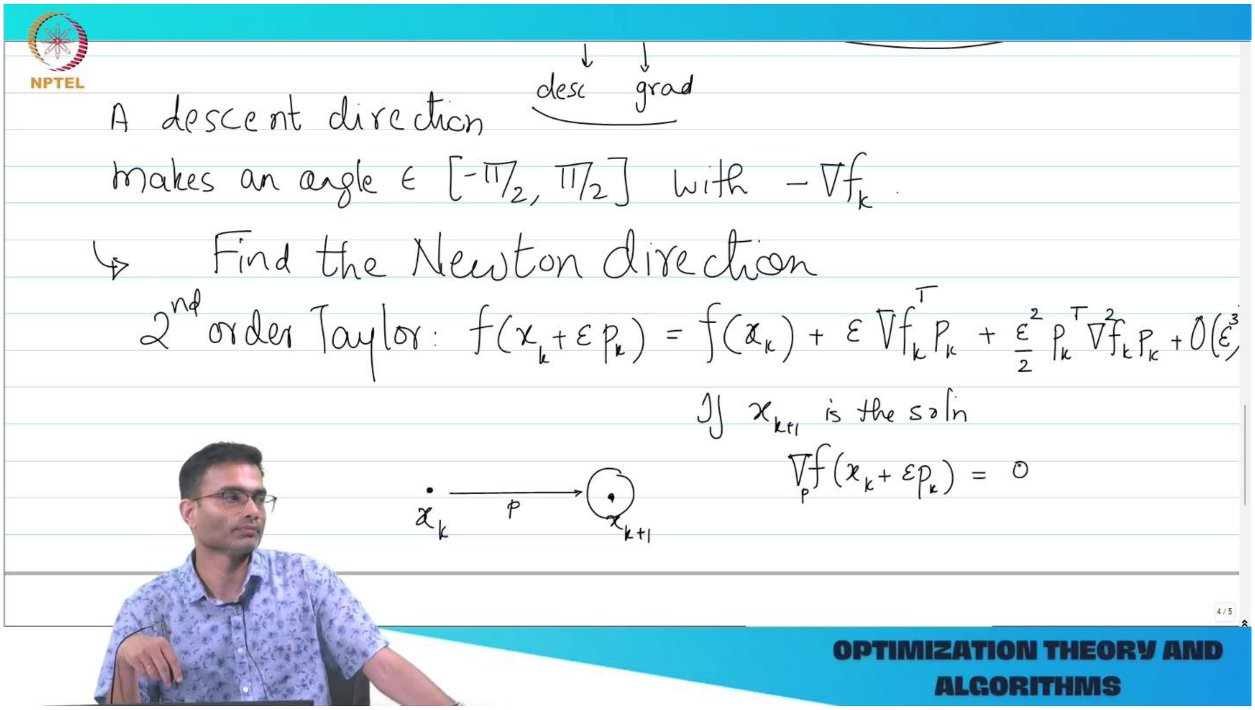
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 03
Lecture - 20

Unconstrained Optimization - 6 - Properties of Descent Directions and Newton Direction

So, we have spoken about. When I say legitimate descent direction, all that you have to ensure is the angle is less than 90 degrees with the negative gradient, then you are fine. Then you can call it a descent direction.

Now, let us prove or let us find the Newton direction. What is the main tool that we used in the previous development? Taylor's theorem to first order. Now, we know that the Newton method is a second order method. So, what should be our starting point? Second order Taylor's theorem, right?



Let us write that out:

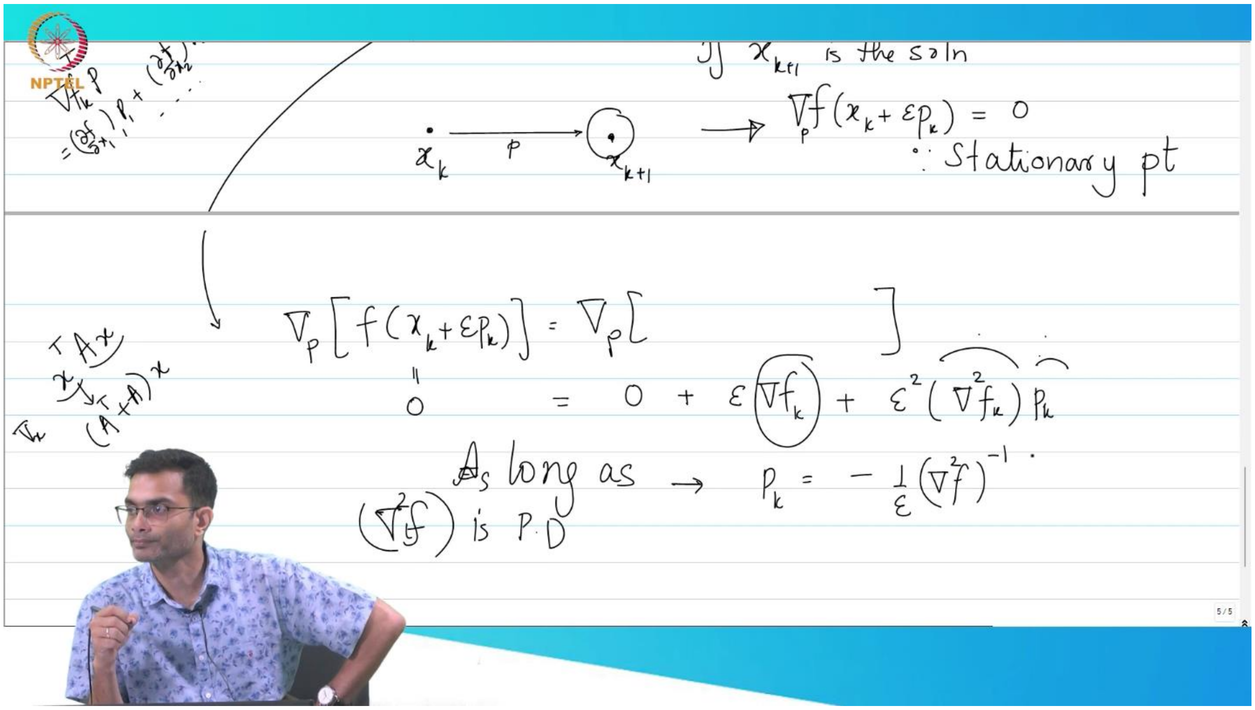
$$f(x_k + \epsilon p_k) = f(x_k) + \epsilon \nabla f(x_k) + \frac{\epsilon^2}{2} p_k^\top H(x_k) p_k + O(\epsilon^3)$$

where $H(x_k)$ is the Hessian matrix. This is a good approximation as long as ϵ is small, as it takes into account the second-order term.

Now, let us take a step from x_k to x_{k+1}. In the dream world, if x_{k+1} is the solution, then the gradient at this point should be zero:

$$\nabla f(x_k + \epsilon p_k) = 0.$$

Here, ϵp_k is the step we are optimizing. We want to reach a stationary point.



Using the Taylor expansion:

$$\nabla_p f(x_k + \epsilon p_k) = \nabla_p \left(f(x_k) + \epsilon \nabla f(x_k) + \frac{\epsilon^2}{2} p_k^\top H(x_k) p_k \right) = 0.$$

Now, let us take the gradient with respect to p_k term by term.

For the first term, $f(x_k)$ does not depend on p_k, so its derivative is zero.

For the second term, we get:

$$\epsilon \nabla f(x_k).$$

For the third term, applying the product rule, we have:

$$\epsilon^2 H(x_k) p_k.$$

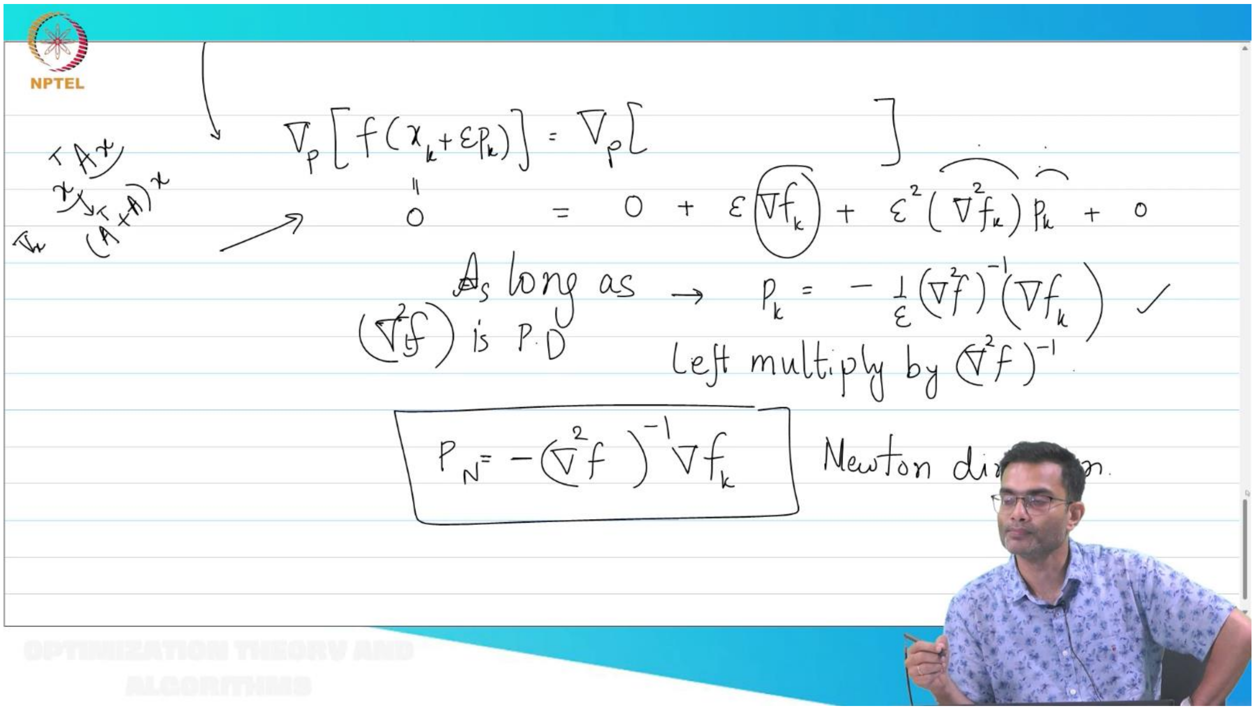
Since we are at a minimum, we can set this expression to zero:

$$\epsilon \nabla f(x_k) + \epsilon^2 H(x_k) p_k = 0.$$

Multiplying both sides by the inverse of the Hessian, we get the Newton direction:

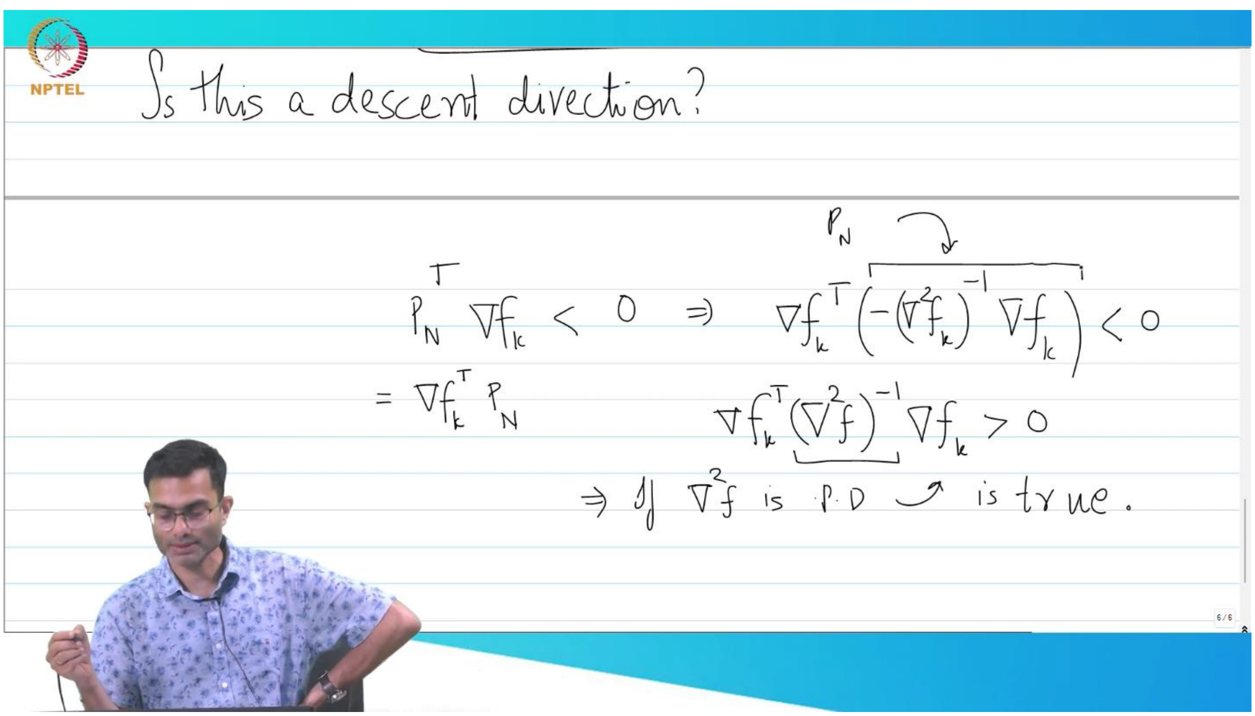
$$p_k = -H(x_k)^{-1} \nabla f(x_k).$$

Now, for the direction to be legitimate, we must ensure that it is a descent direction, which means the angle between p_k and the gradient $\nabla f(x_k)$ must be less than 90 degrees. This condition can be checked by verifying that:



$$p_k^{\top} \nabla f(x_k) < 0.$$

Substituting $p_k = -H(x_k)^{-1} \nabla f(x_k)$:



$$\nabla f(x_k)^\top H(x_k)^{-1} \nabla f(x_k) > 0.$$

This condition holds if the Hessian is positive definite. If the Hessian is not positive definite, the Newton direction may not be a descent direction, and we may ascend instead.

In practice, if the Hessian is not positive definite, we can switch to other methods like gradient descent or conjugate gradient.

Finally, note that for the Newton direction to be valid, the Hessian must be invertible, though it does not need to be positive definite at all times. However, for the direction to guarantee a descent, the Hessian should be positive definite.