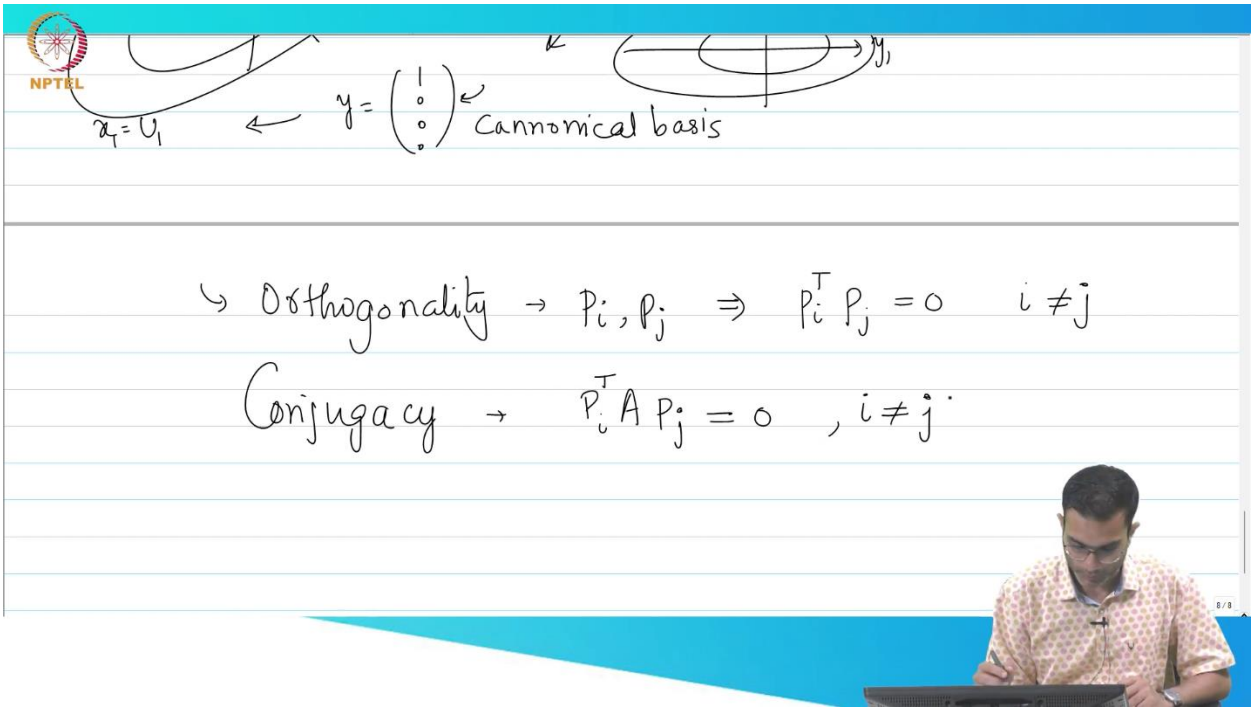
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 05
Lecture - 35

Orthogonality and Conjugacy

Now as I mentioned at the start of this module that we are going to try something a little bit more clever than just going along a descent direction. So, to build up to that we want to generalize our ideas particularly of orthogonality ok. So, when we said orthogonality let us say of p_i and p_j, what did it imply? If I said that p_i and p_j were orthogonal very simply it just means that



$$p_i^T p_j = 0, \quad i \neq j.$$

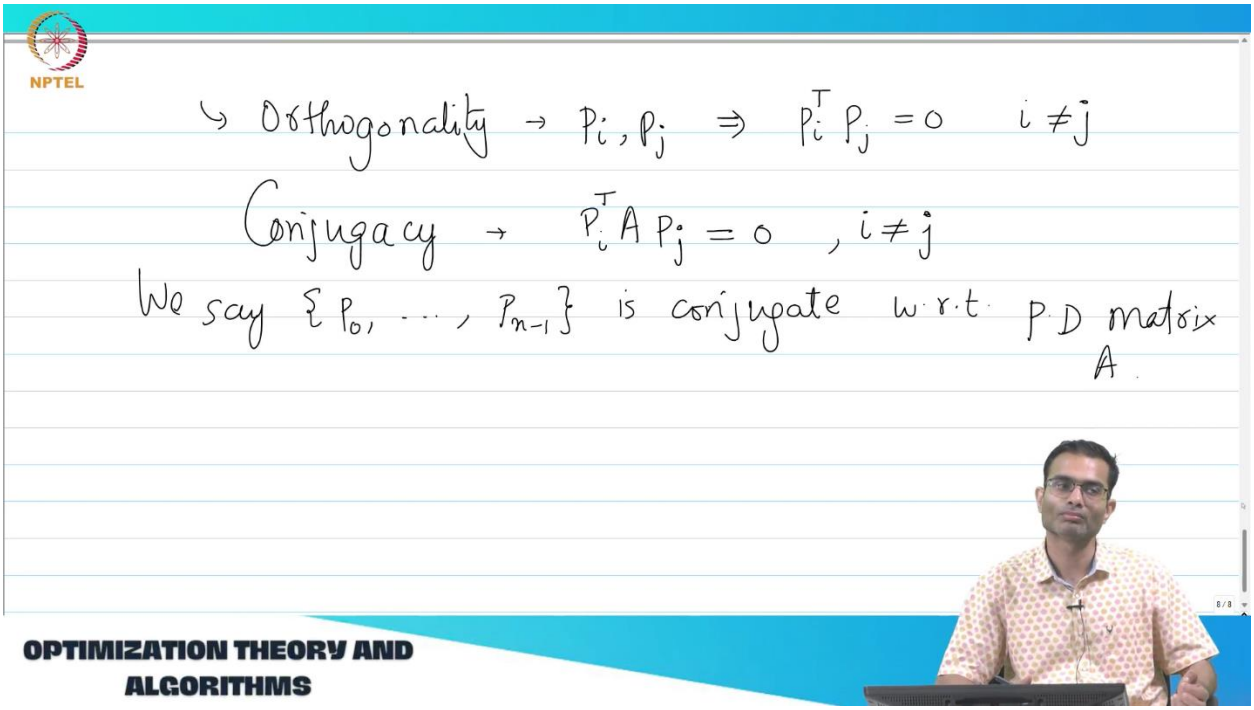
Now, very similar to this property is a property called conjugacy. The first c of the name of the method has the word conjugate gradient.

What does conjugate mean? So, conjugacy simply means that

$$p_i^T A p_j = 0 \quad \text{for all} \quad i \neq j.$$

If this is true, then we say that these two vectors p_i and p_j are A-conjugate with respect to each other. So, if I take all of these p's starting from p_0 up to p_{n-1}, I have n vectors ok. We say that this set is conjugate with respect to a positive definite matrix. That is just a definition.

We are not saying anything about the case when $i = j$, whether it should be 1 or something else that really does not matter. All we need is that it should be not orthogonal, but conjugate with respect to A. Now, you might ask, "So what? I have defined another kind of a look like a generalized what?" Can I write this expression in terms of a norm? Just to brush up linear algebra you had it in the quiz as well right. If A is positive definite, is the square root of A defined? What is it?



$$A^{1/2} = U\Lambda^{1/2}U^T$$

that is the eigenvalue decomposition, the square root just comes like this. So, I could write this and let us call this Q, right.

So, then if I define the norm with respect to this, what happens over here, right? I am going to get $x^T Q$, wait, wait a minute. So, I do not actually need to take square root of Q, ok. So, let us just call this A. Kind of like generalizing a product also.

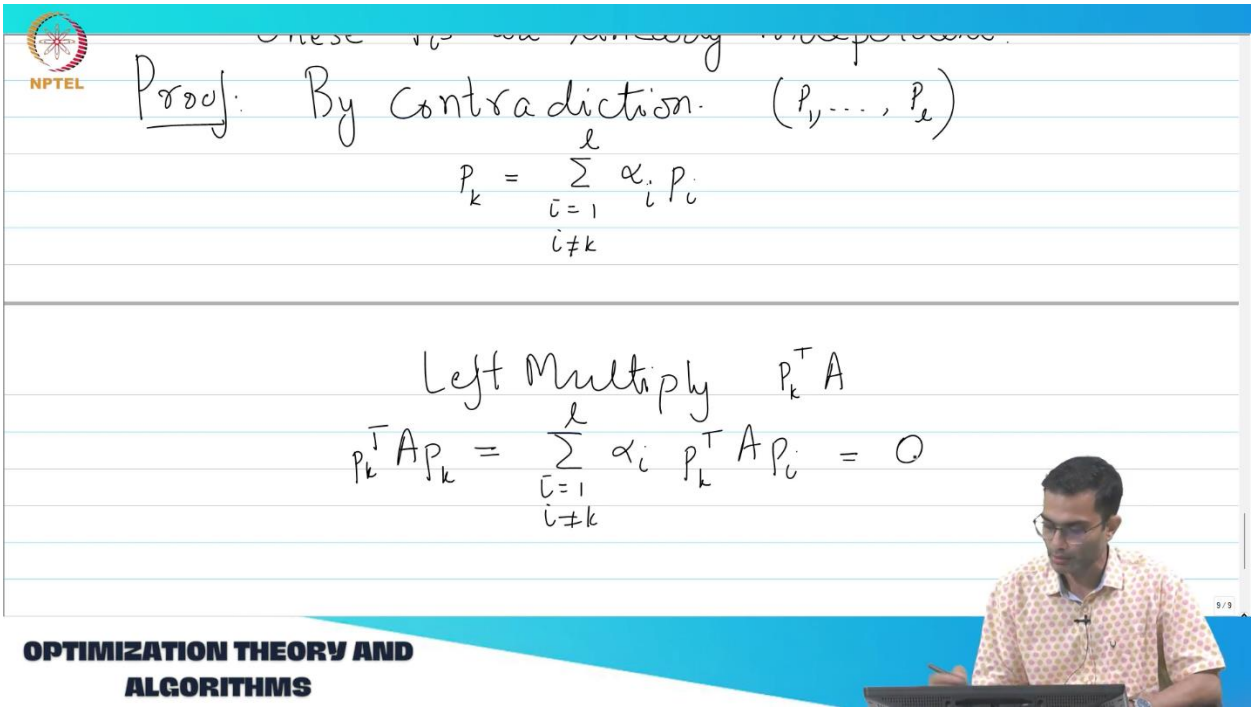
Anyway, this is not important, let us scratch that. So, getting back to these newly defined conjugate vectors, first question is, who cares? What is so special about them? Now, it turns out that if these n vectors are conjugate with respect to a positive definite matrix A, there is a surprising property that comes which is going to be the basis of the conjugate gradient method. So, let us note it down. If these vectors are conjugate with respect to A, it turns out—any guesses?—it is surprising, you would not expect it. These p_i's end up being linearly independent, right? Not at all clear from the definition. These p_i's are linearly independent because it is not satisfying to just have this property listed out like this. We want to know how this happened, right.

So, when we were doing proofs, what was one of the first tricks in the bag to prove something like this? Contradiction, right. So, let us see if our old friend helps us out over here ok. So, I am

going to take, let us say I am going to take l such p_l-vectors. Now, what does proof by contradiction mean here? What is the contradictory statement? They are not linearly independent. That means they are linearly dependent. That means I can express what? I can express any one of the vectors as a linear combination of all the others. That is the meaning.

So, let us put that down, right. So, that simply means that if I take, say some p_k, this is going to be written as a linear combination of what? p_i. This is almost correct, but what should I do in this summation? $i \neq k$, because it does not make sense. I mean, I want to write p_k in terms of all other vectors. So, I will just exclude k from the summation. If this is the case, then p_k is linearly dependent.

It can be written in terms of other vectors, and that is what we want to, right. So, this is what we want to see. Does this lead to a contradiction? What would your next step be? What do you know about these p's? I need to somehow get A into action, right? So if I want to get A into action, what do I do? Left multiply by A and then left multiply by some other p right to get this into action, right. So, the best thing to left multiply would be.



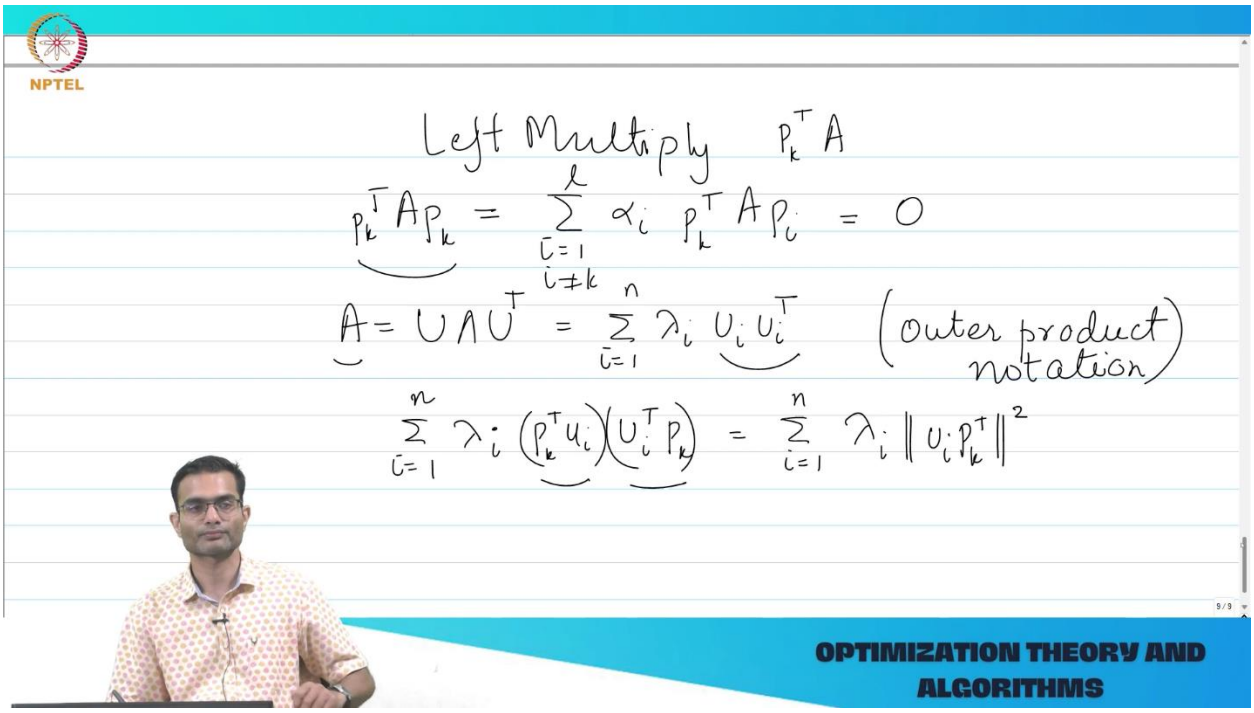
So, left I am going to put up $p_k^T A$. If I stick this in it is going to be great, right, because what will the left-hand side be? I am going to get $p_k^T A p_k$. What happens to the right-hand side, right? So,

$$\sum_{i=1}^{l, i \neq k} \alpha_i p_i.$$

So, the property of conjugacy is going to give the right-hand side to be obviously 0 because there is no p_k left to act on, this is going to be 0, ok.

Now, if this is 0, let us open this up. Now, what is the best way of doing this? For example, A, I could have written it in terms of $U\Lambda U^T$. I can also write it in the outer product notation. Does anyone remember what the outer product way of writing it is?

$$A = \sum_{i=1}^{n} \lambda_i U_i U_i^T$$



That is one way of writing it, right. It does not matter, I mean, you can write it like this also. Let us open this up. So, this is also equal to

$$A = \sum_{i=1}^{n} \lambda_i U_i U_i^T.$$

Is everyone familiar with this? You should have seen it during linear algebra. This is called the outer product. What is nice about it is that I have written a rank-n matrix as the sum of n rank-1 matrices. What is the rank of A? If it is positive definite, it is rank n. What is the rank of $U_i U_i^T$? It is rank 1. Why? Because every row is a multiple of each other. Therefore, it is just rank 1.

So, I am writing a rank-n matrix as the sum of n rank-1 matrices. This is actually very, very useful in things like dimensionality reduction and so on. In image processing, these kinds of ideas are used heavily, ok.

Now, if I open this up over here, now if I stick $p_k^T p_k$ on both sides, what am I going to get? I am going to get a summation

$$\sum_{i=1}^{n} \lambda_i \, p_k^T U_i U_i^T p_k,$$

which is nothing but λ_i times the norm squared of the vector. Why? Because I see the same vector with the transpose. So, therefore, this can be written as

$$\lambda_i(\| \, U_i p_k \, \|^2).$$

If I take the transpose of $U_i^T p_k$, I get $p_k^T U_i$. So, therefore, this is the norm squared of $U_i p_k$ transpose squared.

Do you see the contradiction now? It is a positive definite matrix. Therefore, the λ_i's are strictly greater than 0. This is a norm square, right?

I am obviously assuming that the p's are not 0, that p_k is not 0, that is not a useful thing to consider. So, this is also greater than 0, yet this summation is equal to 0. So, as simple as that. This is a contradiction.

Any questions on this? So, most of what we have—not most, entirely everything—that we have done so far has just been simple linear algebra starting with eigenvalue decomposition, and what we have proven is an interesting result: if you give me the conjugacy condition, it turns out that the p's are going to be linearly independent. That is interesting. That is probably... Have you encountered any other way of generating linearly independent vectors? Gram-Schmidt is one process.

Gram-Schmidt actually went several steps ahead, right? It made them orthogonal to each other, right. Here, you give me an input matrix A, I am not telling you the algorithm, but there exist n linearly independent vectors which satisfy this property, ok. So, that is something new.