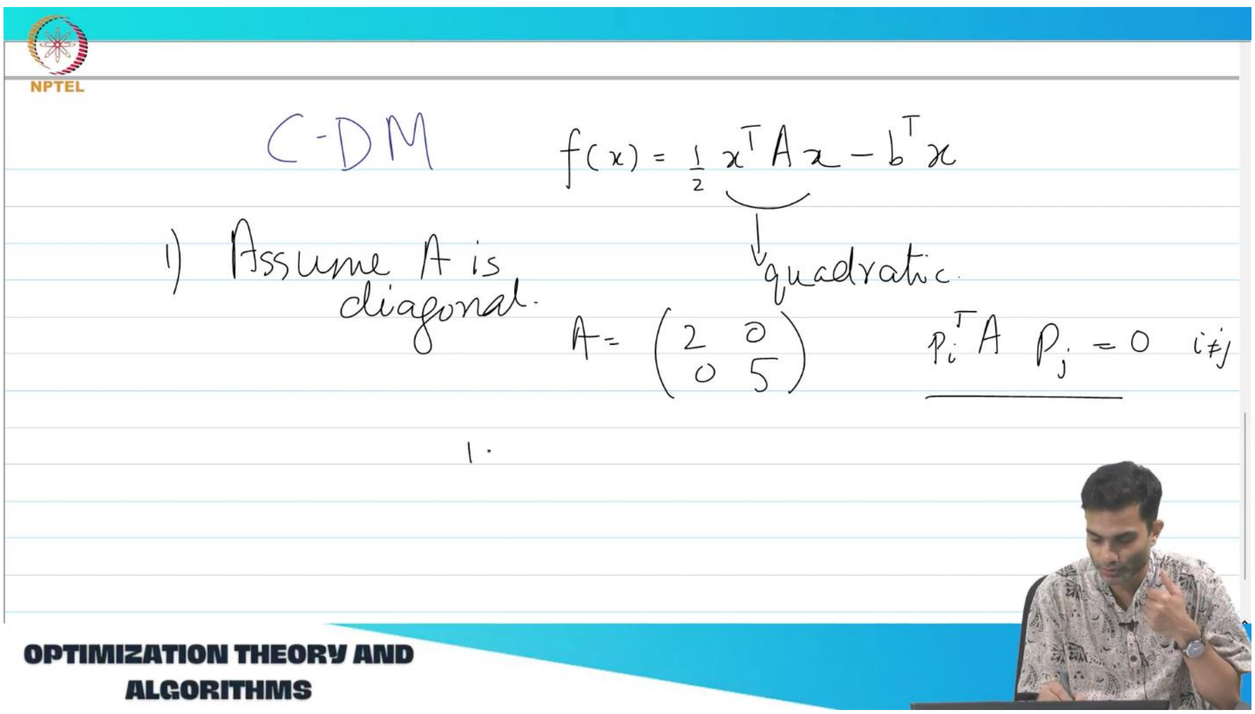
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 06
Lecture - 38

More on Conjugate Directions Method

So, we are doing CDM for conjugate direction method first, then we will get to conjugate gradient method, ok. So, the cost function which I had, which was half of $\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x}$, if I want to visualize this cost function, of course there is the $\mathbf{b}^T\mathbf{x}$ which is simply a linear part. Let us have a look at the more interesting part, which is $\mathbf{x}^T A\mathbf{x}$, the quadratic part, right. So, let us see, ok. And to help us visualize things as quickly and easily as possible, I am going to make life very simple. I am going to assume A is a diagonal matrix, ok.

A is diagonal, ok. Fine. So, this is my very simple straightforward matrix, ok. And let us look at the conjugacy condition.



What is the conjugacy condition? $\mathbf{p}_i^T\mathbf{p}_j = 0$, $i \neq j$, ok. This is a diagonal matrix. Can you immediately guess one set of $\mathbf{p}$'s such that this is satisfied? $\{1,0\}$ and $\{0,1\}$, right. So, supposing I have $\{1,0\}$ for this matrix A and $\{0,1\}$, you can see that this is going to be equal to 0, right. Also, the other way of seeing it is what are the eigenvectors of this matrix? $\{1,0\}, \{0,1\}$, because it is a diagonal matrix and the eigenvalues are 2 and 5, right.

So, automatically I can see that $\{1,0\}, \{0,1\}$ are the eigenvectors of this situation, ok. Now, in this case, if I plot $\mathbf{x}^T A\mathbf{x}$, what is the kind of geometric figure that I get? We did this a couple of classes ago. If I plot $\mathbf{x}^T A\mathbf{x}$ in the x_1x_2-plane, what do I get? I get an ellipse, right. Now, if I look at the conjugate direction method, what is it saying? The first direction in which I must go is which direction? The first conjugate direction $\{1,0\}$, for example. So, let us call this $\mathbf{p}_0$ and let us call this direction $\mathbf{p}_1$, ok.

So, $\{1,0\}$. So, let us say I started from here. Some random starting point. What is the solution to this point? Assuming, well I do not need to assume, this is a convex function. The red dot is the solution.

It is the global minimum to this problem. What is my first move going to be? Along the x_1-direction. The second move is going to be along which direction? x_2-direction. So, if I want to get to the red dot and the first move is only along the x_1-direction, where do I move? I move horizontally and I move where? Hit the x_2-axis, right. So, the first step is going to be actually boom here, right.

The second step is going to be only along the x_2-direction, there is no choice, where do I go? Straight down to the solution, right. So, you notice that my emphasis here was to look at the underlying geometry which was the conjugate directions, go along the conjugate direction. I did not see whether, you know, how perpendicular I am to the contours of the cost function, which I did in the gradient descent method, right. If I were doing gradient descent method, what would I do? I would do something like this. Steepest descent means go in the direction that is along the negative gradient, but I did not do that.

So, my motivation is to cover the ground in each of the directions so that I reach here. So, you can see that in two steps I am at the solution. And you can also see that if you got lucky with a starting point, supposing I started along the x_1-axis or along the x_2-axis, what will happen? One shot I reached that because there is now no other ground to be covered in the next step, right. The other case, which is quite simple to understand, is if A is non-diagonal, ok. So, if A is not diagonal, we also did this, yes, question.

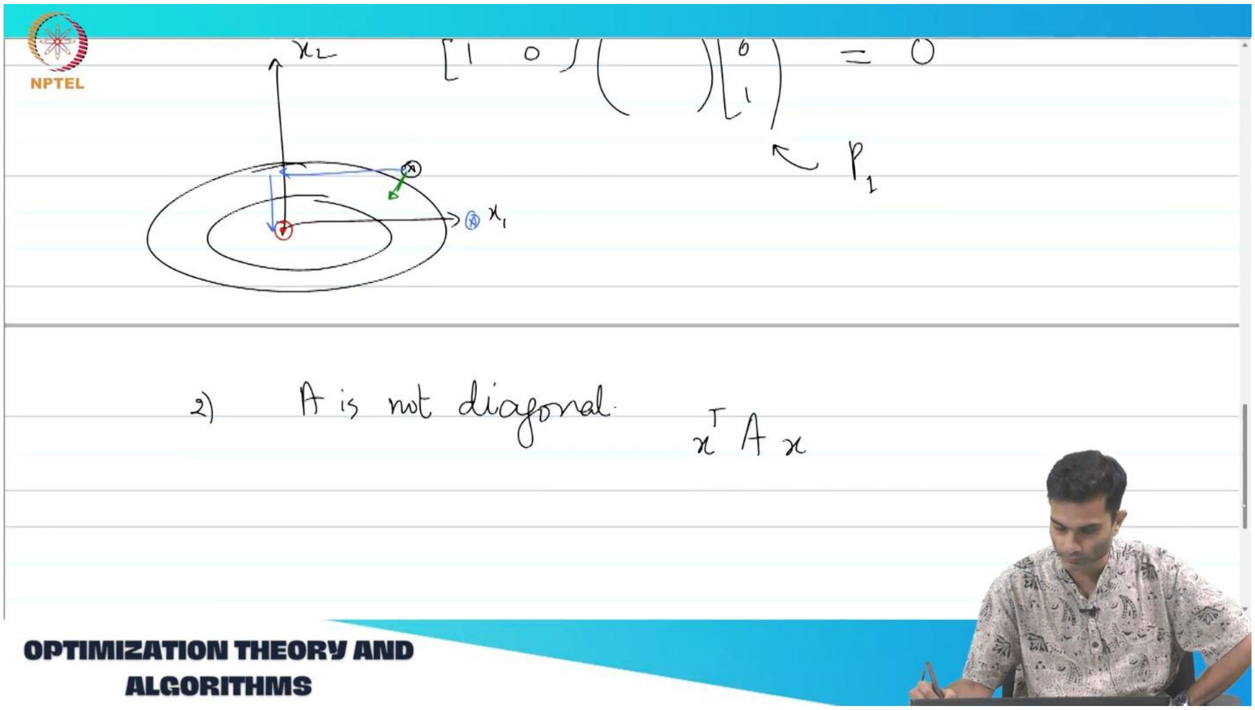
Is there a clever way to choose an initial value? Can someone answer that question? What is the cleverest way of choosing an initial value? Why is it necessary? Okay, so if I were to like in this example, if I chose my point to be along x_1 or x_2, I am saving one step. So, if I get somehow my luck or my skill has to be in figuring out these directions and going along. Ultimately, it may not matter so much its order. I mean there are only n steps so I will get there. Okay, when A is not diagonal how do we visualize? We have seen it is a rotated ellipse, right. But that was, what was the rotation with respect to? With respect to the eigenvectors, right.

But now I want to visualize in terms of what? The $\mathbf{p}$'s. I am given the conjugate direction. I want to work with conjugate directions. Eigenvectors happen to be one candidate, ok. So, this expression over here I can insert identity here and here in a clever way, right.

So, identity I can write as. So, let me take my P-matrix as $\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{n-1}$. This is my P-matrix. I got all my conjugate directions over there, ok. What would be a nice way of writing identity? I can write it in terms of P and P^T because it is invertible, right.

So, if I insert it in this way I can write $PP^\top$ or I can write it as PP^{-1}, the same thing, ok. Actually, it is not the same thing, they are not orthogonal. So, if I give me P, let me write it as PP^{-1}, that is identity, ok. So, it is not transpose. So, what will this become? This will become $APP^{-1}x$.

Can you guess now what I should do for the identity on the other side? There is going to be a transpose operation over there. I can also write identity as $P^\top P^{-1}$ is also a legit way of writing identity, right. This guy, this guy I can insert over here. So, what am I going to get? $P^\top P^{-1}x$, ok. Now, this guy, if I call this guy to be y, I am going to move this over here, this is also correct right, and there is a transpose missing over here.

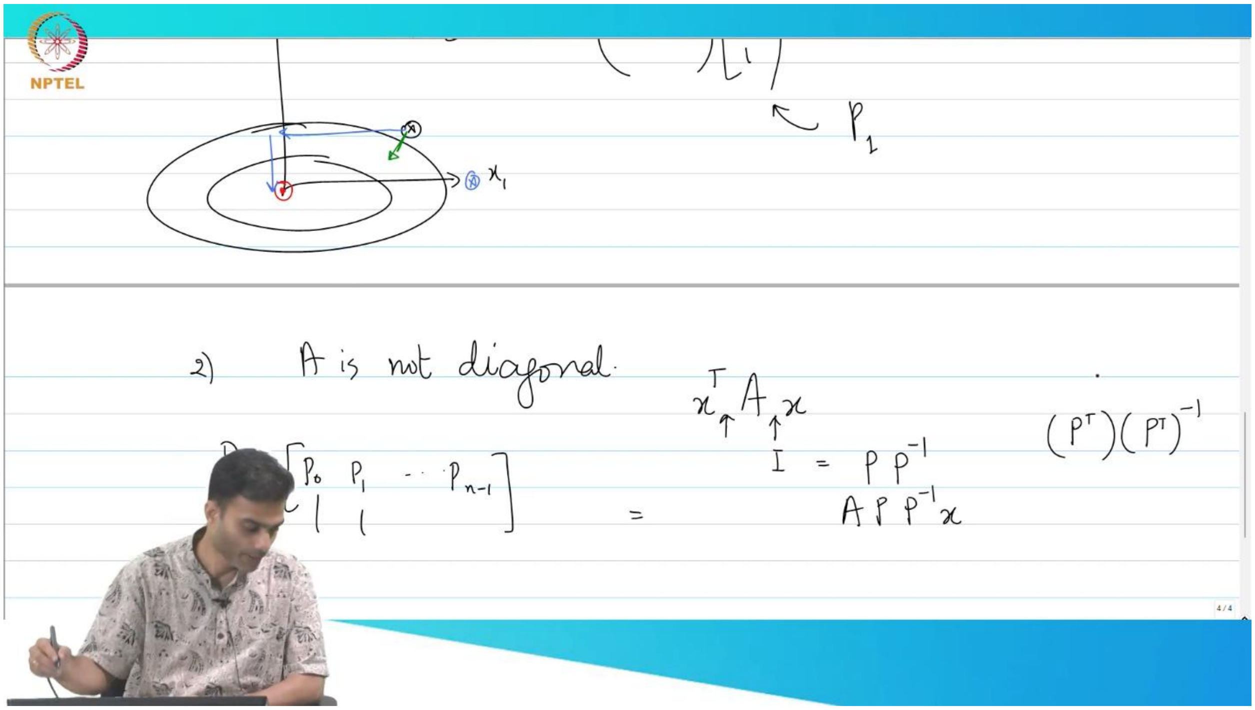


So, this term over here, what is it? If $y = P^{-1}x$, what is $y^\top$? $x^\top P^{-1\top}$. So, this becomes $y^\top$, ok. Now, what we are left with is this matrix $P^\top AP$, where P is, you know, what the matrix over here. Can there be anything special about this matrix? Is it necessary? These are not eigenvectors, these are only conjugate vectors. So, is it necessary for it to have λ? Is diagonal.

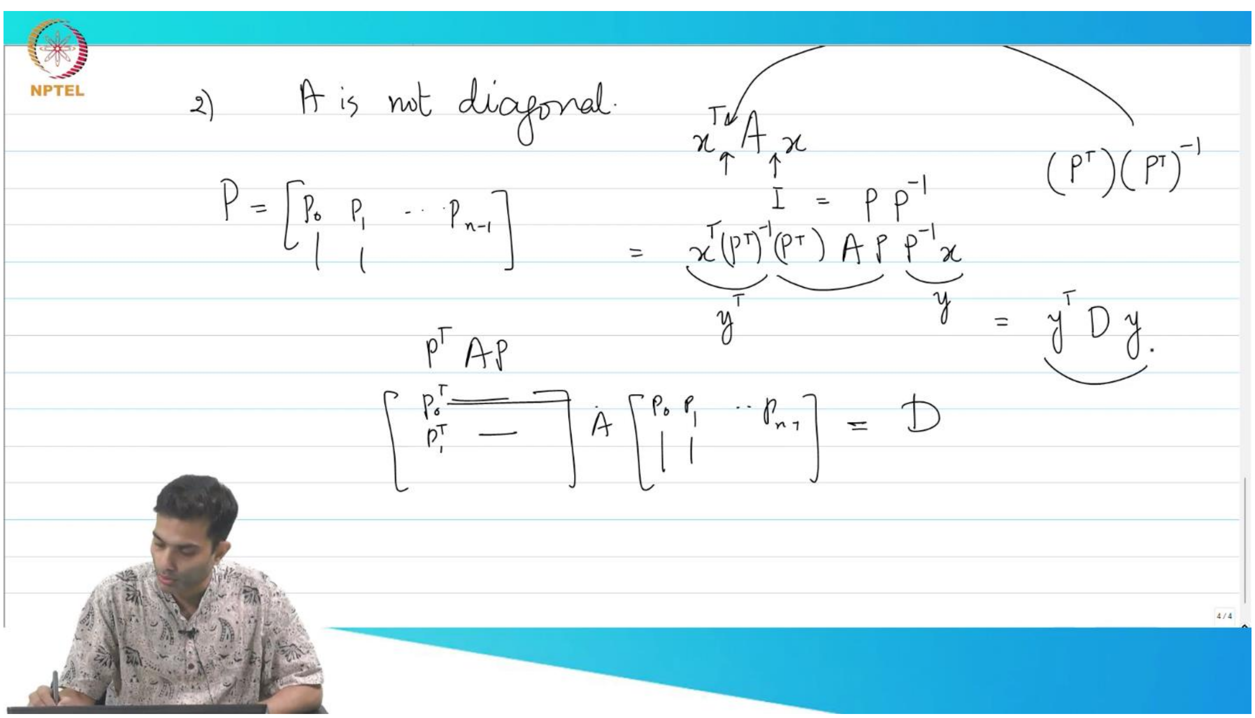
Why is it diagonal? Because if you look at $P^\top$, what is $P^\top$? p_0 like this, $p_1^\top$ like this, right. This is my $P^\top$, this is my A and this is my p_0 this way, p_1 this way, up to p_{n-1}, right. So, when I take this multiplied by A multiplied by any other column, I am going to get 0, right. $p_0^\top Ap_2$, what is it? 0. The only term that is going to survive is $p_0^\top Ap_0$, that is 1 1.

Similarly, $p_1^\top Ap_1$, that is the only term that is going to survive. So, this is actually going to be a diagonal matrix. So, this is a little bit more clever because I am not insisting on eigenvectors, right. I just need A-conjugacy and I am getting this to be, therefore, this is $y^\top Ay$, not Ay, Dy, right. This entire expression simplified over here $y^\top A^\top$ sorry $P^\top AP$ became Dy. What is this looking like now? This looks like case 1, right. In the y-coordinate axis, what is it? If I what shape is it in the y-coordinate axis? An ellipse, right. So, if I were to plot this along $y_1 y_2$, I am going to get ellipses, right. And we already have $y = P^{-1}x$ implies that $x = Py$, ok.

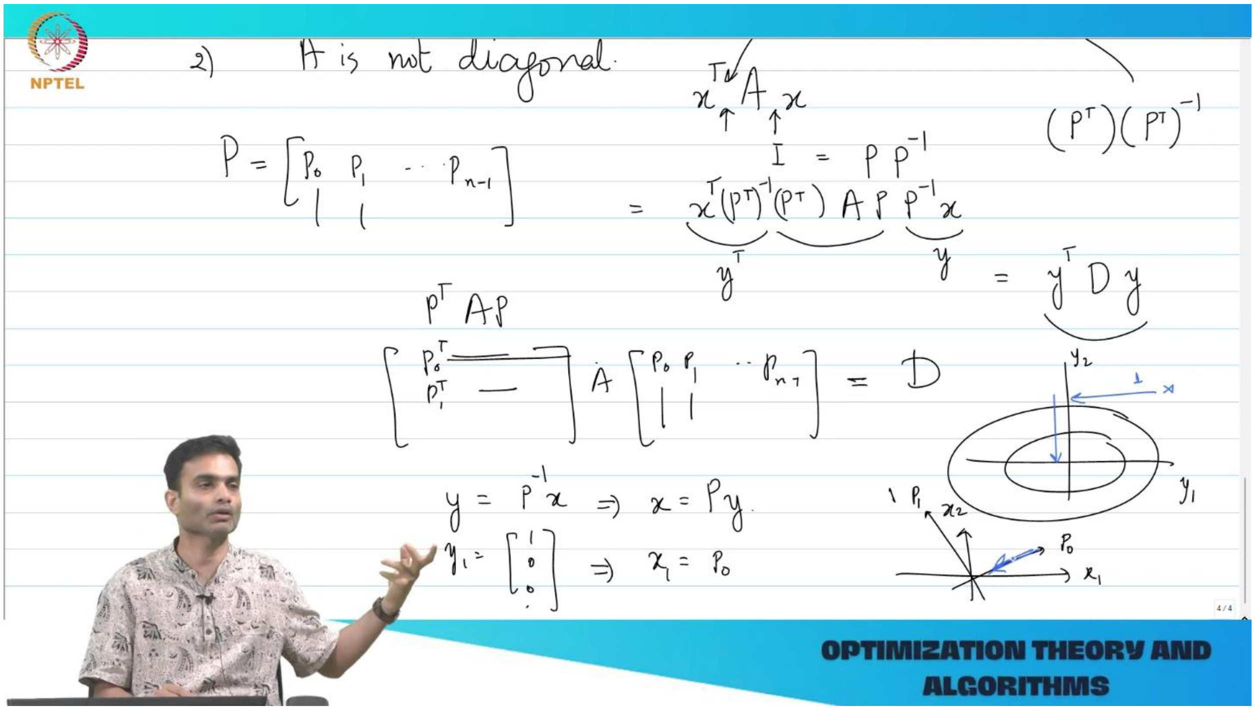
So, in the y-coordinate frame, the easiest way to solve the problem is to, so supposing now I am solving this problem in the y-coordinate axis, this is exactly like case 1, right. First do this, then do this, right? So a step along the y_1-axis is what step in the x-axis, in the x-space? So let's take y_1 is going to be given by $(1,0,0,0)$, right? This is my y_1. If I stick this into here, what am I



going to get x as? Which column of P will get extracted when I multiply by y_1? Only the first column. The first column is what? p_0, right. So, if I am walking along the y_1-axis in the x-x_1, x_2 let us call this p_0, need not be perpendicular to p_1, right. So, the first move over here is actually going to be along this direction. So, every walk along the coordinate axis in the y-space is a walk along the conjugate directions in the x-space. The next step that I walk is going to be y_2, which is $(0,1,0,0,0)$, that is in the x-space p_1 and p_2, p_n, right. So, every step in the y-axis along the coordinate axis is a step along the conjugate directions.



Right. So, is that clear? This is a very important, this is the geometrical intuition of the conjugate direction method, right. Going from non-diagonal to diagonal is trivial. A simple linear algebra inserted identity, we split it, we saw that what is happening keeping track of x-space and y-space, right. So now this gives you the motivation for, you know, whoever came up with this method probably, right. They saw this geometry and said, hey, why bother with descent directions?



If I know the solution is here, the best way I could, well, one clever way in which I could do it is walk along each step. You notice I am never retracing my steps. If I walk 5 centimeters along the x_1-axis, I never revisit the x_1-axis ever again. Whereas in the gradient descent method, it had a zigzag kind of a trajectory in that I am revisiting each direction multiple times and possibly at every iteration I am walking along each of the coordinate axis, right. But here it is very efficient in that way that just walk along the axis once.

At least that's my understanding of how the motivation would have been.