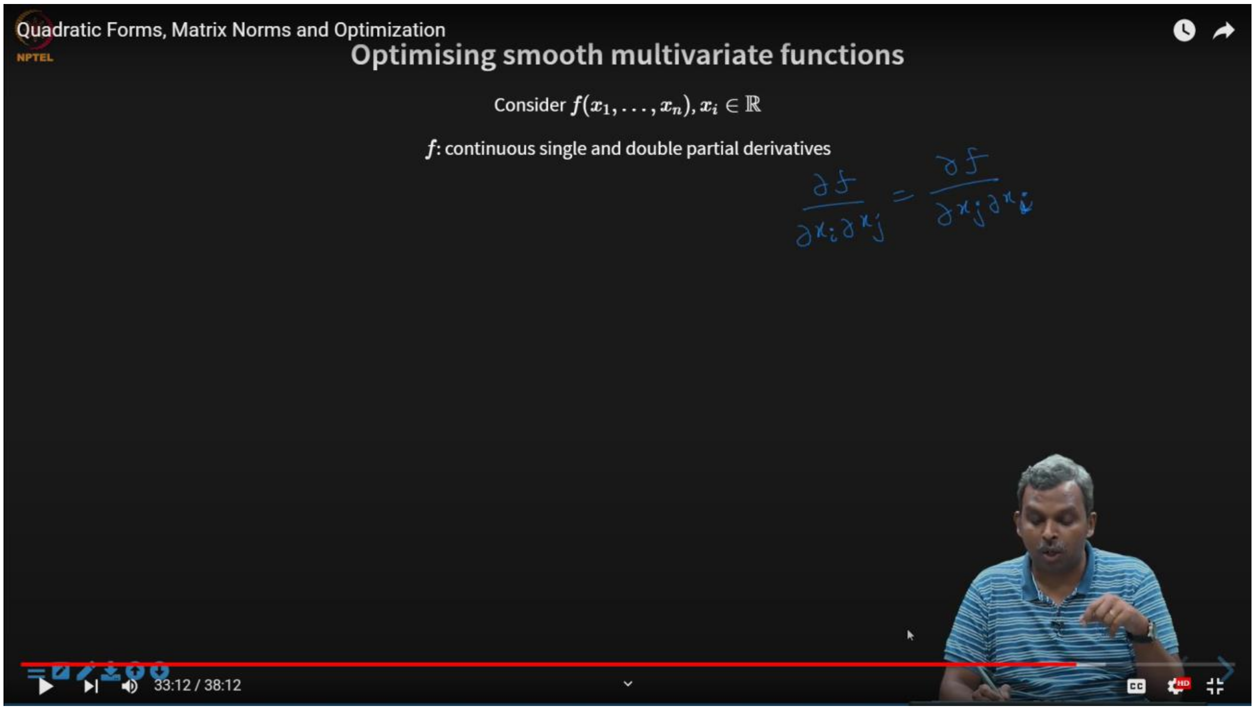
Once again the spectral theorem comes to our rescue in characterizing norms in terms of eigenvalues and eigenvectors. But remember, A is $m \times n$, where will eigenvalues come from, okay? So notice how they enter the picture here. So there is a connection between matrix norm, quadratic forms and all we have been studying. So for that, you just look at norm squared, right? So norm squared is going to be max$||Ax||^2$, okay? So once you go to $||Ax||^2$, I can write in terms of inner product, right? Now $||Ax||^2$ is $< Ax, Ax >$. What is $< Ax, Ax >$? It's $x^T A^T Ax$, isn't it? Okay? So it's quite easy to see why this is true. So this is nothing but a quadratic form defined by $A^T A$. Notice $A^T A$ is symmetric. It is in fact even positive, okay? So we can come to that later. We know its operator-adjoint product, right? So it is a positive operator, it is so symmetric, definitely. So this $x^T A^T Ax$ is a quadratic form now. So this problem of computing norm for an operator is actually the same as constrained maximization of a quadratic form, okay? So all these problems are sort of similar and they are all defined from the same sort of assumption and they all tie up into the eigenvalue of symmetric operators, okay? So it's all very interesting how these things show up. So that's something to keep in mind. So while we define norm as some, you know, maximum amplification that is produced by an operator on the vector x, it is the same as the maximum quadratic, you know, quadratic form that can take with the constraint. And the quadratic form is defined by $A^T A$ even though A is $m \times n$, $A^T A$ is going to become $n \times n$ or something like that, okay? So you get something very reasonable here, okay? So let us say the largest eigenvalue of $A^T A$ is $\lambda_{max}(A^T A)$, and that's going to be non-negative, right? So because it's positive, it's a positive operator, all its eigenvalues are non-negative. So it's going to be greater than or equal to zero, okay? So from this we know what is the largest constrained maximization of $x^T A^T Ax$, it is simply $\lambda_{max}(A^T A)$. So this gives you a very succinct, clean formula for the norm of any operator A. The norm of any operator A is $\sqrt{\lambda_{max}}$ of $A^T A$, okay? So it's a very clean nice relationship for the norm of A. And once again you can see the spectral theorem and positive operators and all of these things play a role here, okay? So that's all I want to say about matrix norm. And maybe if you take some advanced courses in linear algebra, you would study more about matrix norms and its applications in general, okay?

(Refer Slide Time: 30:52)



So the last thing I want to point out in this lecture is basically, you know, an application of positive operators in optimization, okay? So this shows up when you look at smooth multivariate functions, arbitrary functions. Let us say some function $f(x_1, \ldots, x_n)$, each x_i is a real variable let's say. But f is smooth. What do I mean by smooth? I will assume it has continuous single and double partial derivatives. You know what partial derivatives are, right? Partial derivative of f with x_1 or x_2 or so on. And also double derivatives, you can define derivative, partial derivative twice, say with respect to x_1, or once with respect to some x_i, and another time with respect to some other x_j, okay? So all these are continuous. We will assume so, you know, you have relationships like this being true, right? The partial derivative, if you do in any order, they will be equal, right? So when you are smooth, you can make all these assumptions, okay? All these are true, bounded, everything you can assume if you want. So basically smooth nice functions like this, when you want to optimize them… So optimize meaning you want to find… So now you have to sort of imagine in your head. So you have all this. Think of three dimensions or four dimensions or something. All these x_1 to x_n are living in this big space. At every point I've defined a function for you, okay? And this function is going to go up in some ways, come down in some way. If you're in any point $(x_1, \dots, x_n)$, it may in, you can go in so many different directions and in every different direction this function may increase or decrease or behave in various ways, okay? Now I want to optimize this function. So if I want to find a point where this function, wherever we go, it's only going to go down, it's like a hill, you know? Hilltop, right? So you want to find maximum value or minimum value, okay? Bottom of a valley so to speak, right? So you come somewhere and then whichever direction you go, you can only go up, right? So I want to find those kind of points, local minima, local maxima they are called for a multivariate function. And this is a, this is at the heart of all optimization problems. Many engineering problems, wherever you do, ultimately you will end up with one big function of several variables and you want to find local maxima, local minima, understand this function, understand the landscape of this function, how does it look from different points of view. So this is a crucial thing in all engineering applications, okay?

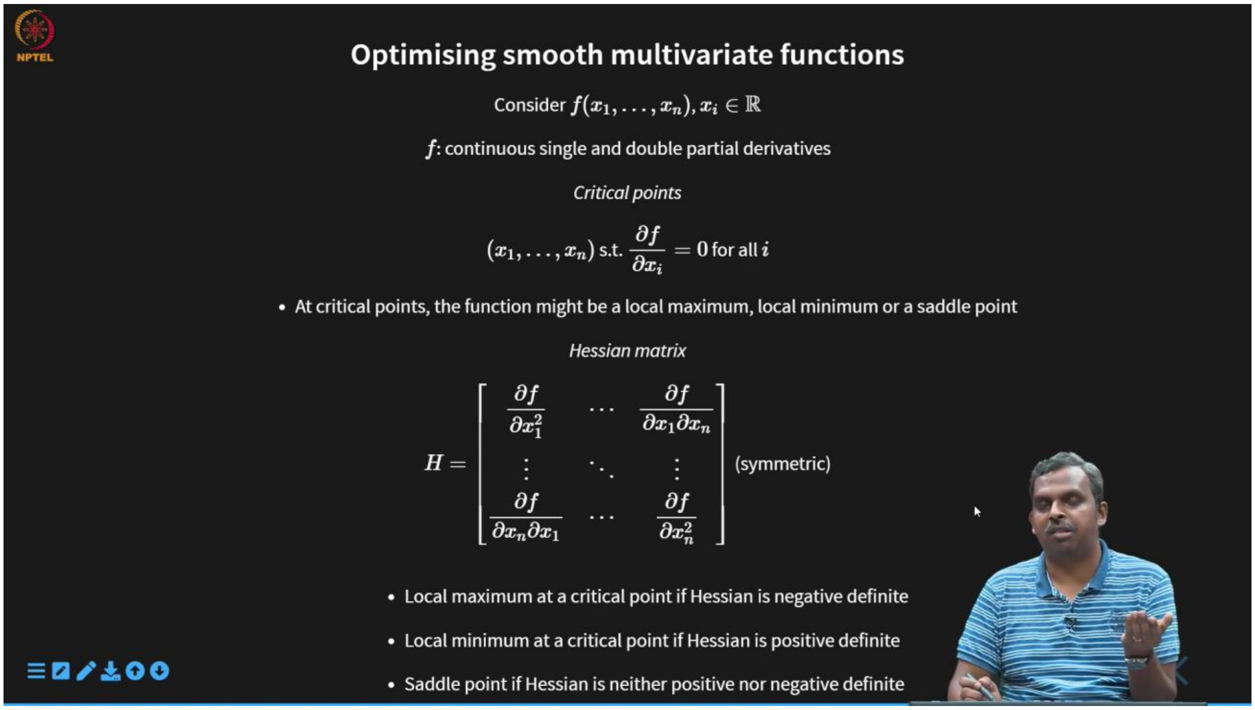
(Refer Slide Time: 33:12)



So what is the connection between this and linear algebra and positive operators? Here is the connection, okay? So we won't prove this result, it's way beyond the scope of this class, I'll just point it out and later on if you take a course in optimization you will study more about this. The first thing is: there is this notion of critical points for such functions. What are the critical points?

The points at which all partial derivatives vanish, okay? So the derivative with respect to one variable is, sort of tells you how the function behaves if you move in that direction, right? So if my derivative goes to zero, it means that function is not really changing that much in that direction, okay? Neither increasing nor decreasing, right? So that's sort of what this means. It's sort of flat in that direction. Now if all partial derivatives are 0, in all directions it has that behavior, okay? So you expect something to happen there, right? In all directions I am not seeing any increase or decrease, right? So that should correspond to a local minimum, local maximum, or it could also correspond to something called a saddle point, okay? So the saddle point is a little bit more complicated. As in local maximum means everywhere it decreases, right? Local minimum means in, everywhere, in every direction you go, it only increases. But saddle point, in some directions it may increase, some directions it may decrease, okay? So it's sort of, it's a little bit more than a general point. You might say, okay, what is the difference between a general point and a saddle point? Saddle point is a critical point as in the partial derivatives vanish, okay? So it's sort of flattish, but in some directions it will go flattened up in some other direction it will go flattened down, okay? So that's why you can imagine the saddle on a horse, right? If you think of a horseback, and there is a saddle, in some directions you go down, some directions you go up, right? So that's the saddle point. It's still critical, as in the partial derivatives vanish. In general points the partial derivatives may not vanish. So there's no flat behavior in the neighborhood. But this has, this, it's not a maximum or a minimum, okay? So this is, classifying critical points into maximum minimum and saddle point is very important. You need to get a sense in your function of multiple variables, how it behaves in different points, okay? And once again linear algebra comes to your rescue through this notion of a Hessian matrix, okay?

(Refer Slide Time: 37:26)



This is the hessian matrix. It is, basically you collect all the possible partial derivatives and put them together in a matrix, right? So for every x_i and x_j, you have a double partial derivative, okay? You, might be x_1^2. you put it on the diagonal. If it is $x_i x_j$, you put it in the $(i, j)^{\text{th}}$ element. $x_j x_i$ will be in the $(j, i)^{\text{th}}$ element, okay? Because of these continuous assumptions, okay? It is very reasonable to expect the Hessian to be symmetric, okay? For most functions, you will get a symmetric Hessian, okay? So this symmetric operator or matrix, at every point you can evaluate it, right? Of course, in the critical points also you can evaluate it, okay? So all this is sort of similar to your double derivative method for detecting local maxima, minima, right? Functions of one variable you might have heard you would take a first derivative, if it is equal to zero you get critical points. And how do you find whether it's maxima or minima? You take the double derivative, okay, and check its sign. If it's, double derivative is positive it's a minima, double derivative is negative it's a maxima. So sort of like that, this is the extension or generalization of that to n dimensions. And positivity of the symmetric operator now will play a role, okay? In one dimension, it was just positivity or negativity of the double derivative at the critical point. Here you have to look at the Hessian at the critical point. It will be a symmetric operator. You have to check whether it is positive or not, okay? The result, which is true, we won't prove this, is that a local maximum at the critical point if Hessian is negative definite, local minimum at a critical point if Hessian is positive definite at that point, and if it is neither positive nor negative, it can also happen right, then, all operators need not be positive definite or negative definite, then you get a saddle point, okay? So this is a nice characterization, and this is how positive operators, the notion of positivity in operators is used in optimization. And if you take further courses in optimization, you will see such ideas explored in great detail. Maybe you will see a proof for this result. So you will see all that, okay? That's the end of this lecture. Hopefully this lecture gave you a sense of how these, you know, spectral theorems are not just idle theoretical results, these characterizations are very powerful. They’re powerful in optimization of simple things in some sense like quadratic forms and even more complicated arbitrary functions, as long as they are smooth in some way, okay? Thank you very much.