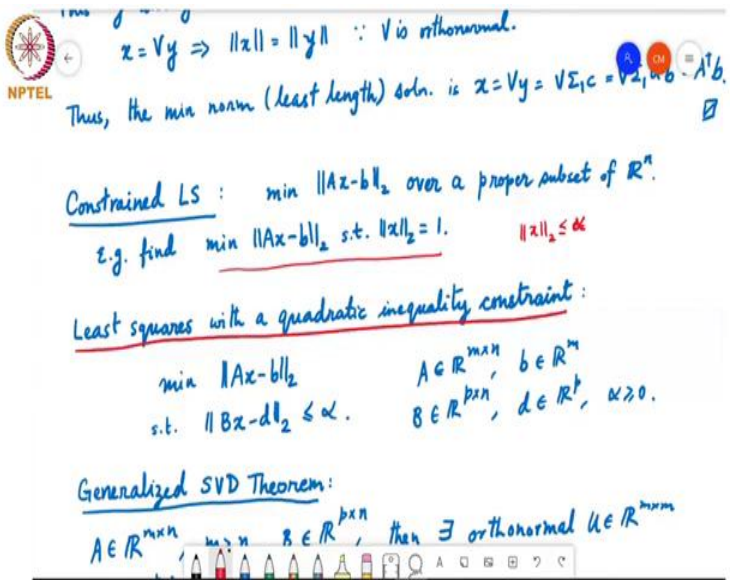
Matrix Theory

Professor Chandra R Murthy
Department of Electrical Communication Engineering
Indian Institute of Science Bangalore

Constrained least squares

(Refer Slide Time: 0:17)

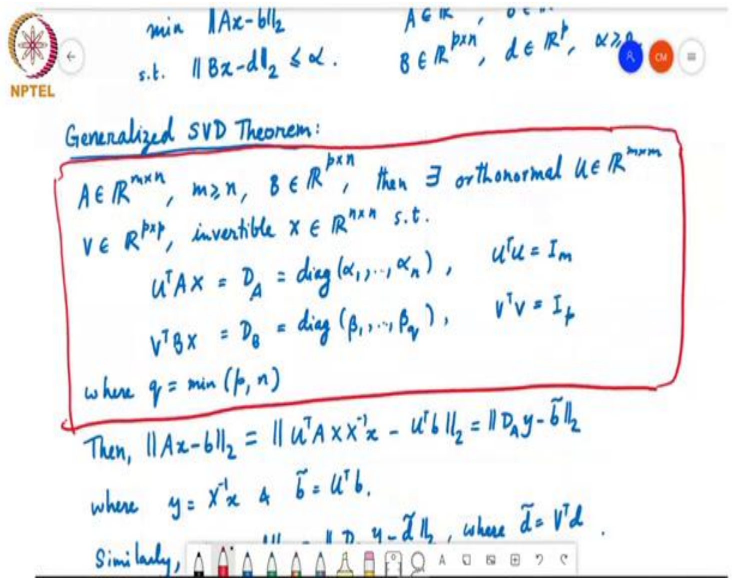


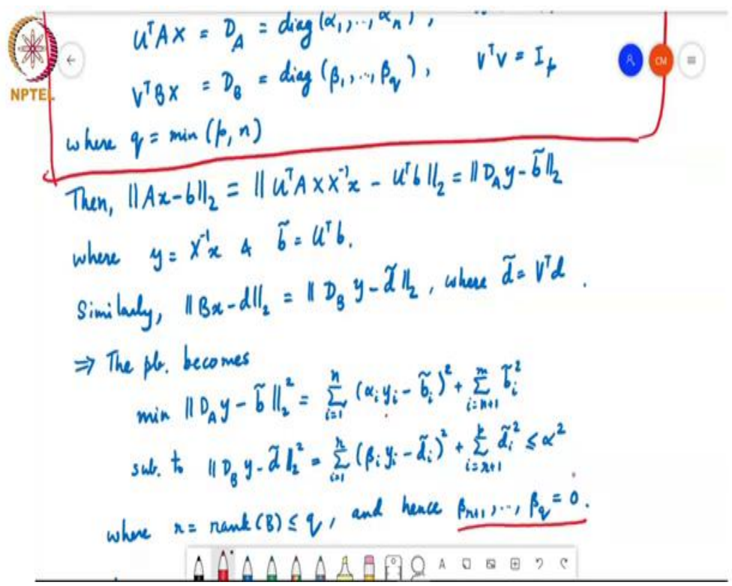
So, the next kind of problems we may be interested in solving are to minimize this the Euclidean norm between, Euclidean norm of Ax minus b over some subset of R to the n, not over, not the unconstrained problem over all possible x's. So, for example, you may want to solve for the minimum x, Ax minus bl2 subject to non x2 equals 1, this is called least square minimization over a sphere.

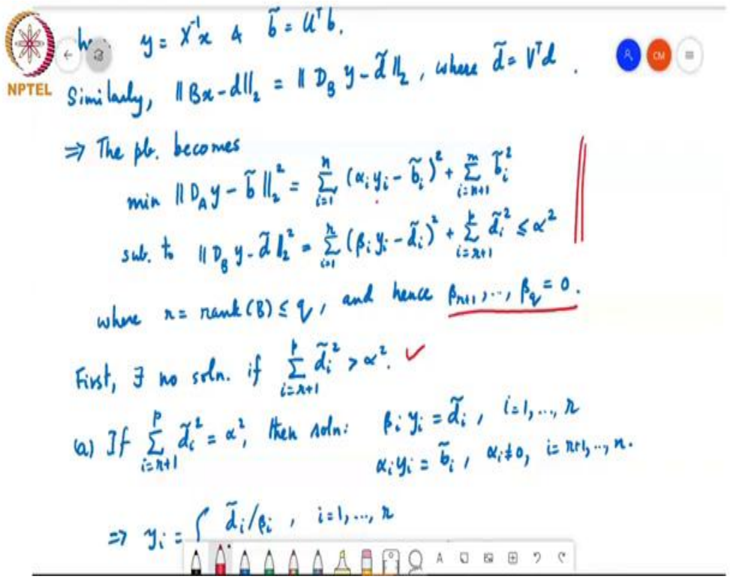
So, before we look at this particular problem or problems of this kind, or it could be for example, norm x2 less than or equal to alpha. So, this is like looking among points within a certain sphere. So, instead of before looking at this particular problem, let us look at least squares problems with a quadratic inequality constraint. So, this the general form of such a problem is given like this, minimize the l2 norm of Ax minus b subject to the l2 norm of Bx minus d is less than or equal to some value alpha.

So, here A as usual is of size m by n, b as usual is of length m. And this capital B matrix here is of size p cross n and d is of size, is of length p, and alpha is something which is greater than or equal to 0. So, clearly if alpha is less than 0, I am asking that some norm should be less than or equal to a negative number, which is never possible. So, there will not be any solution. This problem is meaningful only if I said, if I use some threshold, I mean, if I, if they bound on Bx minus d2 is less than or equal to some positive, at least not negative number alpha.

(Refer Slide Time: 2:24)







Now, in order to solve this problem, we need one more result, which is called the generalized SVD theorem. And again, I do not have time to prove this theorem for you. But this is the result. So, we will just utilize this result to find the solution to that problem. So, if you are given two matrices A which is of size m by n, m greater than or equal to n, B which is a size p cross n.

Then you can find two orthonormal matrices u of size m by n and V of size p cross p and an invertible matrix X of size n cross n, such that u transpose Ax is a matrix DA, which is a diagonal matrix containing alpha 1 to alpha n and V transpose Bx is DB, which is a diagonal matrix containing beta 1 through beta Q and here Q is equal to the min of p n, so, yeah Q is the min of Bn, and u and v are orthonormal matrices.

So, this is the theorem which we are going to utilize. So, for now, it is just some notation, but then the point is that this X here is invertible. So, that is one thing to keep in mind. So, if so, using this theorem, I can write Ax minus bl2 norm to be, I can pre multiply by u transpose and I can insert an XX inverse here. So, I get u transpose AX, X inverse times small x minus u transpose bl2 norm. And u transpose AX is this matrix DA, which has entries alpha 1 to alpha n on the diagonal, and u transpose b I will define to be B tilde. And this X inverse times small x I will define it to be y.

So, just doing a variable transformation here. And similarly, I can write Bx minus d to be DB times Y minus d tilde and so I am pre multiplying by V transpose and so d tilde is V transpose times small d. And DB is this matrix diag of beta 1 through beta q. So, with this the, our objective function or the optimization problem, it becomes the minimize DAy minus b tilde l2 norm square, subject to DBy minus d tilde l2 norm square less than or equal to alpha square.

But since DA and DB are both diagonal matrices, I can easily expand this out and write it as the summation i going from 1 to n alpha i yi minus bi tilde square. And for the remaining entry, i equal to n plus 1 to m, the yi does not touch those entries. So I will just so this term is equal to 0 and so I am just left with bi tilde square. And similarly here, I will have beta i yi minus di tilde r square.

And what I have done here is to take the upper limit to be equal to r, where r is the rank of this matrix B and so that these beta 1 through beta r are always nonzero. So, if r is the rank of B beta r plus 1 through beta q are always equal to 0. And so for those terms, this term is 0. And so I have i equal to r plus 1 through p di tilde square less than or equal to alpha square, this is my constraint.

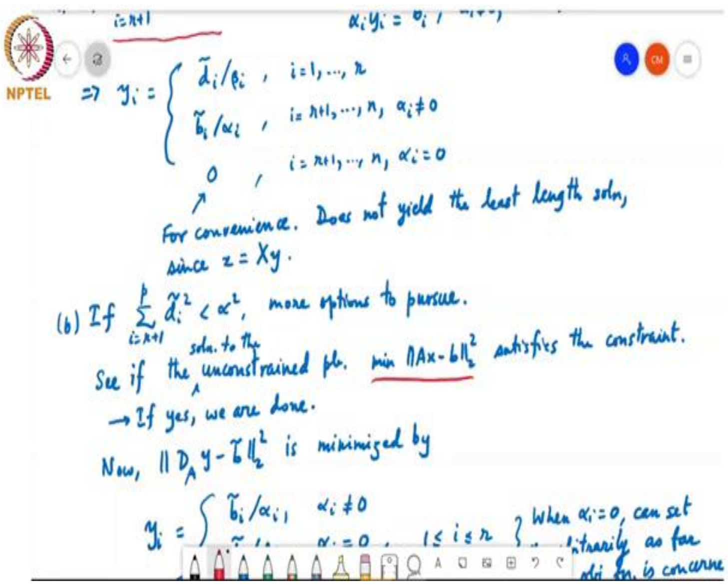
So, this problem looks a little bit easier, but we still have to do some work in order to solve this. Now, first of all, notice that no matter what, y I choose this DB y minus D tilde l 2 norm square is always at least equal to this term, sigma i equal to r plus 1 to p. The internet was working beautiful till will now, I do not know, today, I am having some trouble with my internet. Oh, well, so let us continue.

I was just saying that we reduced the problem to this one where we want to minimize DA y minus b tilde square, which is actually diagonal matrices. So, I can actually expand it out and write it out, subject to Db y d tilde square, which again, expanded and written out is like this less than or equal to alpha square.

Now, no matter what yi I choose this cos function or this part here is at least equal to the second term here, because this is non-negative, and this is non-negative. And so if this itself is bigger than alpha square, then there is no solution. Now, if these two are exactly equal, then we have no choice but to make these two equal.

So, that we means we must choose beta i yi equals di tilde for i equal to 1 to r, or we should choose yi equal to di tilde divided by beta i, for i equal to 1 to r, and we can let alpha i yi equal to bi tilde for i equals r plus 1 to n, because we want to try to minimize this cost here. And we can do that if alpha i is not equal to 0. But if alpha itself equals 0, we do not, again, we cannot really affect this cost by choosing any yi, so we might as well choose yi equals 0.

(Refer Slide Time: 8:21)

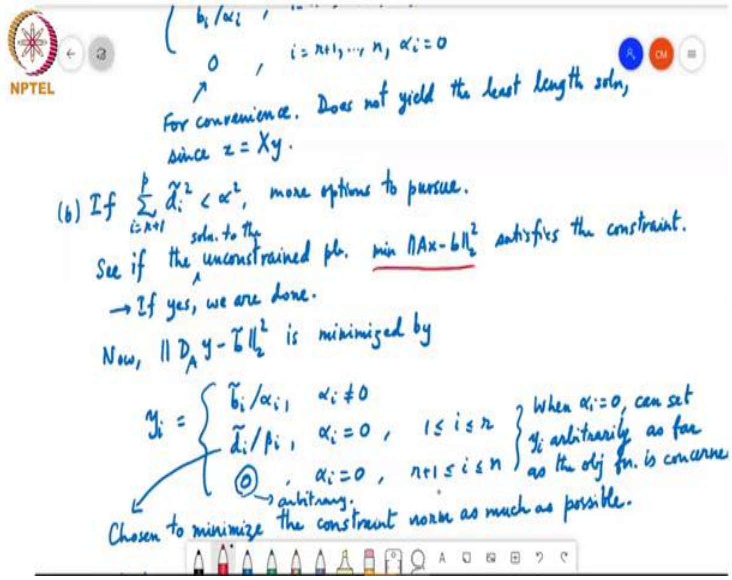


So this is the solution when summation i equal to r plus 1 to p di tilde square is exactly equal to alpha square. It is yi is equal to di tilde over beta i, for i going from 1 to r, and bi tilde overall alpha i for i equal to r plus 1 to n, but alpha i not equal to 0 and 0 for the other cases. So, we were able to solve it in this case.

And so 1, 2, the next case is, if i equal to, i equal to r plus 1 to p di tilde square less than alpha square. In this case, we have more options to pursue because there is some leg room over here that you can possibly choose different yi's which of, all of which will satisfy this inequality constraint.

So, for this case, what we will do is, we will follow kind of the opposite approach. So, we will follow the opposite approach. That is we will solve the unconstrained problem first, just minimize Ax minus b square, and we will check whether that satisfies the constraint, which is that B times y minus d tilde l2 norm square less than or equal to alpha square. If it satisfies that constraint, then we are done. If it does not satisfy the constraint, then we will have to do some more work.

(Refer Slide Time: 9:50)

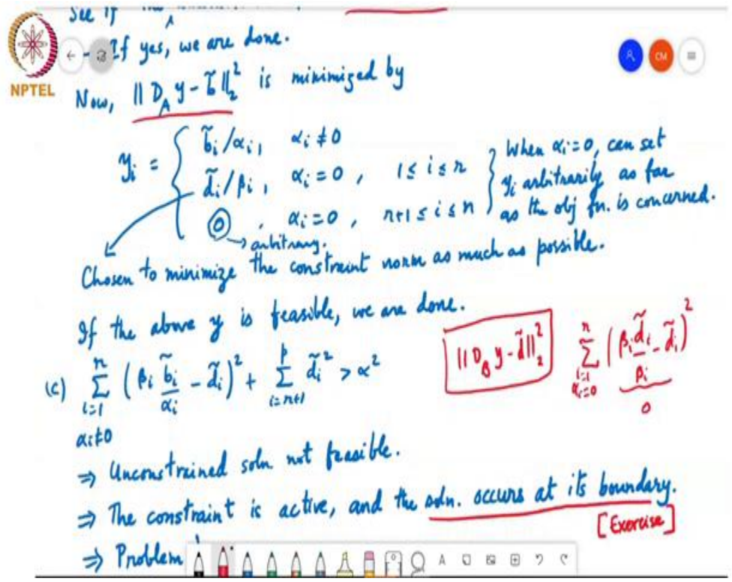


Now, if I look at DA y minus b tilde square that is minimized simply by choosing bi tilde, yi to be equal to bi tilde our over alpha i, for all the alpha i which are nonzero. If alpha i equals 0, I know that I do not care what why i chose, so I might as well choose it to minimize the constraint part which is di tilde over beta i. This is true for i equal to 1 to r.

And otherwise, if alpha equals 0 and it is beyond r i is r plus 1 to n, then I can arbitrarily choose yi equals 0. So, as far as the cost function is concerned, if alpha equals 0, I can choose yi arbitrarily. So, for i equal to 1 to r, I will choose yi to be di tilde over beta i to minimize the constraint as much as possible.

So, with this solution, we check whether this, with this y we check whether DB y minus d tilde square is less than alpha square, if it is true, if so, then we are done, but if not, which is this condition here. So, this is just DB times y minus d tilde a square and these are the other terms of y which do not affect that constraint part. So, if this itself is greater than alpha square, then the unconstrained solution is not feasible. So, we need to actually do some more work here.

(Refer Slide Time: 11:20)



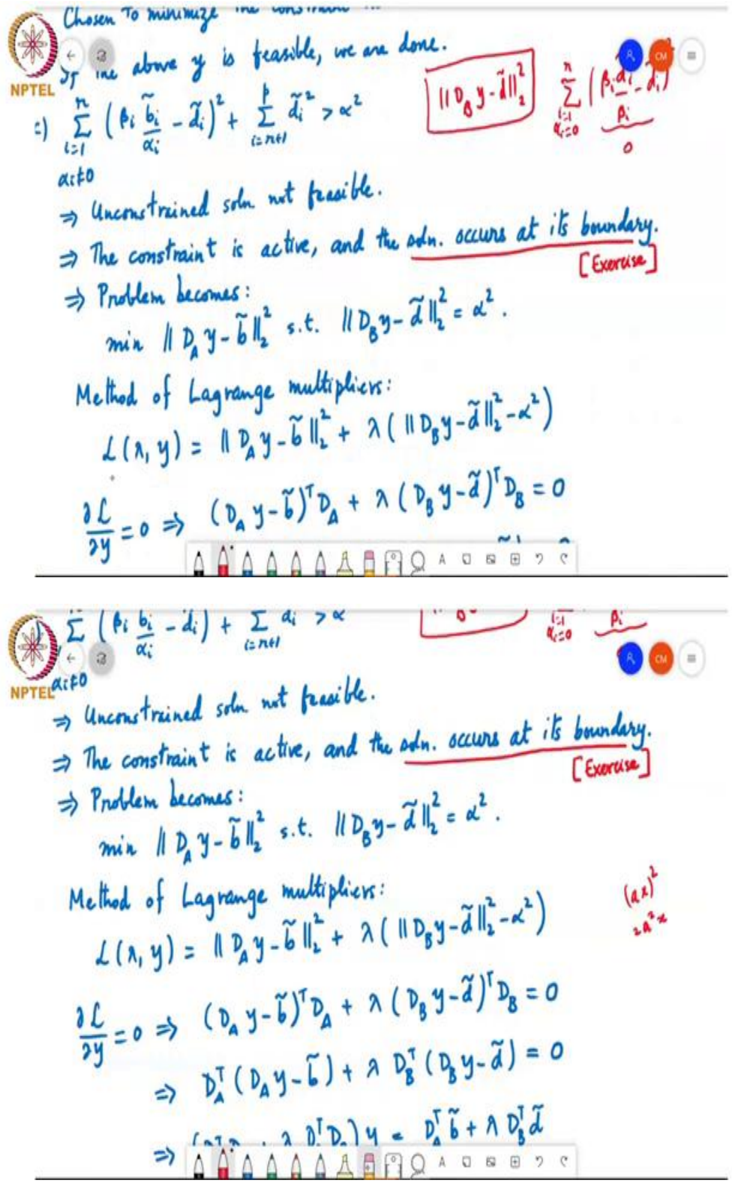
Now, so, here actually, there is another term, if I look at this, just for clarity, if I look at DB y minus d tilde square, then there will be 3 terms, there will be a term like this alpha i not equal to 0, i equal to 1 to r, and then there will be a term sigma alpha i equal to 0 and i going from 1 to r of beta i, but in this case, I am going to choose the yi to be di tilde over beta i minus di tilde square, that is this part here, second case, alpha i equals 0, and i going from 1 to r.

So, if alpha equals 0, this is the y but then you see that these cancel and this is equal to 0. So, that term I am not writing here. But if this is greater than alpha square and the unconstrained solution is not feasible, so, the constraint is active. And in this case, this is a small exercise you have to show this.

The solution actually occurs at its boundary, that is the solution will occur when this constraint is met with equality. The way you show this is that, if you are able to find a solution such that this summation, this part here di, DB y minus d tilde square is less than alpha square, then you will be able to find a solution, a better solution that is, it means you have head room, which you can utilize to further reduce the original cost, which is DA y minus b tilde l2 norm square.

So, the solution occurs at the boundary. So, which means that we need to solve this problem minimize DA y minus b tilde square subject to DB y minus d tilde square is equal to alpha square.

(Refer Slide Time: 13:31)



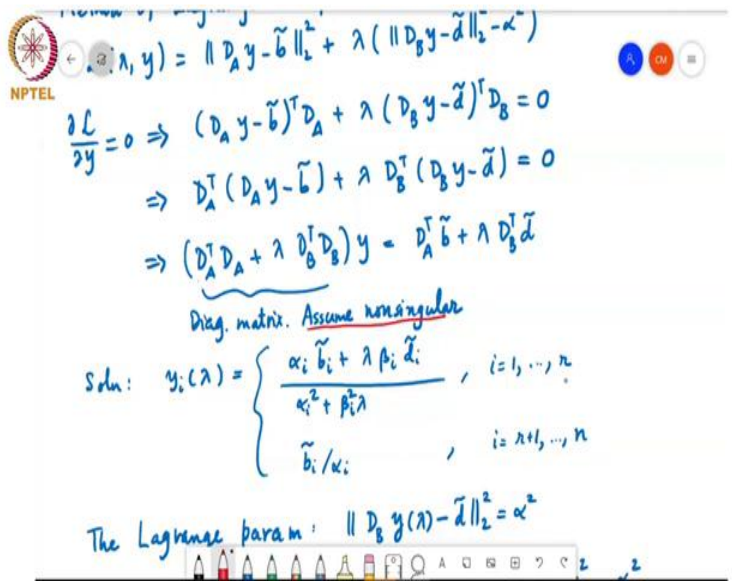
So, we can utilize the method of Lagrange multipliers where we write the Lagrangian function L of lambda y to be DA y minus b tilde l2 norm square plus lambda times DB y minus d tilde square minus alpha square. Once again, differentiating this with respect to y, we just get DA y minus b tilde transpose times DA.

This is again vector derivatives. And these are actually very like the very first or second result you will see, if you start, if you look at vector derivatives, it is very easy. It works similar to scalar derivatives, but you have to keep track of these transposes that show up when you take vector derivatives.

So, it is something like, if I differentiate x square, I will get 2 x I am dropping the 2 part here and writing it as just x. So, if I differentiate Ax the whole square I will just get 2a square times x. So, that is the DA times DA type of term that is coming up. So, coming back to this, this is DA y minus b tilde transpose times DA plus lambda times DB y minus d tilde transpose times DB.

Now, just take the transpose of this, I get DA transpose DA y minus b tilde plus lambda DB transpose times DB y minus d tilde, lambda is just a the scalar here, like Lagrange multiplier. So, combining the terms that involve y, I will get DA transpose DA, plus lambda DB transpose DB times y is equal to I am taking these two terms to the other side, I will get DA transpose b tilde plus lambda DB transpose d tilde.

(Refer Slide Time: 15:35)



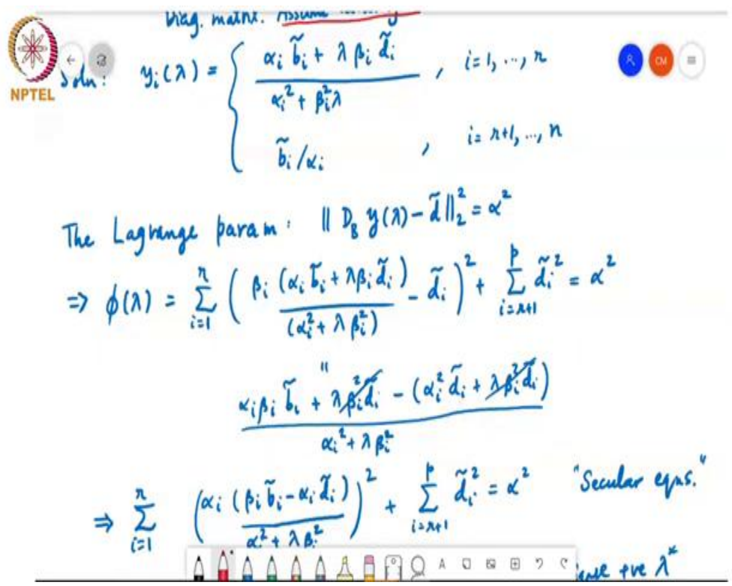
So, now this is, these are just diagonal matrices. And let us assume that this is non singular. In fact, the singular case is also easy to handle. The other thing is, you note that this lambda here is actually a parameter. So, I can choose lambda to be different values. And try to make this non singular. If in spite of choosing different lambdas, if this is singular, then you will have to handle that separately.

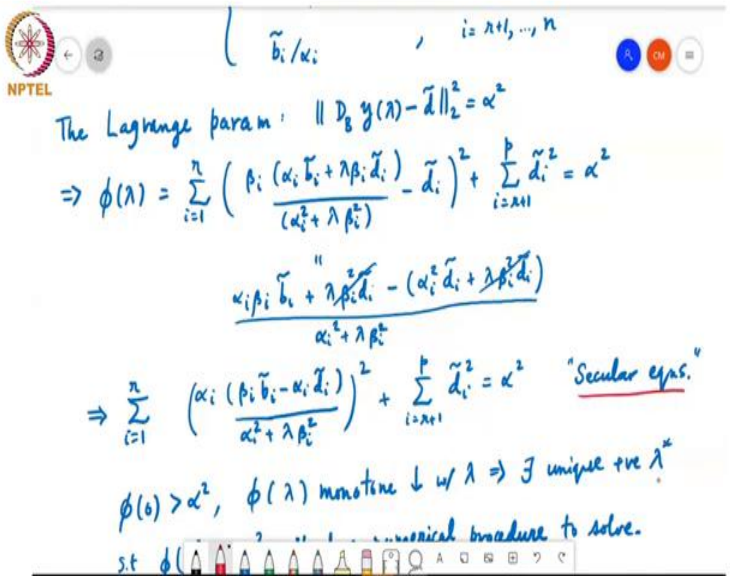
But for the moment, let us assume that this matrix is non singular. In that case, I can just directly write out what yi of lambda is, this just a diagonal matrix, it is just the diagonal entry of this, which is alpha i times the ith entry of this is alpha i, bi tilde plus the ith entry of this is beta i times di tilde divided by the ith entry of this is alpha i square. And the ith entry of this is lambda beta i square.

And this is true for i going from 1 to r, beyond i equal to r, the beta i are equal to 0, so I am just left with bi tilde divided by alpha i, so i equal to r plus 1 to n. So, and how do I find this lambda? I need to solve for the lambda which satisfies the constraint DB times y of lambda minus D tilde square is equal to alpha square.

So, if I write, if I just expand this out and call this function, phi of lambda, I will just get i equal to 1 to r beta i times yi of lambda, this quantity here, minus di tilde square plus i equal to r plus 1 to p. For those terms, this y of lambda will multiply 0 and so I am just left with di tilde r square is equal to alpha square.

(Refer Slide Time: 17:23)





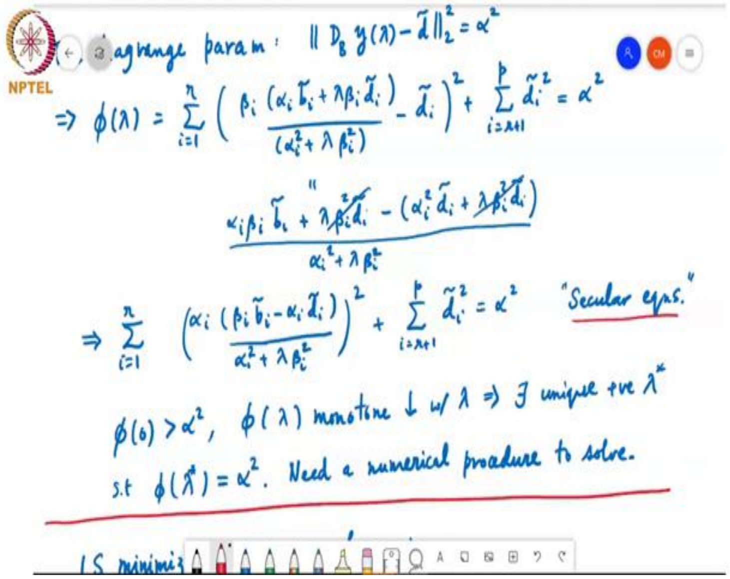
And if I take this alpha square plus lambda beta square common between these two terms, and in the numerator, I will get alpha i beta i times bi tilde plus lambda beta i square because there is beta i here, times di tilde minus di tilde times this denominator, which is alpha i square di tilde plus lambda beta i square di tilde are these two of course cancel.

And so what I am left with is sigma i equal to 1 to r alpha times, alpha is common between these two terms, I have taken it out, alpha i times beta i bi tilde minus alpha i di tilde divided by this quantity alpha i square plus lambda beta i square square plus summation i equal to r plus 1 to p di tilde square equal to alpha i, alpha square.

These equations are called, this equation is called the secular equation, the name apart, the point is that lambda is sitting in the denominator here, and you have a sum of r terms here. So, you cannot solve this in closed form. But when I said lambda equal to 0, phi of 0 is actually the solution I would have obtained by solving the unconstrained problem.

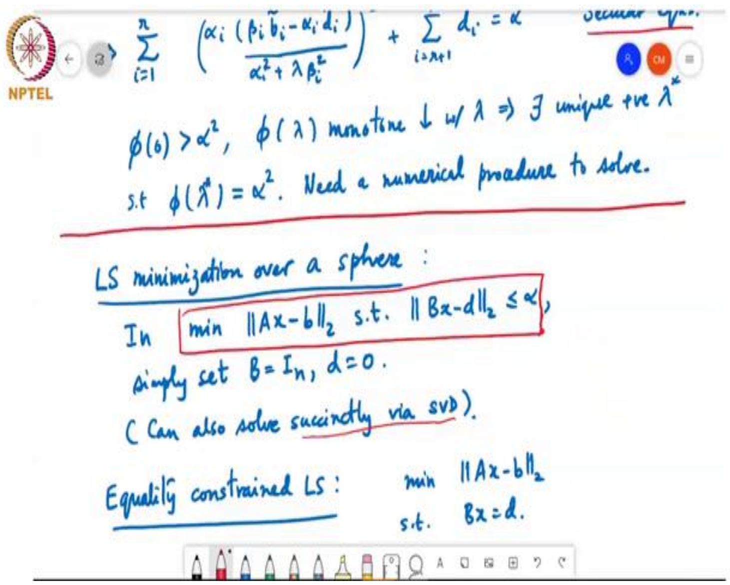
And we know that phi of 0 is, I mean that is why we came to, came into all this because phi of 0 must be bigger than alpha square. And because lambda is sitting in the denominator here, phi have lambda is actually a monotone decreasing function of lambda. So, as I increase lambda this phi of lambda will keep decreasing and then at some point, there will be a unique positive lambda star such then phi of lambda star is exactly equal to this alpha square here.

(Refer Slide Time: 19:09)



But you will need to use some numerical recipe to solve this equality condition. So, that completes my description of how to solve least squares problems with quadratic inequality constraints. So, we just learned how to solve this problem.

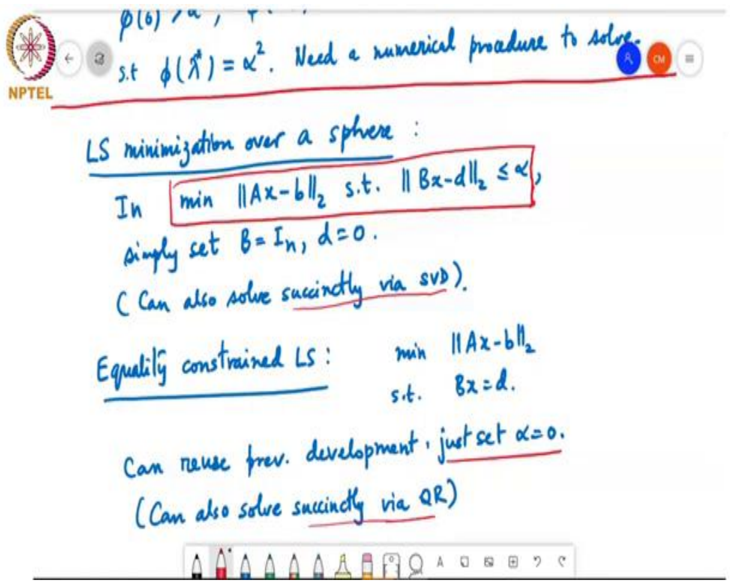
(Refer Slide Time: 19:40)



Again, then, knowing how to solve this is actually very powerful, because so for instance, if you want to perform least squares minimization over a sphere, that is, you can do that by just setting B equal to the identity matrix, and D equal to 0, then it just becomes norm x2 less than or equal to alpha.

So, you want to minimize the l2 norm error between Ax and B subject to norm x less than or equal to alpha. There is also another succinctly solution you can obtain by SVD. But I will not discuss that here. Clearly if you want to do, solve an equality constraint least squares problem that is minimize Ax minus b l2 norm subject to an equality constraint Bx equals d. Then all I have to do is to set alpha equals 0 here.

(Refer Slide Time: 20:29)



So, I just set alpha equals 0 and I can use the solution developed here to solve it. But there is also another succinct solution via QR that I will not discuss here. So, basically this brings me to the end of what I wanted to discuss in this course. Are there any questions?