Optimization Algorithms: Theory and Software Implementation

Prof. Thirumulanathan

Department of Mathematics

Institute of IIT Kanpur

Lecture: 49

Hello everyone. This is lecture 4 of week 10. In the previous two lectures, we were looking at the affine scaling method. If you recall, this is a method where we start from an interior point and try to find a direction that is both a descent direction and ensures the point lies inside the feasible polytope. We were also discussing the stopping criterion. We learnt a little bit about duality and we said that if the primal value and the dual values are close enough, that is within a tolerance, then we will stop the iterations. That will be used as a stopping criterion.

If you recall where we ended up in the last class, please note that if the primal is

min $c^T x$

subject to

$A x = b, x \geq 0$,

we saw that the dual is actually maximizing $-\mu^T b$ subject to $A^T \mu \geq -c$.

There were a few errors that I would like to correct.

When we write $\mu^T(A x - b)$, we would actually get $-\mu^T b$ and not $\mu^T b$.

If you recall what has happened, it is we are minimizing:

$c^T x + \mu^T(A x - b) - \lambda^T x$.

If you differentiate this with respect to x, you get:

$c + A^T \mu - \lambda = 0$.

This means $c^T x + \mu^T A x - \lambda^T x = 0$.

The remaining part is just $-\mu^T b$. I made a mistake; I did not include the minus. The problem thus becomes:

maximize $-\mu^T b$

subject to

$A^T \mu \geq -c$.

Suppose I write μ as $-\mu$.

Note that μ is not constrained. If instead of μ being 5, I write it as -5, or any μ as $-\mu$, this problem would equivalently be written as:

max $\mu^T b$

because μ has become $-\mu$, and the constraint becomes $-A^T\mu \geq -c$.

Multiplying throughout by a minus,

you will get $A^T\mu \leq c$.

So, the dual is:

max over μ $\mu^T b$

subject to

$A^T\mu \leq c$.

This is more pleasing to the eye than writing maximum of $-\mu^T b$. We will consider this as the dual.

The moral of the story is that if the primal problem is:

min $c^T x$ subject to $A x = b$, $x \geq 0$,

then the dual is:

max $\mu^T b$ (or $b^T\mu$) subject to the constraint that $A^T\mu \leq c$.

What this means is that if we get a μ in such a way that, if you look at the theorem, when you have a linear programming problem where a solution exists, then the values of the objective functions of both the primal and the dual are one and the same.

If there exists an x which is primal feasible, meaning it satisfies $A x = b$, $x \geq 0$, and there exists μ which is dual feasible, meaning μ satisfies $A^T\mu \leq c$, then we have the solutions for both primal and dual problems.

The good part about linear programming is that when your primal is a linear programming problem, the dual also turns out to be a linear programming problem. In this case, the primal was a linear programming problem, and the dual turned out to be a linear programming problem in μ.

One of the ways of finding the solution for the primal is to find an x that is primal feasible and a μ that is dual feasible such that $c^T x = b^T\mu$.

If you can find an x and μ such that all of these conditions hold, then we actually have the solution for both the primal and dual problems.

This is how we choose the stopping criterion. If you recall the steps in the affine scaling method, we start with an initial point, we do a centering, and then for the centered vector, we move along the direction of the steepest descent direction—not just the steepest descent, but it is actually a projection of the steepest descent onto the null space of the matrix A. We move along the direction and after moving, we de-center it to get back the value in the original space.

We can keep continuing that until some stopping criterion, and we were trying to figure out the stopping criterion using the duality method. We just saw that $c^T x$ will be equal to $b^T \mu$ if x is the minimizer.

We will stop if the modulus of $c^T x - \mu^T b$ is less than or equal to some tolerance, say 10^{-6}.

If it is greater than the tolerance, then continue doing the loop. When x and μ are such that the difference between $c^T x$ and $\mu^T b$ is within some tolerance, then we stop.

This is basically the idea, but we are still not done yet because it is good that we have all these theorems, but now what is the value of μ?

Suppose you start with some x_0. You do the centering, move along some direction, do the de-centering; all that is fine. But given an x_0 or x_k or any x for that matter, how do I find the corresponding μ? That needs a particular step, which we will do now.

Consider the primal problem: min $c^T x$ subject to $A x = b$, $x \geq 0$.

If you write the Lagrangian for the primal, $L(x, \lambda, \mu) = c^T x + \mu^T (b - A x) - \sum \lambda_i x_i$ for i = 1 to n. To solve this, you take the gradient with respect to x and equate it to 0. This tells us that:

$c - A^T \mu - \lambda = 0$,

so

$A^T \mu + \lambda = c$.

This should be satisfied if x is the minimizer.

We also have the complementary slackness conditions where $\lambda_i x_i = 0$ for all i = 1 to n.

This is all part of KKT conditions.

So, the gradient of L equals 0 and $\lambda_i x_i = 0$ for all i.

This tells us that at least one of λ_i or x_i has to be 0. If x_i is positive, that means λ_i is 0, so $(A^T \mu)_i = c_i$ for that component.

If $x_i = 0$, then the product is zero. What I can write is that $x_i (A^T \mu - c)_i = 0$ for each component i. This is true for all i in 1 to n.

In other words, if capital X is diag(x), I can write this in matrix form as:

$X (A^T \mu - c) = 0$.

This is a condition that relates x and μ.

So, we can find the least squares solution for which μ satisfies $X (A^T \mu - c) = 0$.

When x is the solution, you will see that $X (A^T\mu - c) = 0$.

If x is not the solution, we are going to choose a μ such that $||X A^T\mu - X c||^2$ is minimized.

If $X A^T$ were invertible, I could have just written $(X A^T)^{-1} X c$.

But $X A^T$ need not be a square matrix. Note that A is an $m \times n$ matrix, A^T is $n \times m$, and X is $n \times n$, so $X A^T$ is an $n \times m$ matrix, which is rectangular and not invertible.

What I am going to do is multiply both sides by A X.

So, you will have:

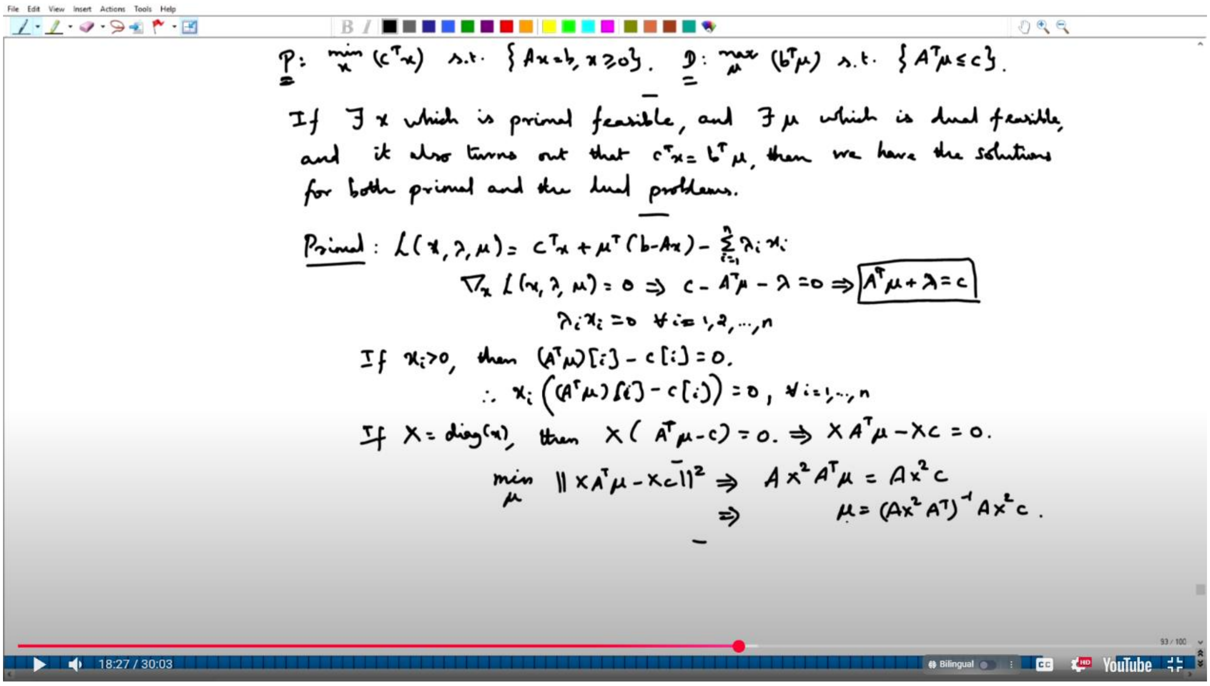
$A X^2 A^T\mu = A X^2 c$.

Now, $A X^2 A^T$ is an $m \times m$ matrix, and it can be shown that it is invertible; its rank is m since the rank of A is m.

Thus, you will have:

$\mu = (A X^2 A^T)^{-1} A X^2 c$.

(Refer Slide Time 18:27)



If x is the solution, then you will satisfy $X A^T\mu - X c = 0$, which means $A X^2 A^T\mu = A X^2 c$, and thus you will get μ such that $X (A^T\mu - c) = 0$.

Since x need not be the solution and could be any arbitrary point, what we are doing is minimizing the norm squared of $X (A^T\mu - c)$, because of which we are getting μ to be $(A X^2 A^T)^{-1} A X^2 c$.

Now, when we are given x, we have a value of μ.

Now, I can write down the steps as follows.

Suppose you have a primal problem of this form:

min $c^T x$ subject to $A x = b$, $x \geq 0$.

The steps for the algorithm are:

1. Convert the given LP to its standard form.

2. Initialize x_0 as an interior point of the feasible polytope. (This means $A x_0 = b$ and $x_0 > 0$, all components strictly positive).

3. Initialize $k = 0$.

4. Write capital X_0 as $\text{diag}(x_0)$ (the diagonal matrix formed from the elements of x_0).

5. Then, compute the initial dual variable estimate:

   $\mu_0 = (A X_0^2 A^T)^{-1} A X_0^2 c$.

While $|c^T x_k - b^T \mu_k| >$ tolerance (your chosen tolerance, e.g., 10^{-6}):

* Centering: Calculate the centered vector:

   $y_k = X_k^{-1} x_k$

(This will be the vector of all ones, 1).

So, note that if you consider this particular problem:

min $c^T x$

subject to

$A x = b$, $x \geq 0$.

The projection of $-c$ onto the null space of A is $-(I - A^T(A A^T)^{-1}A) c$.

In the Y space this is slightly different.

Recall that y is defined as $y_k = X_k^{-1} x$, so $x =$ capital $X_k y_k$.

Substitute capital $X_k y_k$ into the problem.

The objective becomes: $c^T x = c^T (X_k y)$ and the constraint becomes:

$A x = A (X_k y) = b$, $x \geq 0$ which implies $X_k y \geq 0$.

Since X_k is a diagonal matrix with positive entries (as we are in the interior), this is equivalent to $y \geq 0$.

In the y-space:

- The objective vector c has become $X_k c$.

- The constraint matrix A has become A X_k.

The projected steepest descent direction in the y-space becomes: $-(I - (A X_k)^T[(A X_k)(A X_k)^T]^{-1} (A X_k)) (X_k c)$

Note that:

- $(A X_k)^T = X_k^T A^T = X_k A^T$ (since X_k is diagonal and symmetric).
- $(A X_k)(A X_k)^T = (A X_k)(X_k A^T) = A X_k^2 A^T$.

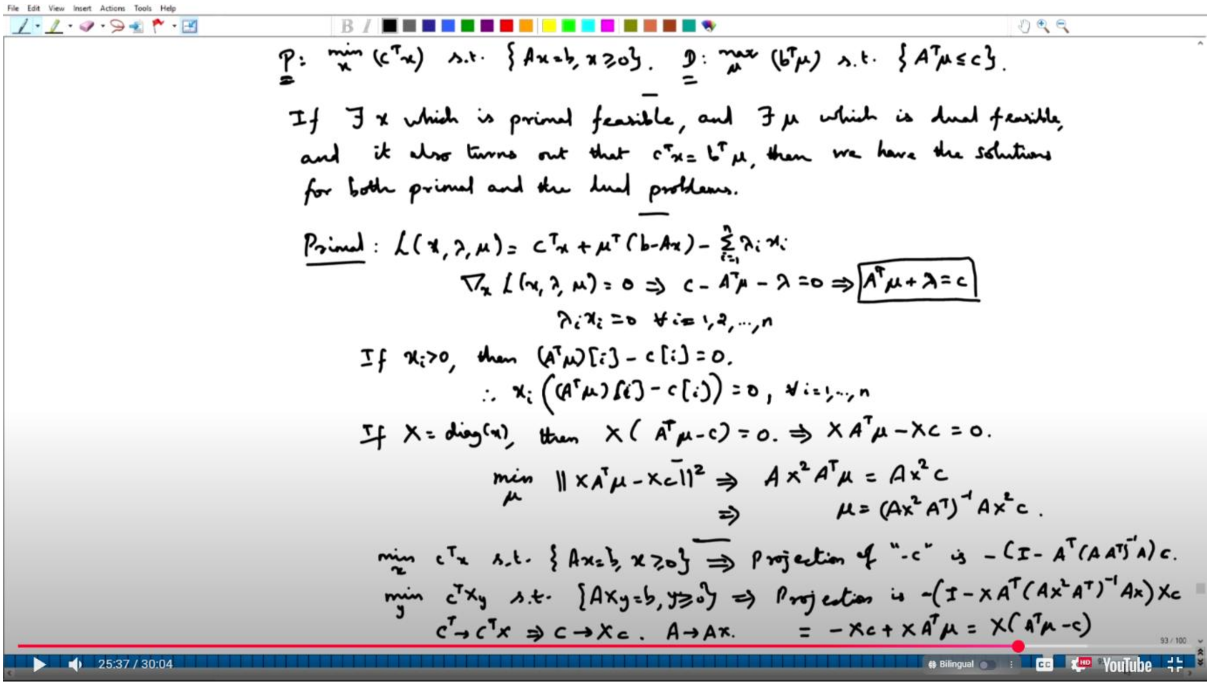
So the expression becomes: $-(I - (X_k A^T)(A X_k^2 A^T)^{-1} (A X_k)) (X_k c)$

This simplifies to: $-X_k c + (X_k A^T)(A X_k^2 A^T)^{-1} (A X_k) (X_k c)$

Notice that the term $(A X_k^2 A^T)^{-1} (A X_k) (X_k c)$ is the least-squares solution for μ derived earlier, $\mu_k = (A X_k^2 A^T)^{-1} A X_k^2 c$.

Therefore, the entire expression simplifies to: $-X_k c + X_k A^T \mu_k$ which is $X_k (A^T \mu_k - c)$. So, the direction d_k is $X_k (A^T \mu_k - c)$.

(Refer Slide Time 25:37)



* Calculate the Direction: Compute the direction vector in the transformed space:

$d_k = X_k (A^T\mu_k - c)$.

* Choose the Step Size α_k: The new point in the transformed space is $y_{k+1} = 1 + \alpha_k d_k$. We require $y_{k+1} \geq 0$ to maintain feasibility in the next step. This requires:

$1 + \alpha_k d_k \geq 0$.

This inequality will be violated if any component d_{kj} is negative and α_k is too large. The most restrictive constraint for α_k comes from the most negative component of d_k.

Therefore, the maximum allowed step size is:

α_k = min over $\{j: d_{k,j} < 0\}$ of $(-1 / d_{k,j})$.

(Often, a fraction of this maximum, like 0.99 * α_k, is used to ensure we stay strictly inside the feasible region).

* Take the Step in the Transformed Space:

$y_{k+1} = 1 + \alpha_k d_k$.

* De-centering: Convert the new point back to the original space:

$x_{k+1} = X_k y_{k+1}$.

* Update: Form the new diagonal matrix from the updated point:

$X_{k+1} = \text{diag}(x_{k+1})$.

* Update the Dual Variable Estimate:

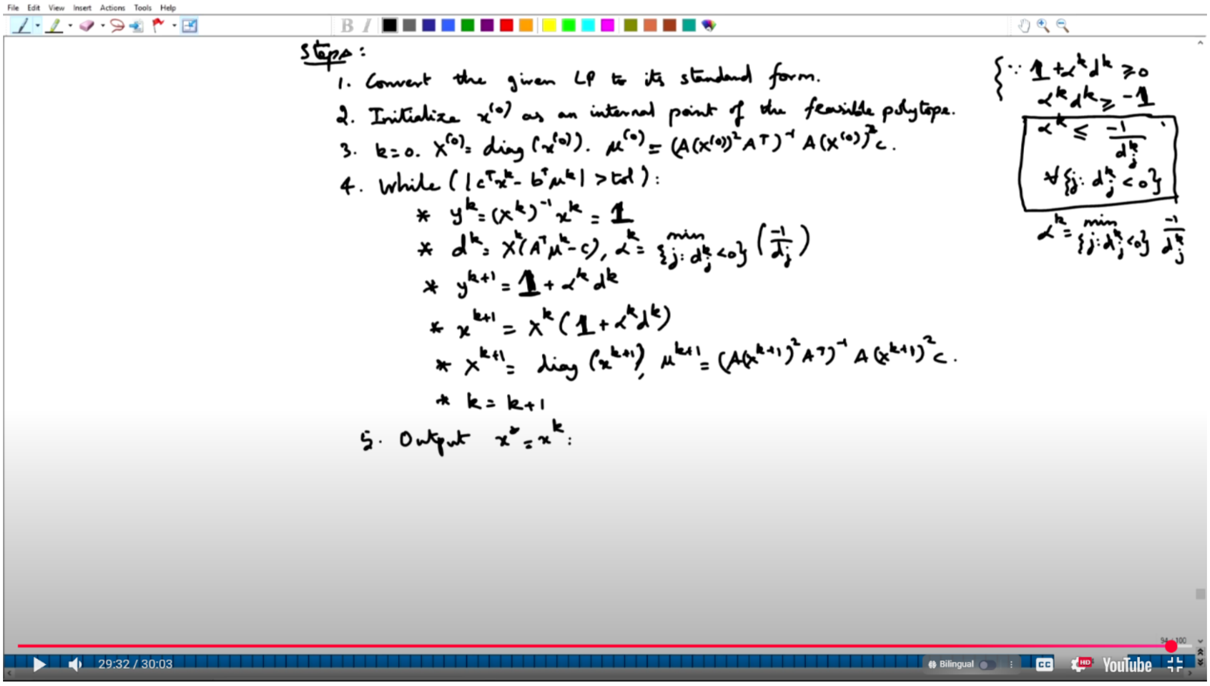
$\mu_{k+1} = (A X_{k+1}^2 A^T)^{-1} A X_{k+1}^2 c$.

* Increment the Counter: Set k = k + 1.

End While

The output is the (approximate) solution $x^* = x_k$.

(Refer Slide Time 29:32)



We will discuss this once more at the start of the next lecture and then work out some examples. Thank you.