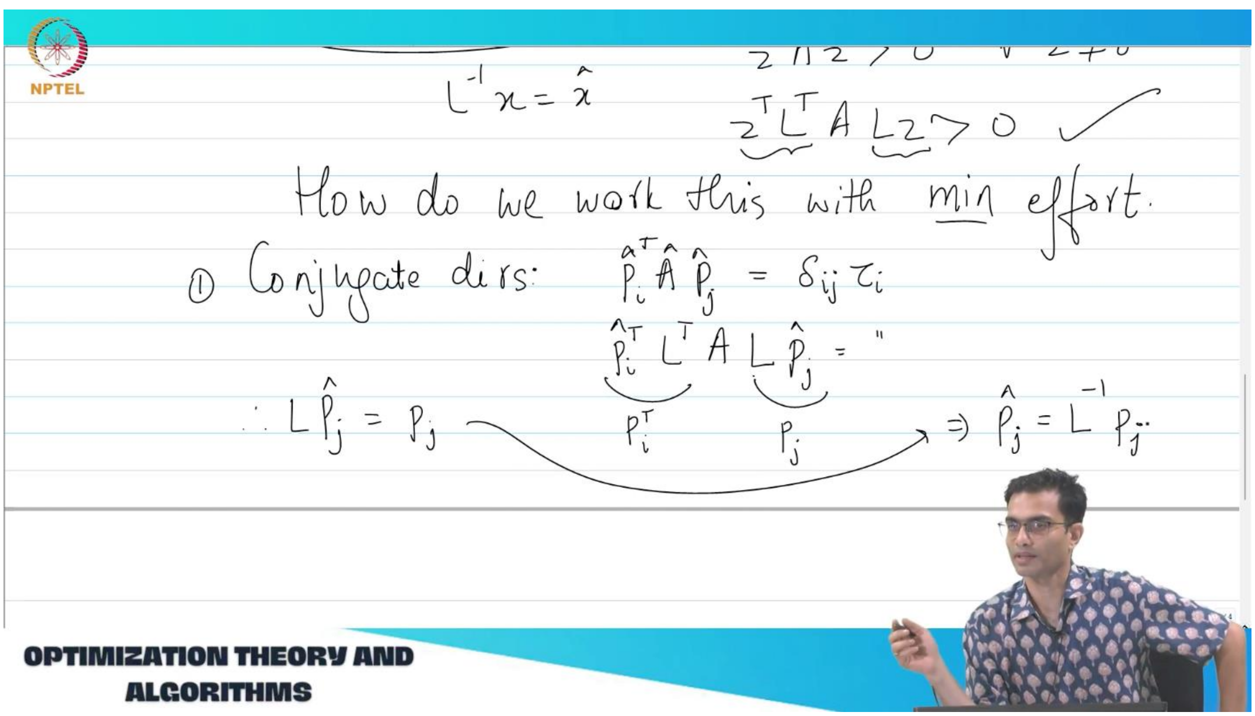
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 06
Lecture - 46

Preconditioned Conjugate Gradient - Part 2

So, how do we work this with minimal effort? Another reason, apart from minimal effort, is that many times you get code available, let's say on GitHub or from a collaborator or a colleague, right? You want to make use of it with minimum modification and editing, right? So this is another motivation for working this out, okay? So let's see how to work this out. So what do we need in a CG method? We need the starting point, which is my p_0. I need from p_0 to get to x_1, what do I need? α, from α I get my next x_1, then what do I need? I need to get the next p_1, to get p_1 what do I need? β, right? So, if I get a few of these parameters correct, I know the rest of the sequence. So, let us work it out correctly, okay?

So, the first thing is the conjugate directions. The conjugate directions, obviously, of $\hat{A}$ will be different from the conjugate directions of A; there is no reason why they should be the same, and we have to also ask whether or not they are conjugate, right? So, let us see what the relation is over here. So, $p_i^T \hat{A} p_j$ this obviously should be equal to δ_{ij}, that is what we expect. So, let us substitute over here. $\hat{A}$ was simply $L^T A L$, right? So, this becomes δ_{ij}, it is 0 if $i \neq j$, right? This is the Kronecker delta.



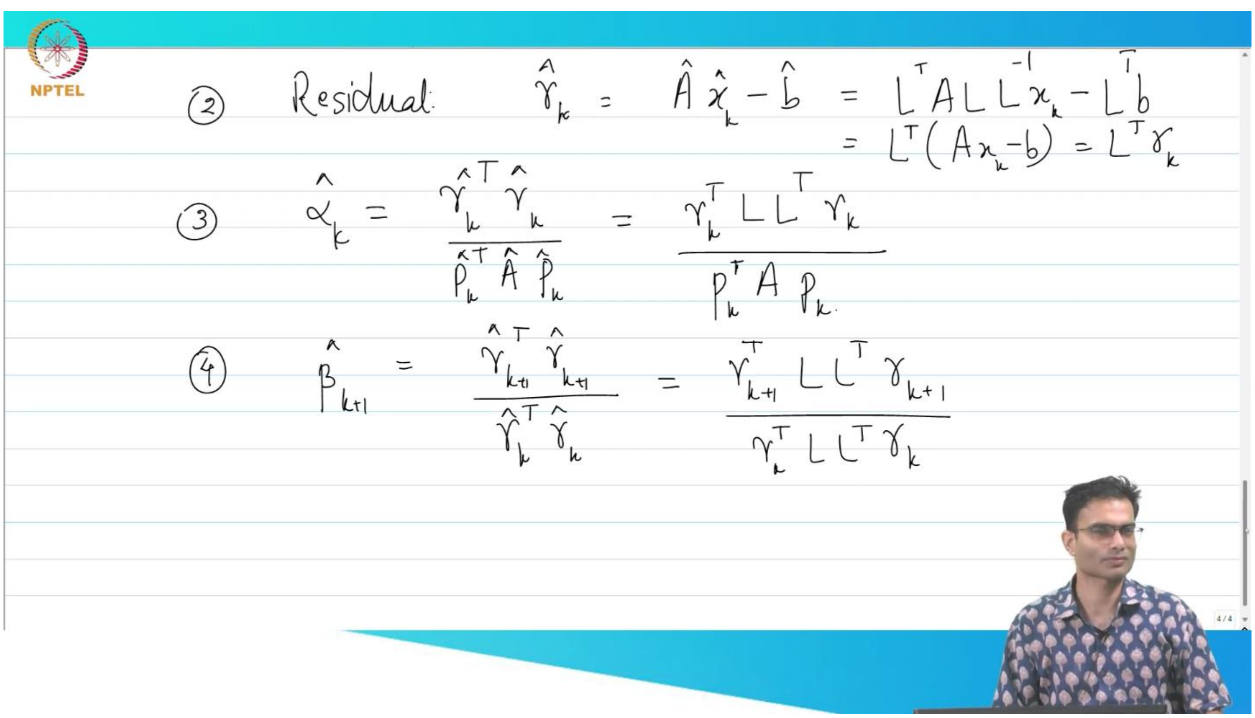
Are they orthonormal or not? They need not be; they can be some orthonormal. Are they orthonormal? I am saying they need not be. So, you can have some factor over here, τ_i, or something, it does not matter, not important, right?

So, if I just wanted to make sure that everything went ahead as before, what would be the most obvious choice for p_i^T? Right, if I made this equal to p_j and this equal to p_i^T, I will automatically make sure or ensure orthogonality and conjugacy, right? Therefore, $p_i^T L p_j = p_j$, okay? So, the new conjugate directions are simply L^{-1} times the old conjugate directions, okay?

So, that was one easy thing out of the way. By the way, it currently looks like I need to evaluate L^{-1}; you should be suspicious of any method that asks you to invert a matrix. So, later on, we will come back and see if we actually need to calculate L^{-1}. Right now it is just in symbolic form; I have written L^{-1}, we will see whether we need to calculate it, okay? Then let us look at the residual. So, r_k^T will simply be $\hat{A}\hat{x} - \hat{b}$. What should this be? Exactly, right? This is going to be, maybe we should just write it out more explicitly.

So, this is $L^T A L x$, and b hat was $L^T b$. So, if I take L^T common, what do I have? $Ax - b$, so, $L^T r_k$, okay?

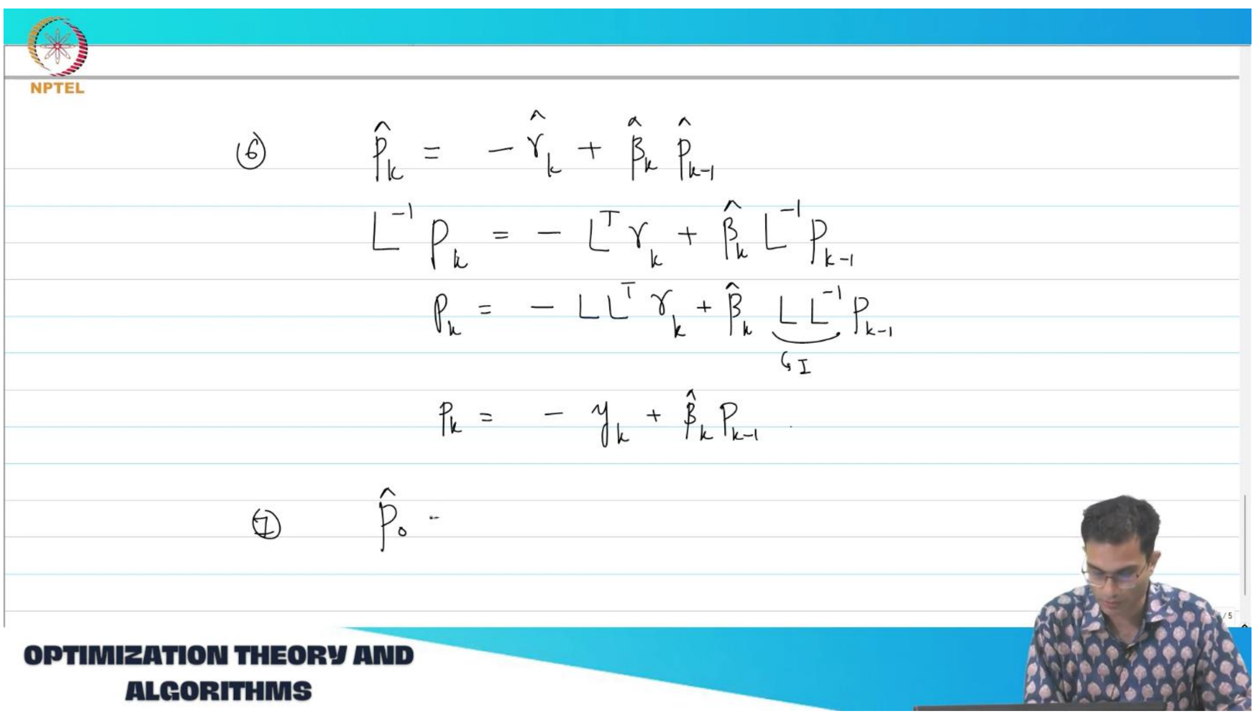
So, the new residual is given in terms of the old residual, and there is a factor L over there, okay. What about α_k? So, if you look at the simplified expression for α, that we derived, what was the expression for α, right? Now, we already know p hats and r hats in terms of the old guys. So, $r_k^T r_k$ ^what would that simplify to? What is r_k^T? Can I write it in terms of the old r? See, I want to get all the new expressions in terms of the old expressions as much as possible. So, I can substitute from step 2. I am going to get $r_k^T L L^T r_k$. What about the denominator?



$p_k^T p_k A p_k$, it remains the same, okay. So, this expression looks a little different from the original expression. The original expression obviously did not have an LL^T hanging out there in between. So, this is the first difference that we notice.

Let us see how we can resolve it. Next, what do I need? I got my p, r, α, what else do I need? I need my β's, right? So, let us look at $\hat{\beta}_{k+1}$. The simplified expression for that was $\frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$. This is very similar to step 3.

I am just going to get $\frac{LL^T r_k}{r_k^T LL^T r_k}$, right? So, you notice that both α and β have a kind of a funny thing happening: this LL^T is being introduced somewhere, and if you look at if you remember our CG routine, I needed the values of α, β to go ahead next, right? So, these two guys seem to be a little bit different, but the difference has a pattern, and the difference is the appearance of this LL^T is always being accompanied by the residual term, okay. So, the clever thing to do would be at this point, I introduce a new variable: let us call this $LL^T r_k$ to be a new variable, let us call it y_k. So, having done that, what happened to my α? This became $\frac{r_k^T y_k}{p_k^T A p_k}$ and this became $\frac{r_{k+1}^T y_{k+1}}{r_k^T y_k}$. So, at least there is a simple-looking form that is coming up, okay.

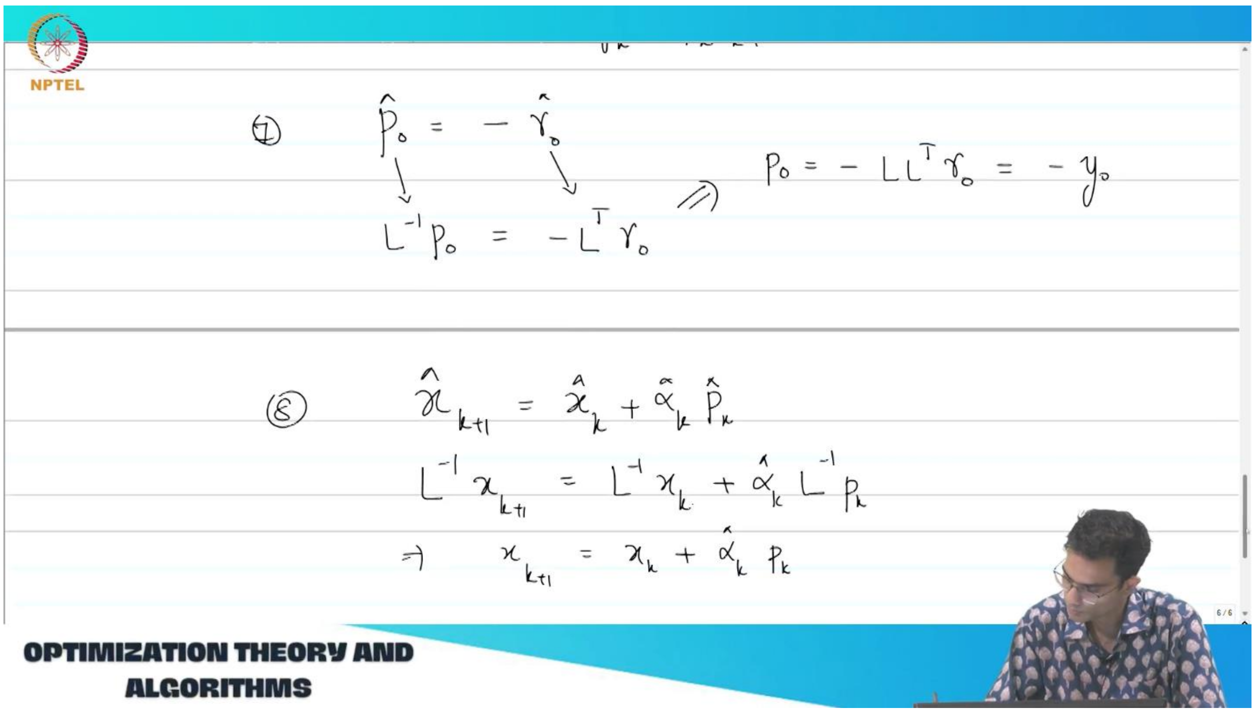


So, I have got my α's, β's, I need to now know how do I go to the next conjugate direction. So, that let us say step 6 is. So, $\hat{p}_k$ I write it in terms of $-\hat{r}_k + \hat{\beta}_k$ times what? The previous conjugate direction. So, $\hat{p}_{k-1}$, okay. Now, $\hat{p}_k$ was what in terms of the old p's? L^{-1}.

L^{-1}. So, this is going to be $L^{-1}p_k = -L^T r_k + \beta_k \hat{p}_k$. Let us leave that as $\beta_k \hat{p}_k$ because I have an expression for it, and $\hat{p}_k$ was again $L^{-1}p_{k-1}$. What do I do here? Multiply by L on the left. So, that is going to give me my p_k. So, minus L, our friend appears again; LL^{-1} cancels out, right?

So, basically, this is the new relation I get: $p_k = -y_k + \beta_k \hat{p}_{k-1}$, right? Not bad. So, this is sort of telling me that even though I have the new system, the "hatted" system $\hat{A}\hat{x} = \hat{b}$, my update equations still seem to work with the original p_k's, but with the exception that there is one new term, y_k, floating around, right?

Now, to confirm whether all of this goes through, I need to look at step 1. Step 1 is the first conjugate direction, right? So, what is $\hat{p_0}$? In any CG method, p_0 would be $-r_0$. So, this is going to be equal to $-\hat{r_0}$, all right.



Now, $\hat{p_0}$ is what in terms of the original p_0? L^{-1}, right. So, this guy is $L^{-1}p_0$. What is this guy? r_0 is $L^T r_0$, so we get $-L^T r_0$, okay? So, these are obviously equal. This implies what? $p_0 = -LL^T r_0$, there it is again. So, this is $-y_0$, okay?

So, I have got my p_0. The final piece of the puzzle is how do I go to the new x? Right, so let us look at that. The new x is going to be, that is my expression for it, right? Now, what is $\hat{x}$ in terms of L? Is it L or L^{-1}? L^{-1}, right. So, this is $L^{-1}x_k + 1$, Lx_k, $x_k \alpha$, I have an expression. I am going to leave it, and $\hat{p}$ is again $L^{-1}p_k$, right. So, basically, I get the same expression.

So, that is quite interesting, right? Because what is it saying? Let us just try to recap from the beginning what we did. We started with a new system of equations, which was $\hat{A}\hat{x} = \hat{b}$. So far, we have not spoken about how do we get L, but let us assume we have it. Then we saw that after having done all of this algebra, the starting direction p_0 was simply $-y_0$, which looked similar to the first CG step, right? After p_0, what did I need to do? For example, to get to the new x, to get to x_1, what would I do? I would use the relation in equation (8) because I have my p_0, I have my α_k expression also, I have my x_0, and I get my new x from equation (6). I get my new p_k's, and I know how to update my p_k's, right?

And these are all the original p_k's of the original A matrix. I have not had to recompute or do anything. These are the original p_k's that were there for the bad condition number, except everywhere we notice that there is one new variable hanging around whose name is y_k. So, there is obviously, there is no free lunch; the price that I am paying is that you need to compute this y_k, right? So, that is the price that I have to pay. If you just look at that, $LL^T r_k = y_k$, okay? Did we ever have to calculate L^{-1}? It appeared as an intermediate step, but we never actually needed it anywhere, right? Everything was updated, and the only extra term was this LL^T appearing everywhere, right? Okay, so this needs to be done.